Fronthaul Constrained Cloud and Fog Radio Access Networks: An Information Theoretic View

Shlomo Shamai

Department of Electrical Engineering
Technion−Israel Institute of Technology

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

IPAM Conference on Emerging Wireless Networks: February 10, 2017

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

Technion
Israel Institute of Technology

Outline

I. Introduction

II. Uplink
   A. Distributed fronthaul compression
   B. Structured coding
   C. Multi-hop fronthaul topology

III. Downlink
   A. Multivariate fronthaul compression
   B. Structured coding
   C. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks
   A. Edge caching and F-RAN
   B. Information-theoretic analysis
   C. Optimization of signal processing strategies

V. Outlook

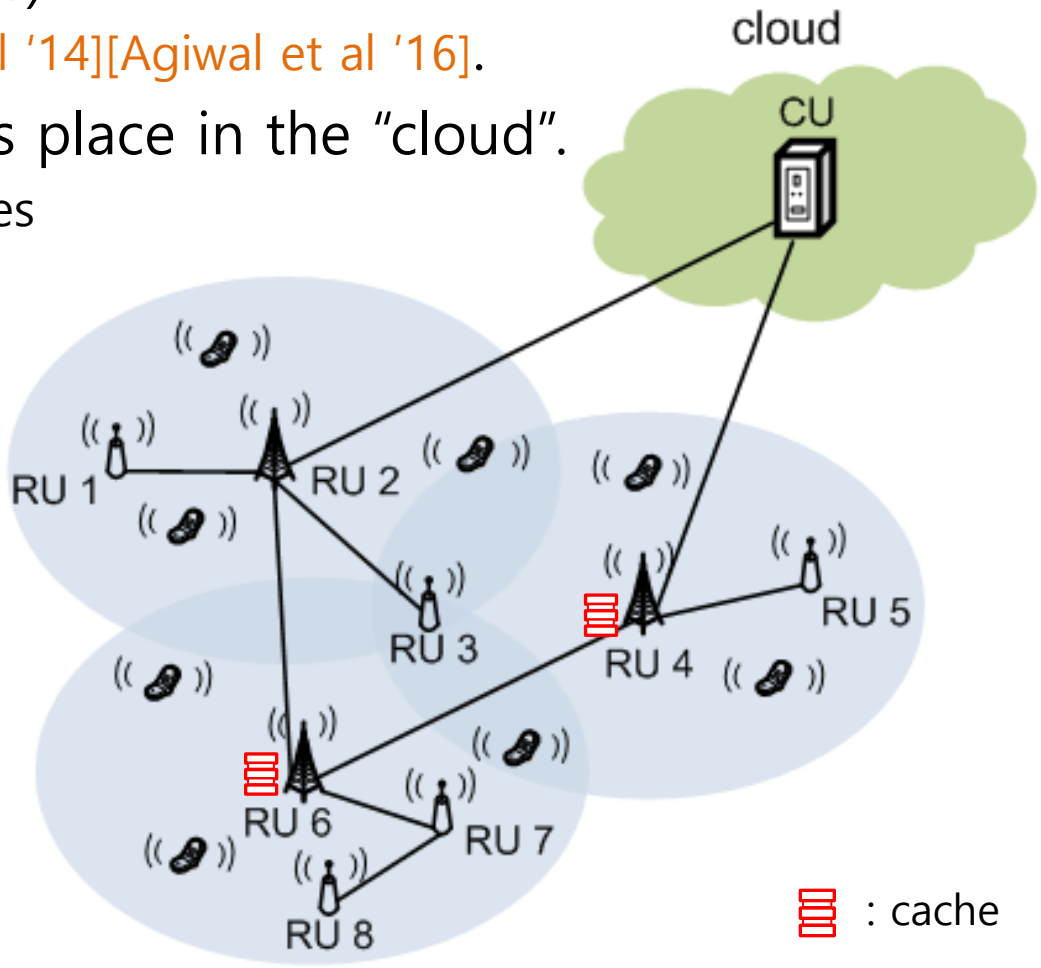
Technion
Israel Institute of Technology

Outline

I. Introduction
II. Uplink
  A. Distributed fronthaul compression
  B. Structured coding
  C. Multi-hop fronthaul topology
III. Downlink
  A. Multivariate fronthaul compression
  B. Structured coding
  C. Inter-cluster multivariate fronthaul compression
IV. Fog Radio Access Networks
  A. Edge caching and F-RAN
  B. Information-theoretic analysis
  C. Optimization of signal processing strategies
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.



Technion Israel Institute of Technology

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]

Technion Israel Institute of Technology

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]

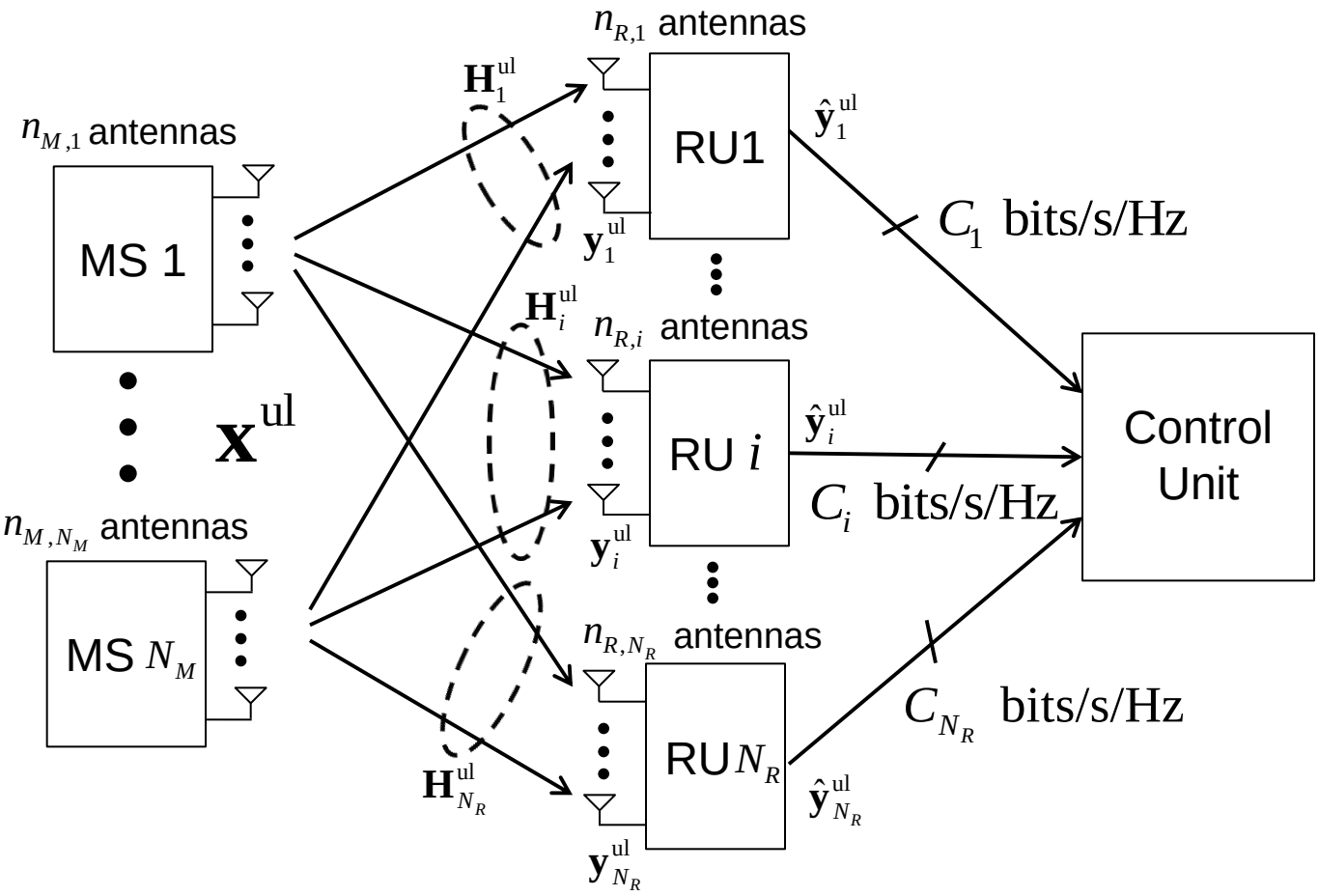
Technion
Israel Institute of Technology

Outline

I. Introduction
II. Uplink
A. Distributed fronthaul compression
B. Structured coding
C. Multi-hop fronthaul topology
III. Downlink
A. Multivariate fronthaul compression
B. Structured coding
C. Inter-cluster multivariate fronthaul compression
IV. Fog Radio Access Networks
A. Edge caching and F-RAN
B. Information-theoretic analysis
C. Optimization of signal processing strategies
V. Outlook

Technion
Israel Institute of Technology

Basic Settings



Basic Settings

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

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

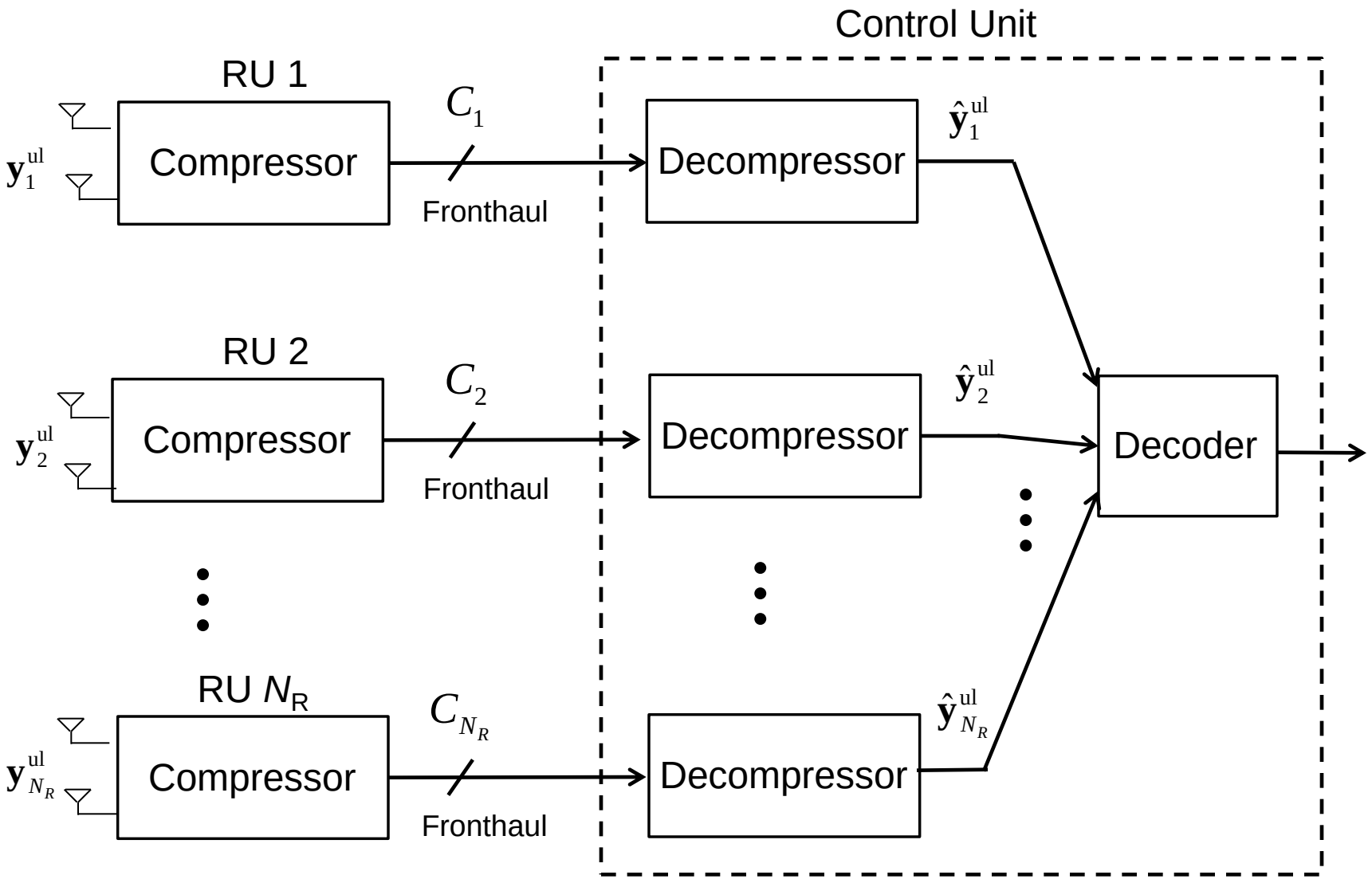
where $\mathbf{H}_i^{\mathrm{ul}} = \left[\mathbf{H}_{i,1}^{\mathrm{ul}}, \cdots, \mathbf{H}_{i,N_M}^{\mathrm{ul}} \right]$: channel matrix toward to RU i;

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

$\mathbf{z}_i^{\mathrm{ul}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{\mathbf{z}_i^{\mathrm{ul}}})$: 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]

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

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

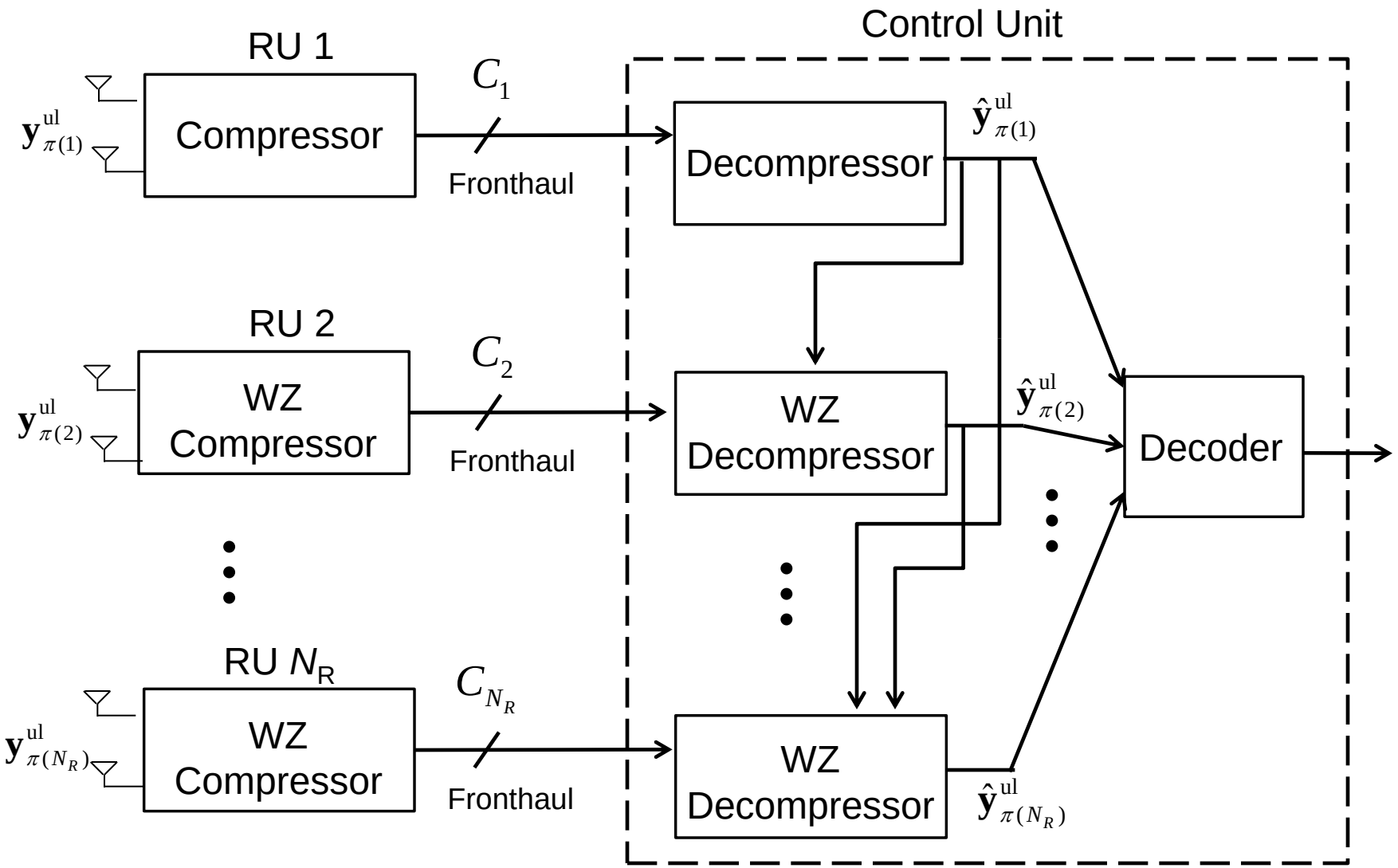
- If the fronthaul capacity C_i satisfies

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

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

Distributed Fronthaul Compression

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



$\pi : \mathcal{N}_R \to \mathcal{N}_R$ denotes a permutation of RUs' indexes where $\mathcal{N}_R \subseteq \{1, \ldots, N_R\}$.

Distributed Fronthaul Compression

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

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

$$I(\mathbf{y}^{\mathrm{ul}}_{\pi(i)};\hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(i)} \mid \hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(1)},\hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(2)},\ldots,\hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(i-1)}) \le C_{\pi(i)}.$$

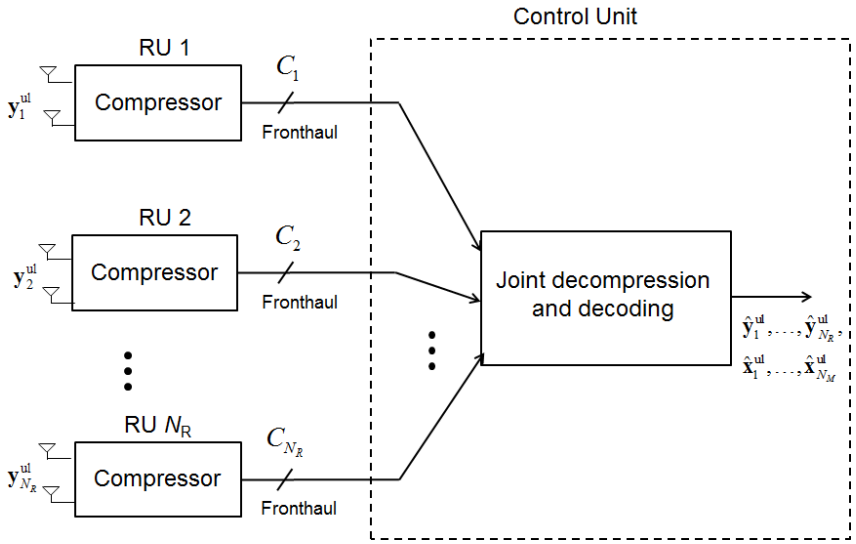
- After the quantized IQ signals $\hat{\mathbf{y}}^{\mathrm{ul}}_1,\ldots,\hat{\mathbf{y}}^{\mathrm{ul}}_{N_R}$ are recovered, the CU then performs joint decoding of the signal $\mathbf{x}^{\mathrm{ul}}$ sent by all MSs.
  - The uplink sum-rate is given by

$$R^{\mathrm{ul}}_{\mathrm{sum}} = I(\mathbf{x}^{\mathrm{ul}};\hat{\mathbf{y}}^{\mathrm{ul}}_1,\ldots,\hat{\mathbf{y}}^{\mathrm{ul}}_{N_R}).$$

Distributed Fronthaul Compression

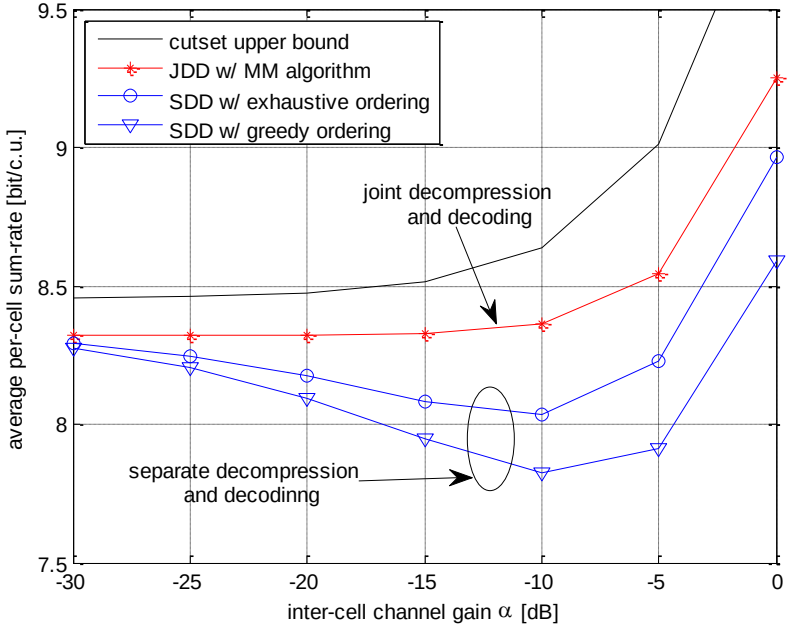
[Sanderovich et al '09] [dCoso-Simoens '09] [Zhou-Yu '11]

- 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]:

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



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 $\mathbf{\Omega}^{\mathrm{ul}}_{\pi(1)},\ldots,\mathbf{\Omega}^{\mathrm{ul}}_{\pi(N_R)}$ following the same order π employed for decompression.
  - At the ith step, for fixed (already optimized) covariances $\mathbf{\Omega}^{\mathrm{ul}}_{\pi(1)},\ldots,\mathbf{\Omega}^{\mathrm{ul}}_{\pi(i-1)}$, the covariance $\mathbf{\Omega}^{\mathrm{ul}}_{\pi(i)}$ is obtained by solving

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

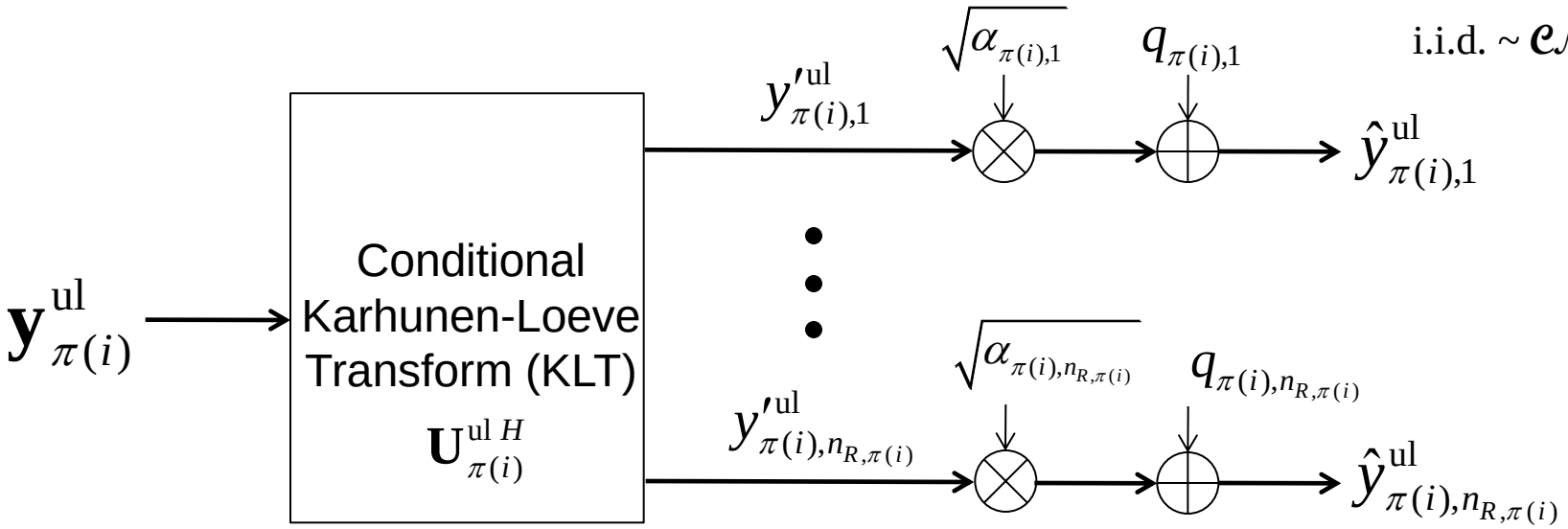
$$\text{s.t.} \quad I(\mathbf{y}^{\mathrm{ul}}_{\pi(i)};\hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(i)} \mid \hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(1)},\hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(2)},\ldots,\hat{\mathbf{y}}^{\mathrm{ul}}_{\pi(i-1)}) \le C_{\pi(i)}.$$

Distributed Fronthaul Compression

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

- Optimal WZ compressor [dCoso-Simoens '09]

$q_{\pi(i),1},\ldots,q_{\pi(i),n_{R,\pi(i)}}$ i.i.d. $\sim \mathcal{CN}(0,1)$.



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

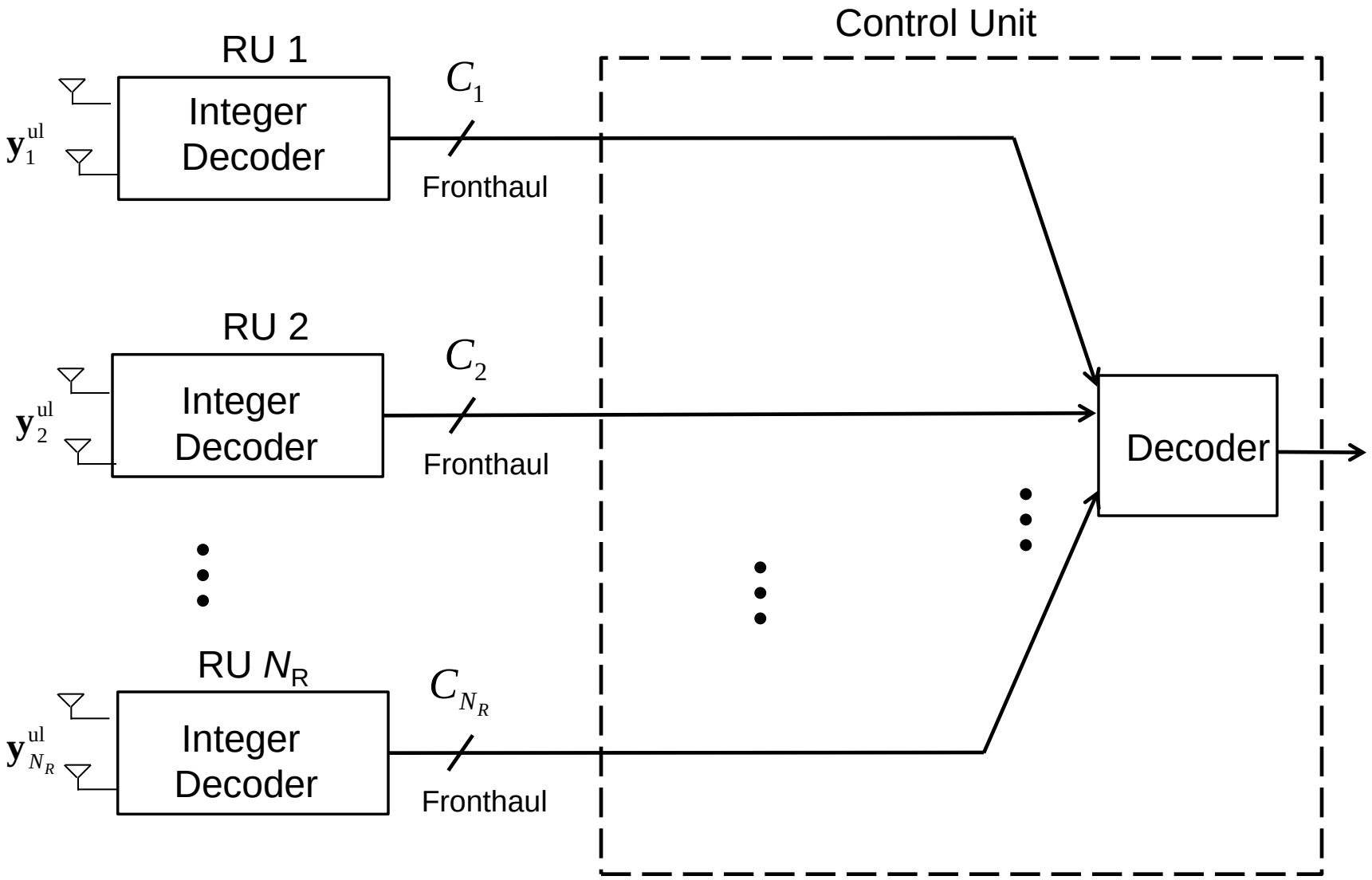
Technion
Israel Institute of Technology

Outline

I. Introduction
II. Uplink
  A. Distributed fronthaul compression
  B. Structured coding
  C. Multi-hop fronthaul topology
III. Downlink
  A. Multivariate fronthaul compression
  B. Structured coding
  C. Inter-cluster multivariate fronthaul compression
IV. Fog Radio Access Networks
  A. Edge caching and F-RAN
  B. Information-theoretic analysis
  C. Optimization of signal processing strategies
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 $C_1 = \ldots = C_{N_R} = C$, achievable rate per MS is given by

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

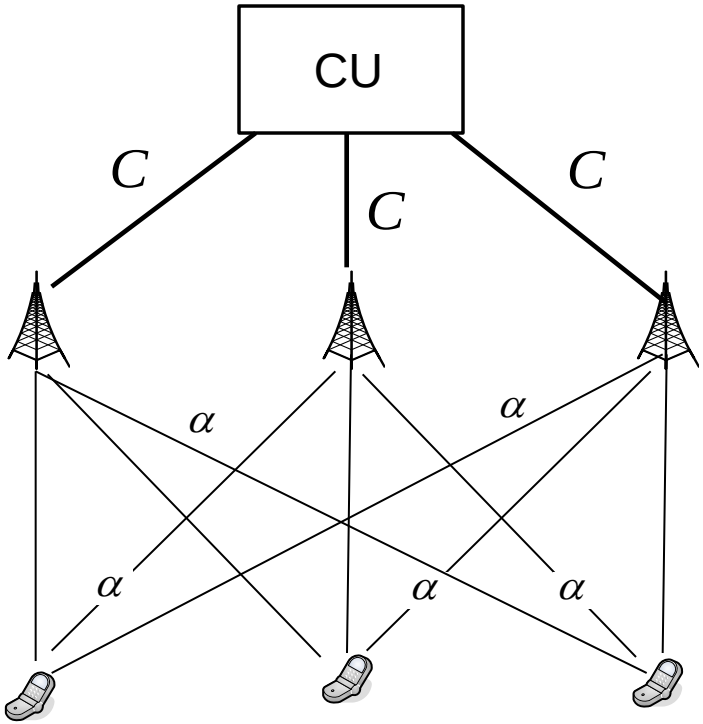
Integer penalty: The signal received at each RU is sum with *non-integer weights* of the codewords transmitted by MSs.

$\mathbf{h}_l$: 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

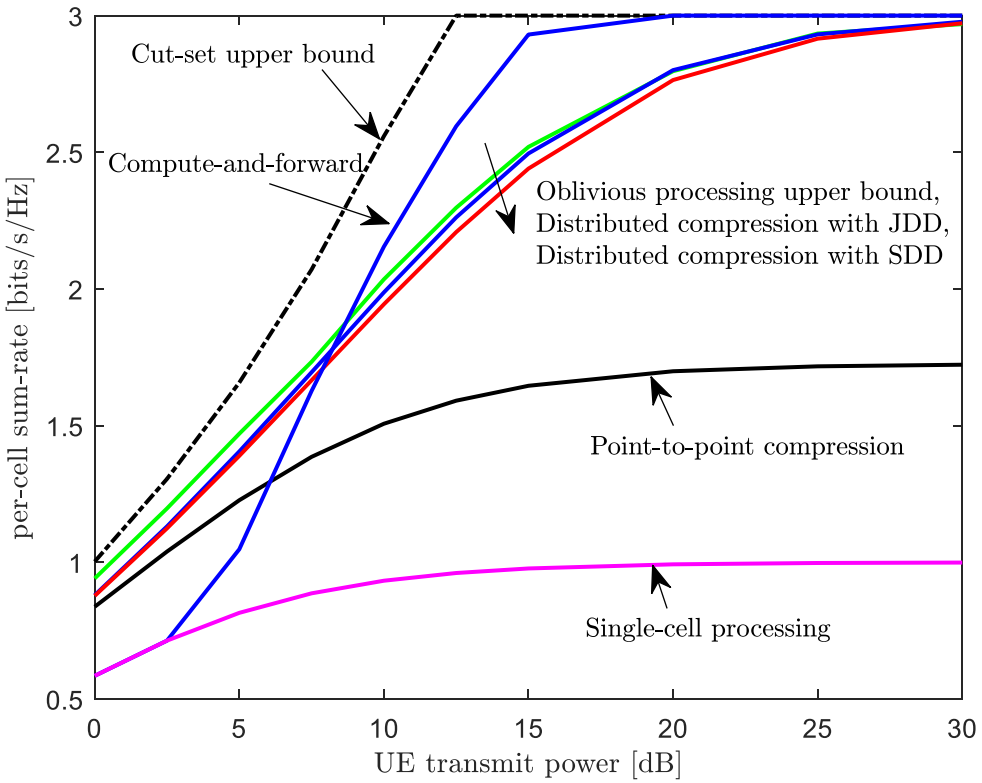
Technion Israel Institute of Technology

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

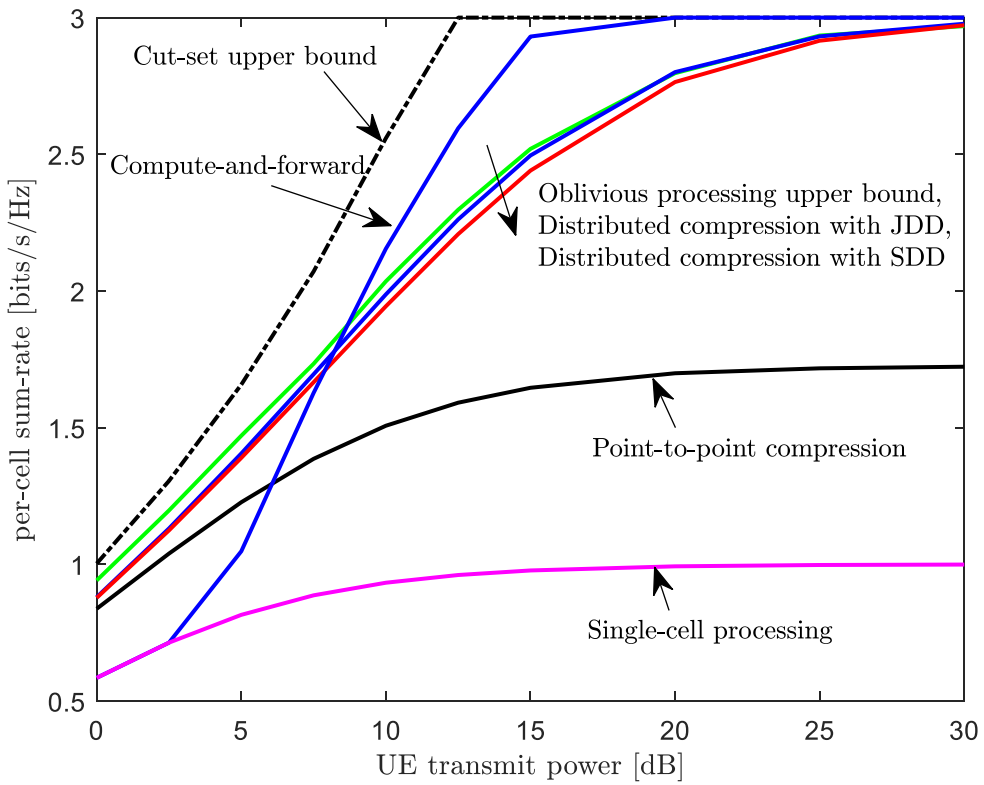
$\alpha = 1/\sqrt{2}$ and $C = 3$ bit/s/Hz



- 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

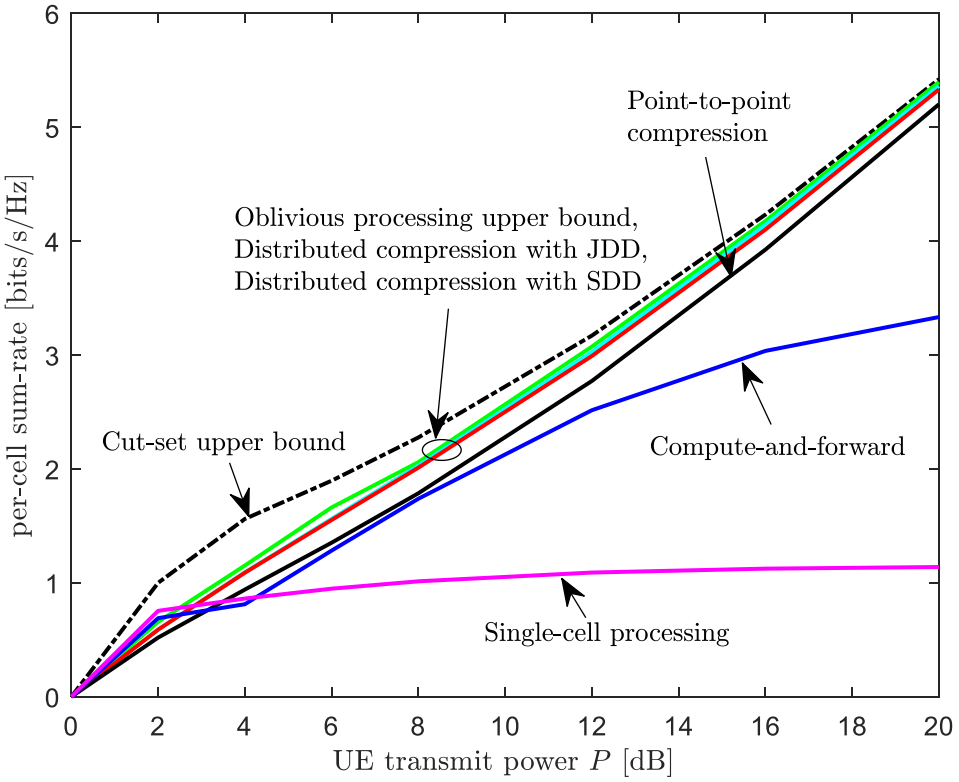
$\alpha = 1/\sqrt{2}$ and $C = 3$ bit/s/Hz



- Distributed compression
  - Joint decompression and decoding does not provide much gain compared to separate decompression and decoding.
  - This is because there is not much room to improve with oblivious processing at the RUs.

Numerical Example

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



- When C increases as log(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.

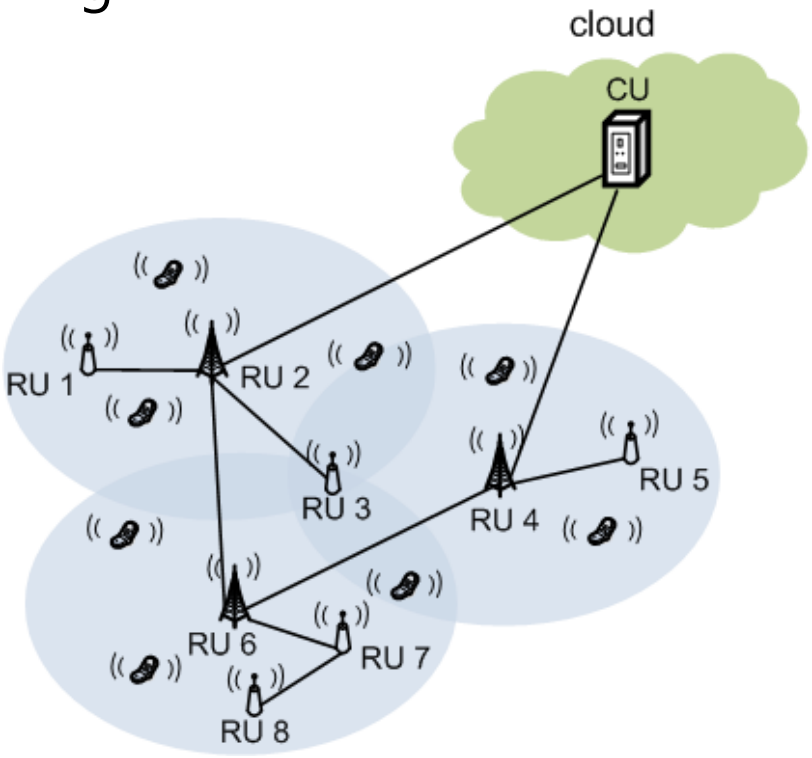
Technion
Israel Institute of Technology

Outline

I. Introduction
II. Uplink
  A. Distributed fronthaul compression
  B. Structured coding
  C. Multi-hop fronthaul topology
III. Downlink
  A. Multivariate fronthaul compression
  B. Structured coding
  C. Inter-cluster multivariate fronthaul compression
IV. Fog Radio Access Networks
  A. Edge caching and F-RAN
  B. Information-theoretic analysis
  C. Optimization of signal processing strategies
V. Outlook

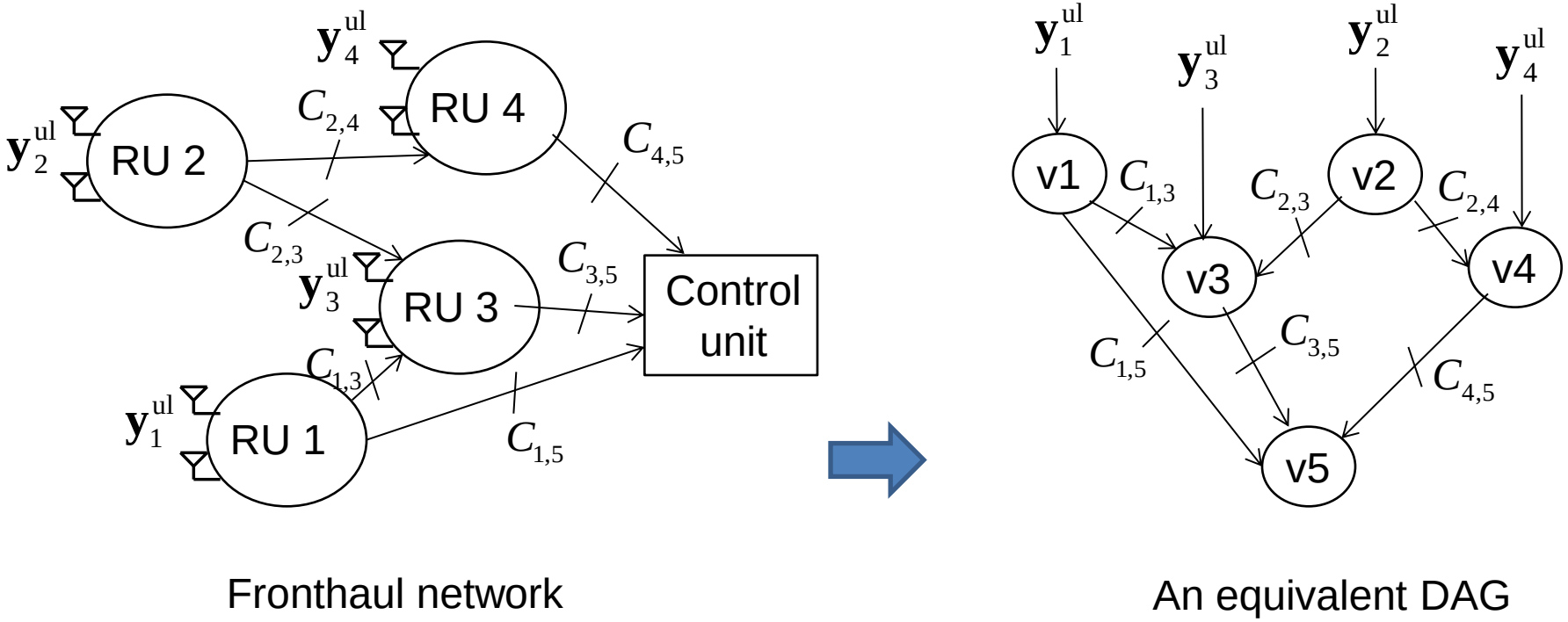
Cloud Radio Access Networks

- 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

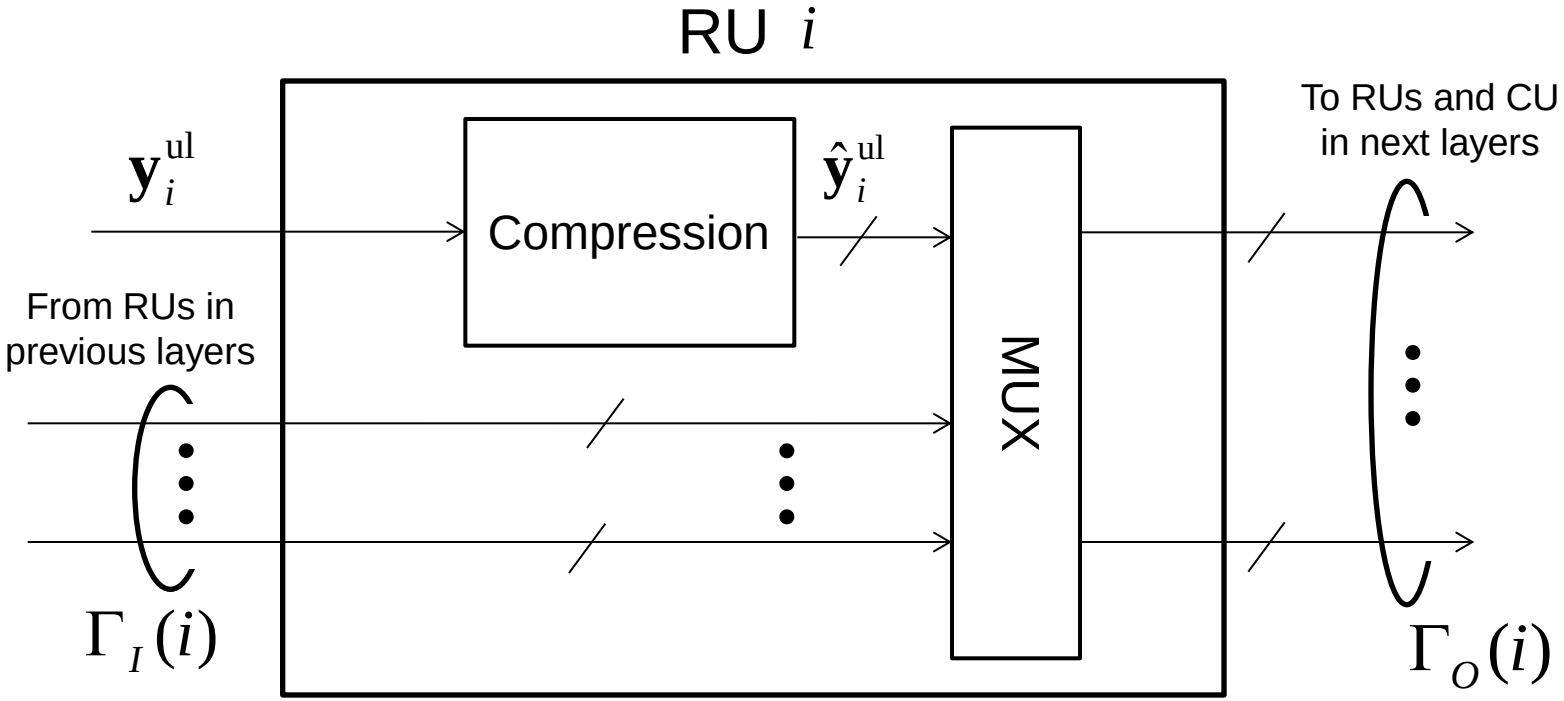
- Multihop fronthaul network modeled as a directed acyclic graph (DAG) [Koetter-Medard '03]



Fronthaul network

An equivalent DAG

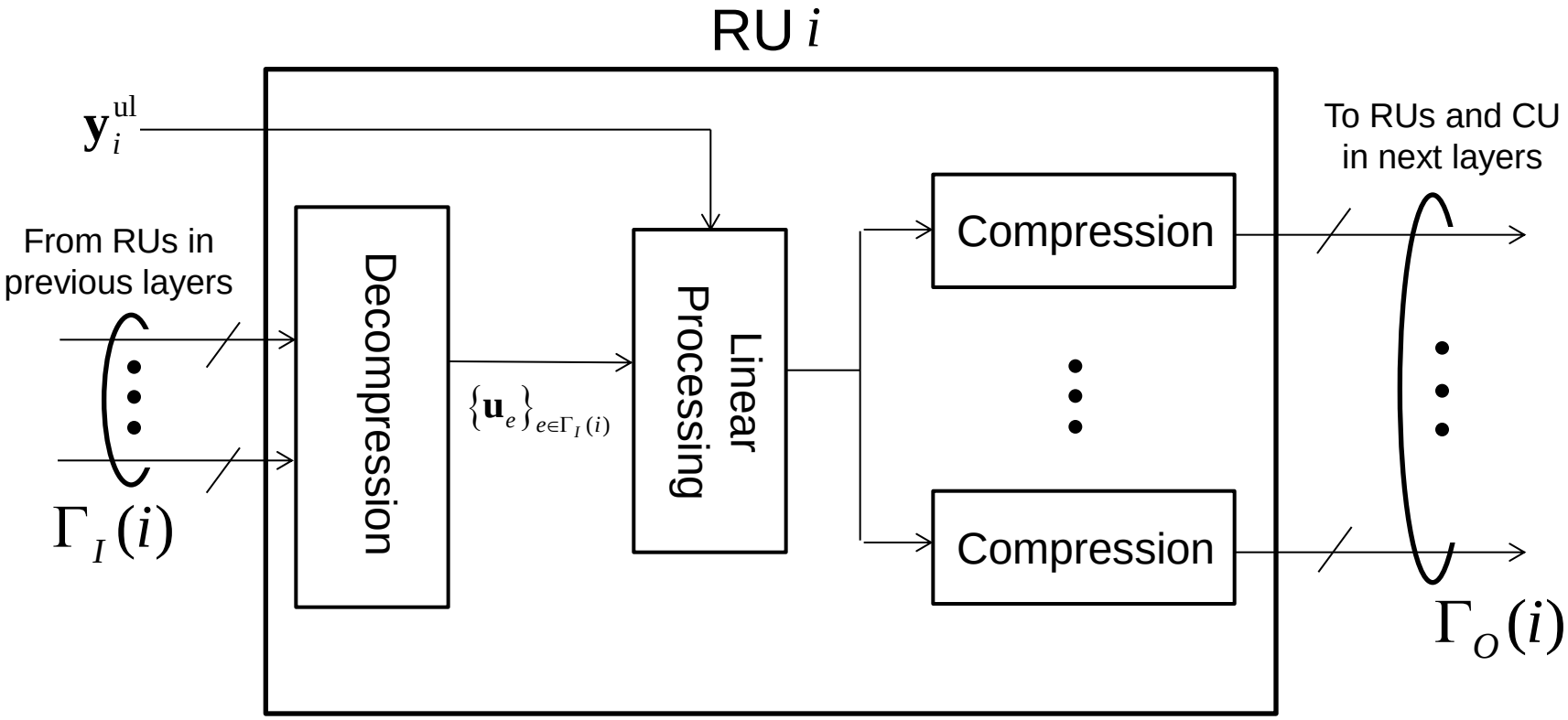
Routing



- 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]



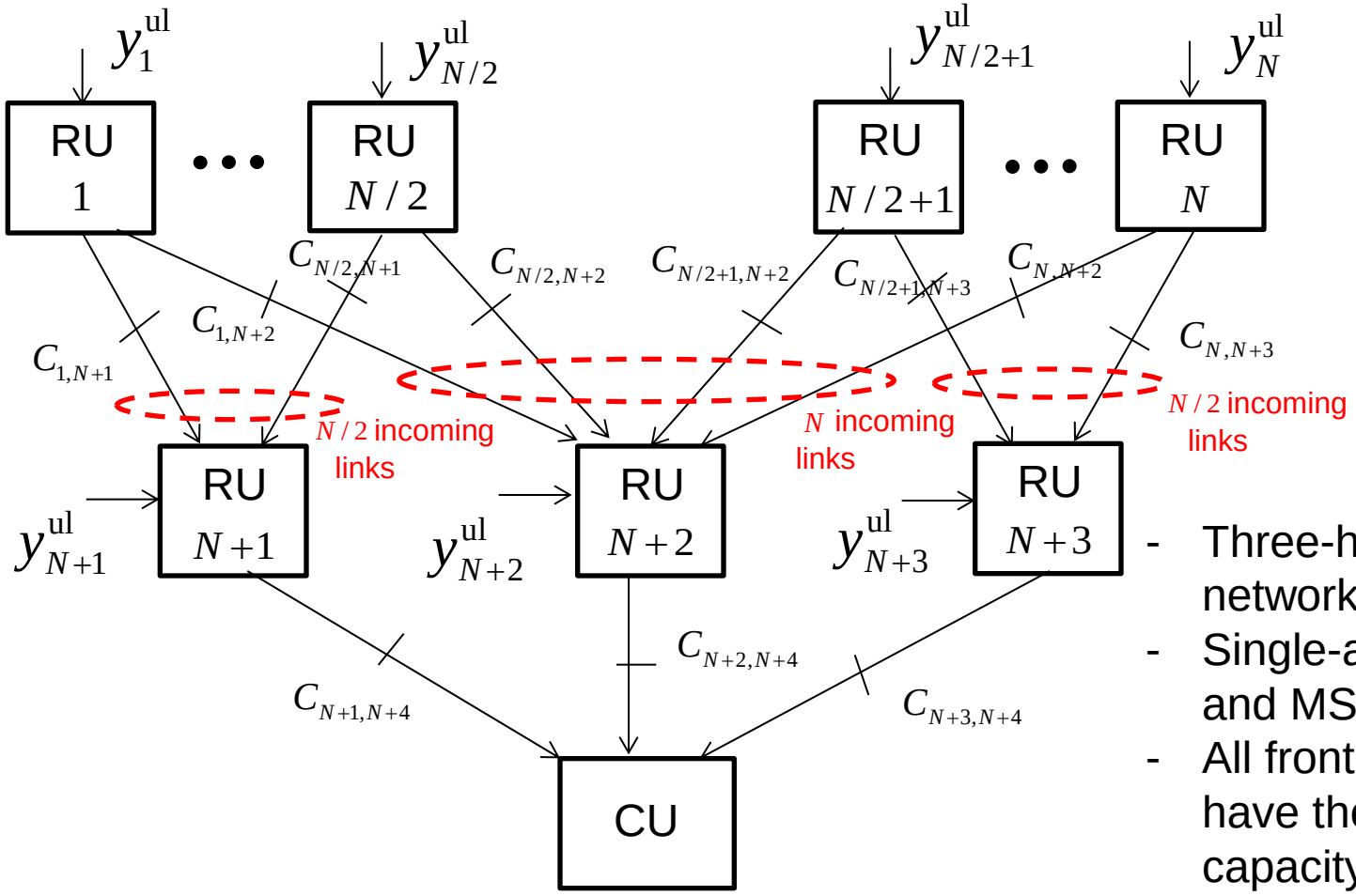
- In a dense deployment of RUs, an RU may be connected to a large number of nearby RUs, all of which receive correlated baseband signals.
- It is possible to combine the correlated baseband signals at the RU in order to reduce redundancy.

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].

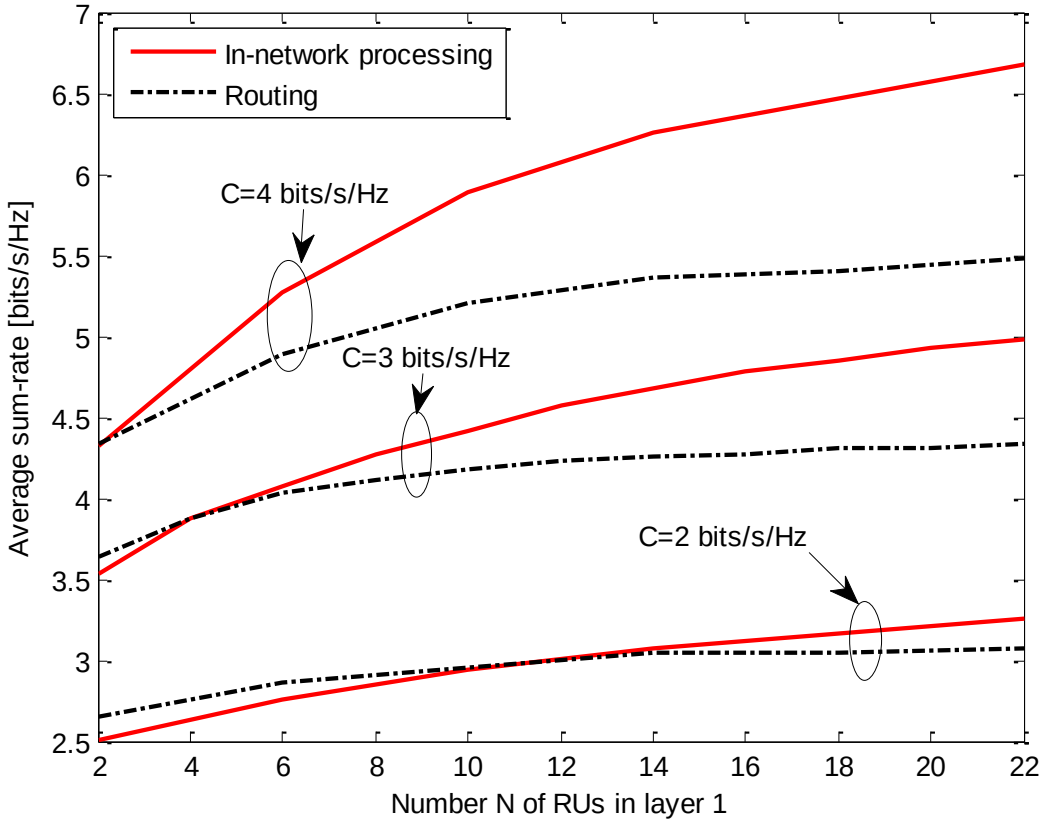
Numerical Example



- 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.

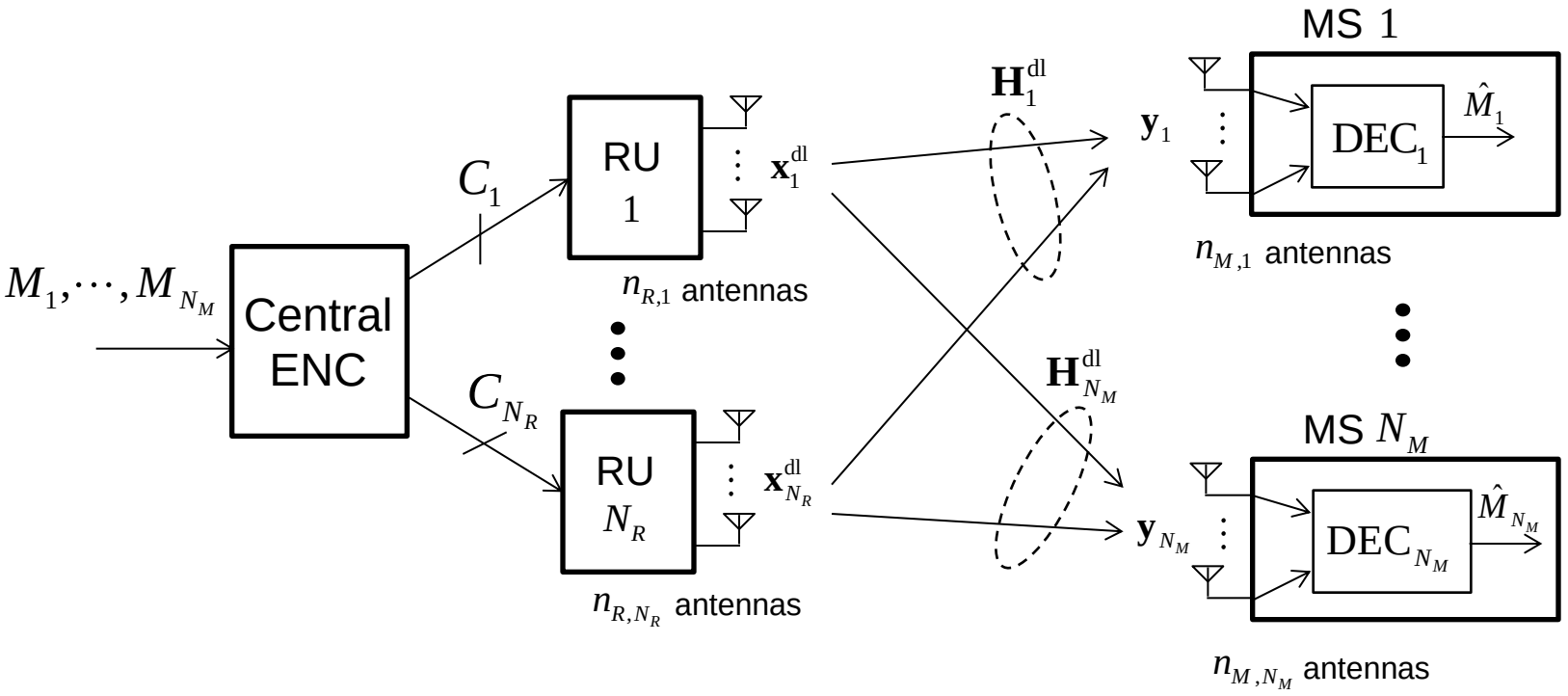
Technion
Israel Institute of Technology

Outline

I. Introduction
II. Uplink
  A. Distributed fronthaul compression
  B. Structured coding
  C. Multi-hop fronthaul topology
III. Downlink
  A. Multivariate fronthaul compression
  B. Structured coding
  C. Inter-cluster multivariate fronthaul compression
IV. Fog Radio Access Networks
  A. Edge caching and F-RAN
  B. Information-theoretic analysis
  C. Optimization of signal processing strategies
V. Outlook

Technion
Israel Institute of Technology

System Model



Technion Israel Institute of Technology

System Model

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

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

where

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

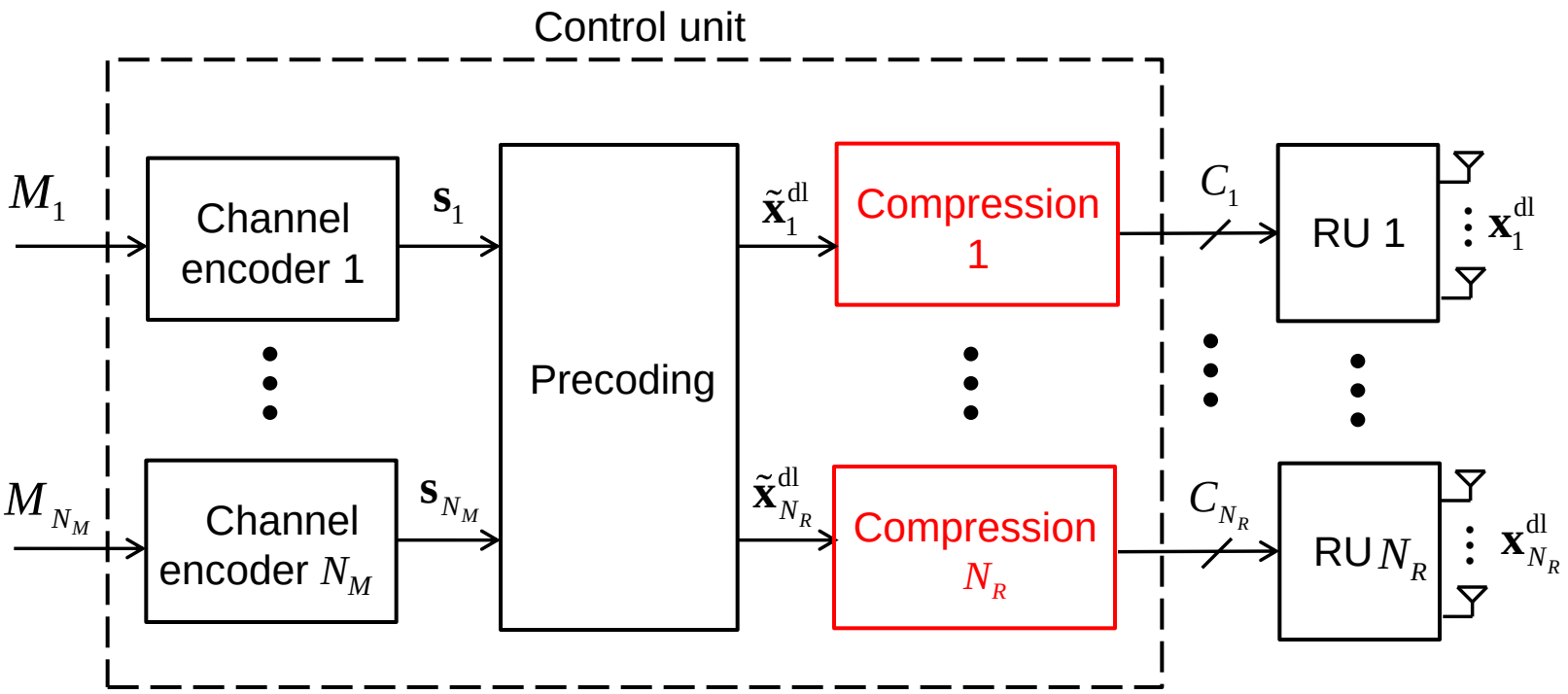
$\mathbf{z}_k^{\mathrm{dl}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{\mathbf{z}_i^{\mathrm{dl}}})$: noise and interference arising from the other clusters;

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

- Per-RU power constraint: $E\left\|\mathbf{x}_1^{\mathrm{dl}}\right\|^2 \le P_{R,i},\ i \in \{1,\ldots,N_R\}.$

Point-to-Point Compression

[Simeone et al '09]

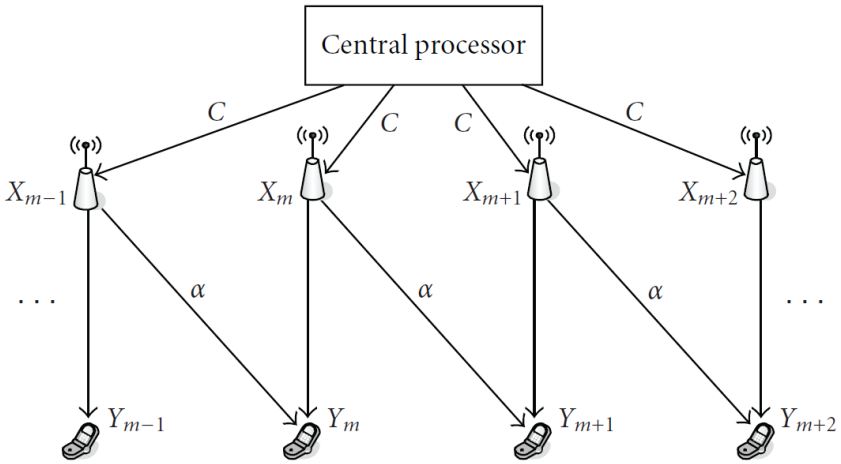


Baseband signals for different RUs are separately compressed.

- 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

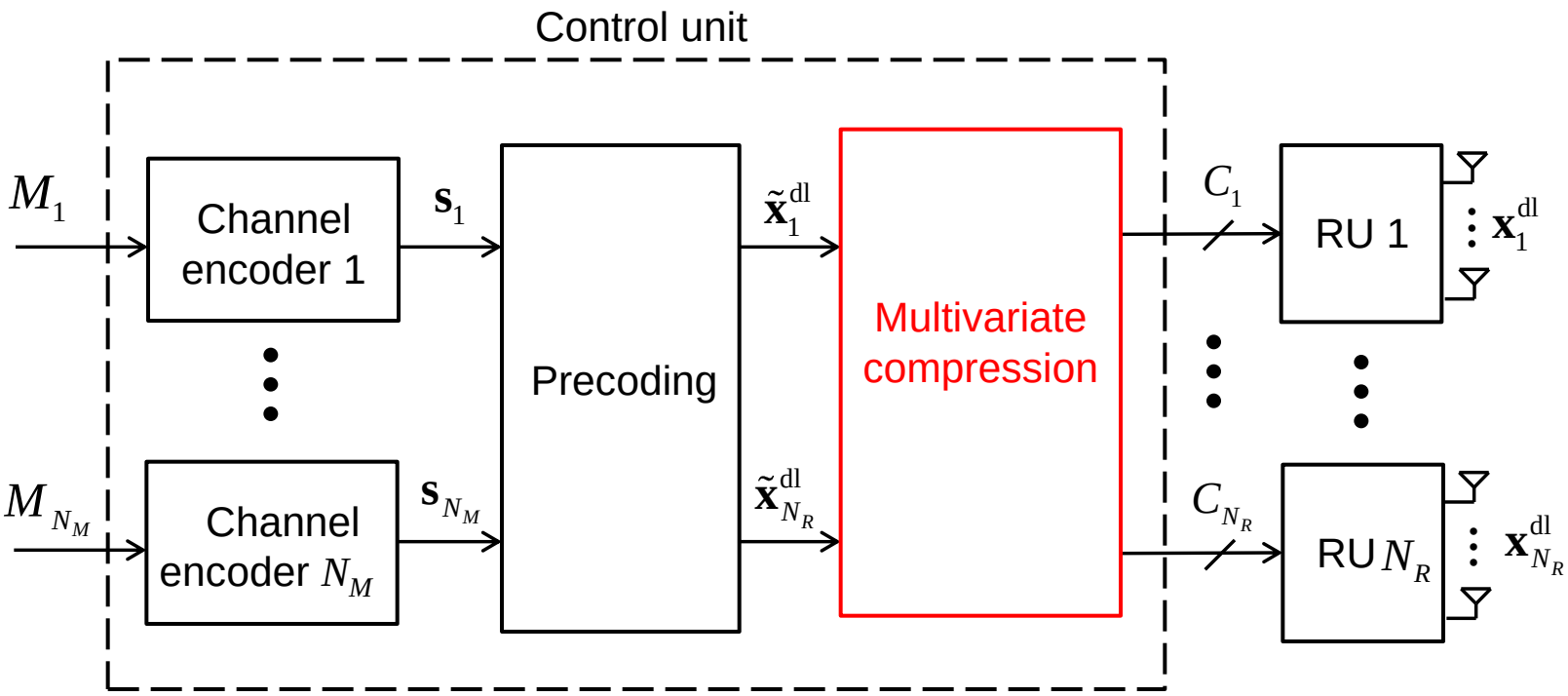
$$R_{\text{per-cell}} = \log\left(\frac{1+(1+\alpha^2)\tilde{P}+\sqrt{1+2(1+\alpha^2)\tilde{P}+(1-\alpha^2)^2\tilde{P}^2}}{2}\right)$$

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

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

Multivariate Compression

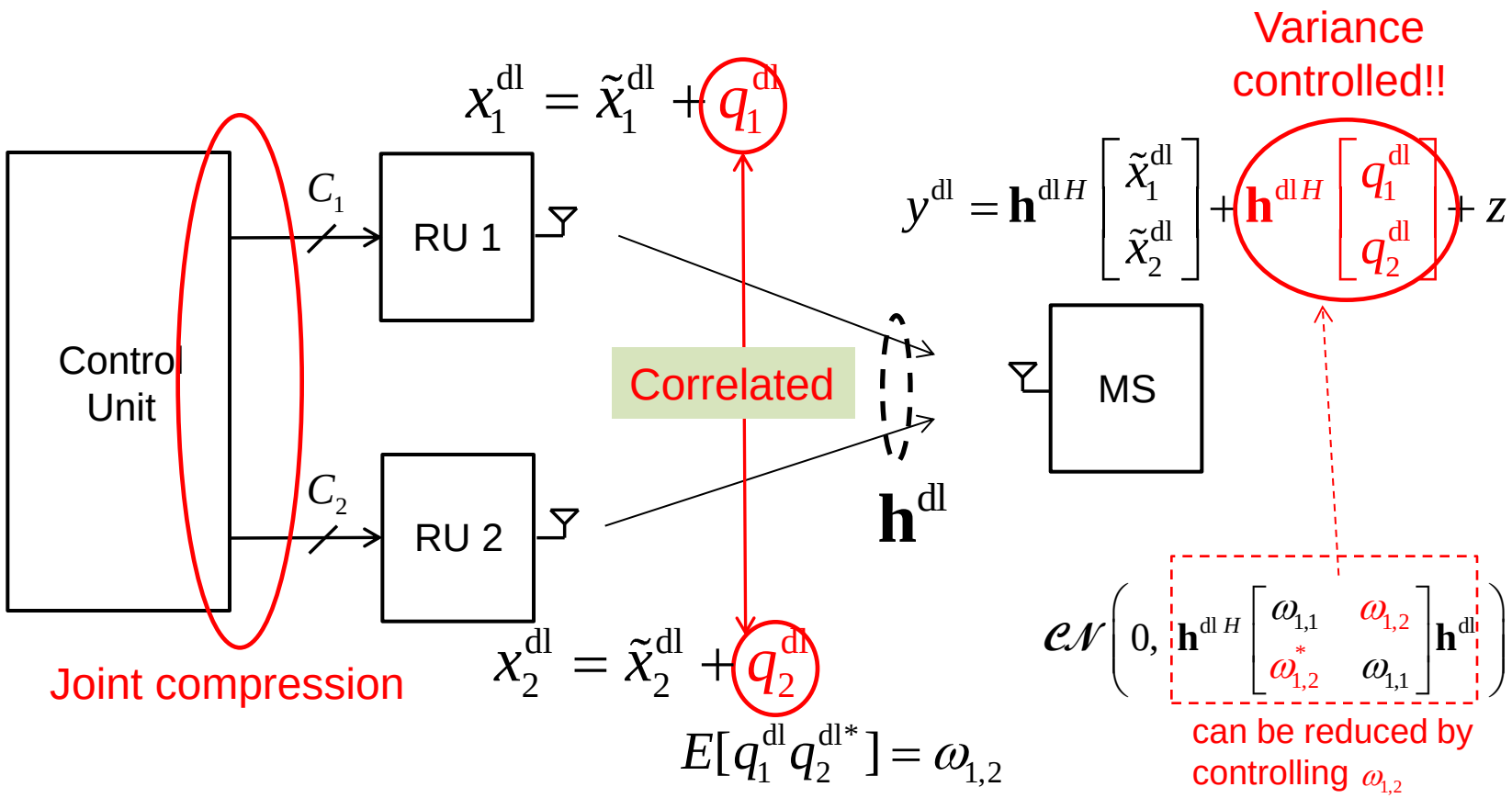
[Park et al TSP'13]



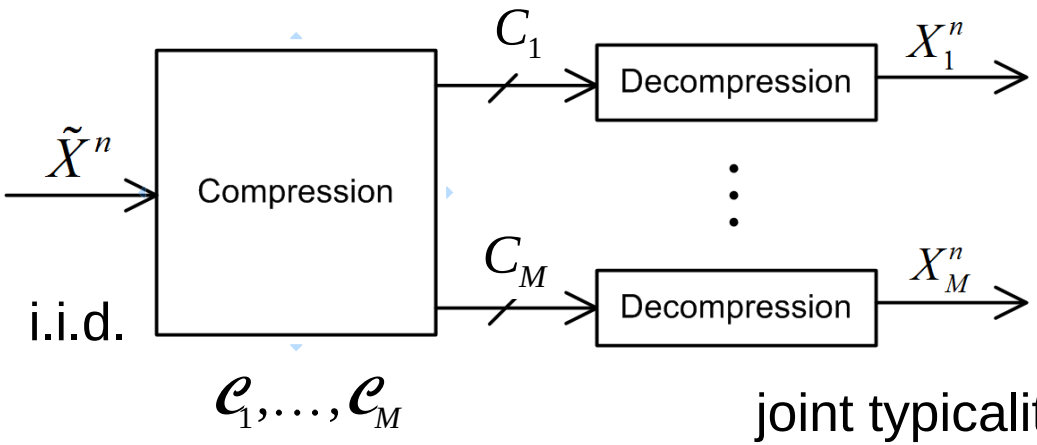
Baseband signals for different RUs are jointly compressed.

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

$$p(\tilde{x}, x_1, \ldots, x_M) = p(\tilde{x})\, p(x_1, \ldots, x_M \mid \tilde{x})$$

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

[ElGamal-Kim '11, Ch. 9]

Multivariate Compression

- Linear precoding (DPC treated in a similar way)

- Gaussian test channel:

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

- The compressed signal $\mathbf{x}^{\mathrm{dl}} = \left[\mathbf{x}_1^{\mathrm{dl}\,H}, \cdots, \mathbf{x}_{N_R}^{\mathrm{dl}\,H}\right]^H$ is given as

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

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

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

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

Technion Israel Institute of Technology

Optimization

- Weighted sum-rate maximization

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

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

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

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

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

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

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

- Difference-of-convex (DC) problem: Local optimum via MM algorithm [Beck-Teboulle '11]

Technion
Israel Institute of Technology

Outline

I. Introduction

II. Uplink

- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology

III. Downlink

- A. Multivariate fronthaul compression
- B. Structured coding
- C. Inter-cluster multivariate fronthaul compression

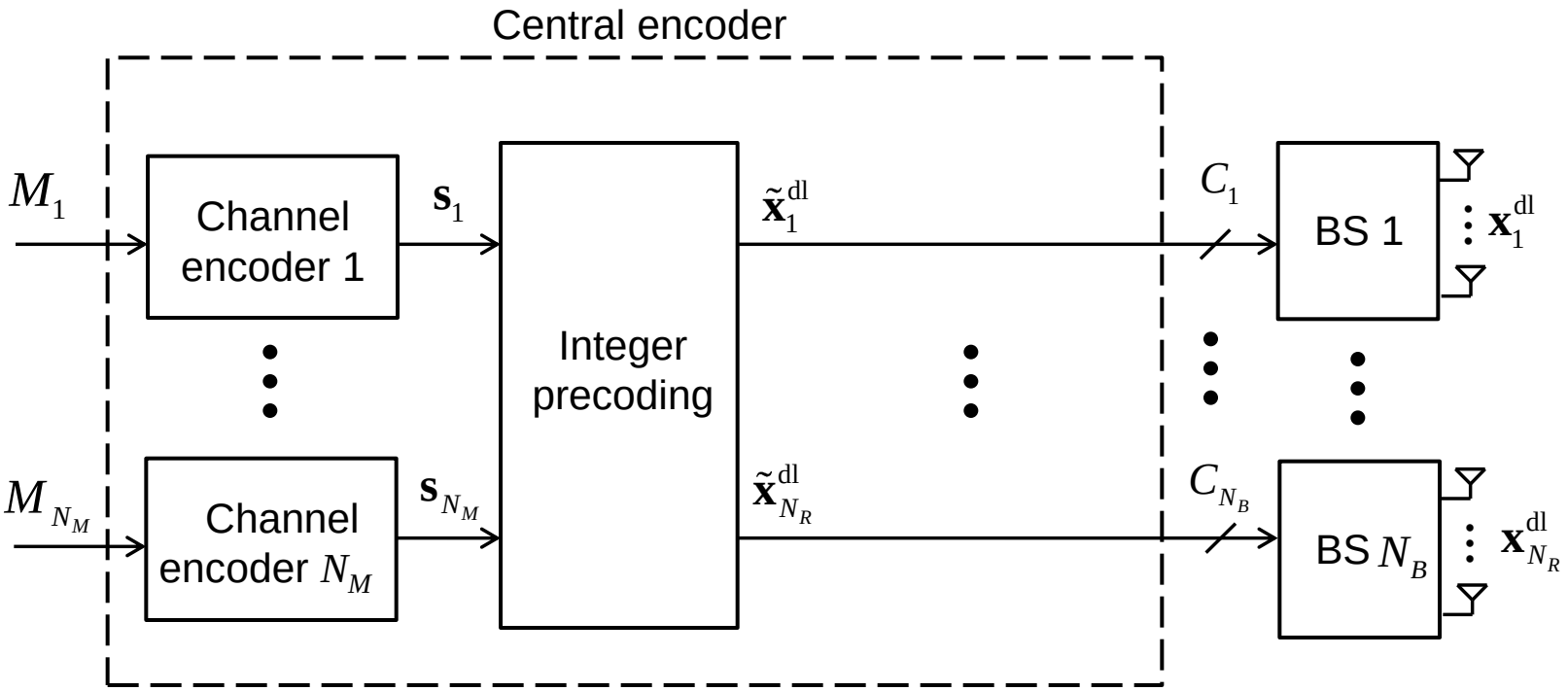
IV. Fog Radio Access Networks

- A. Edge caching and F-RAN
- B. Information-theoretic analysis
- C. Optimization of signal processing strategies

V. Outlook

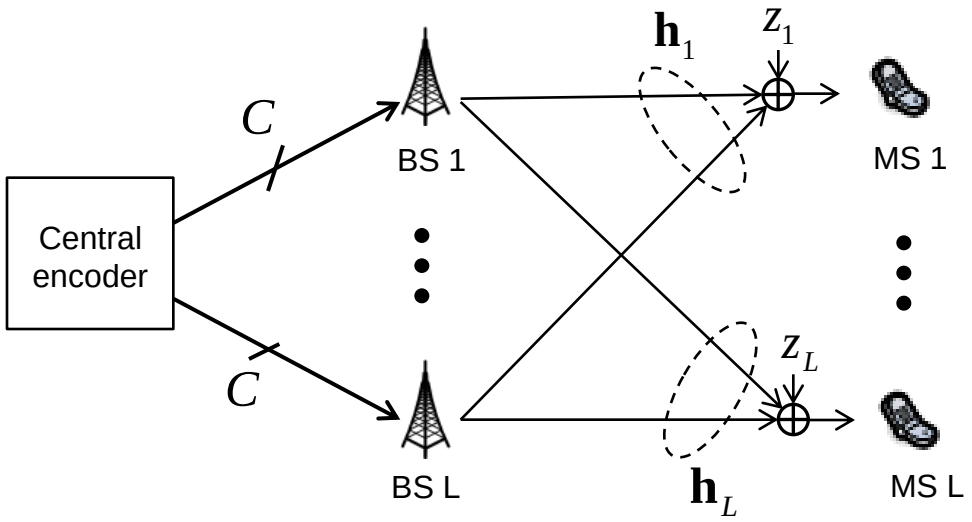
Structured Coding

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



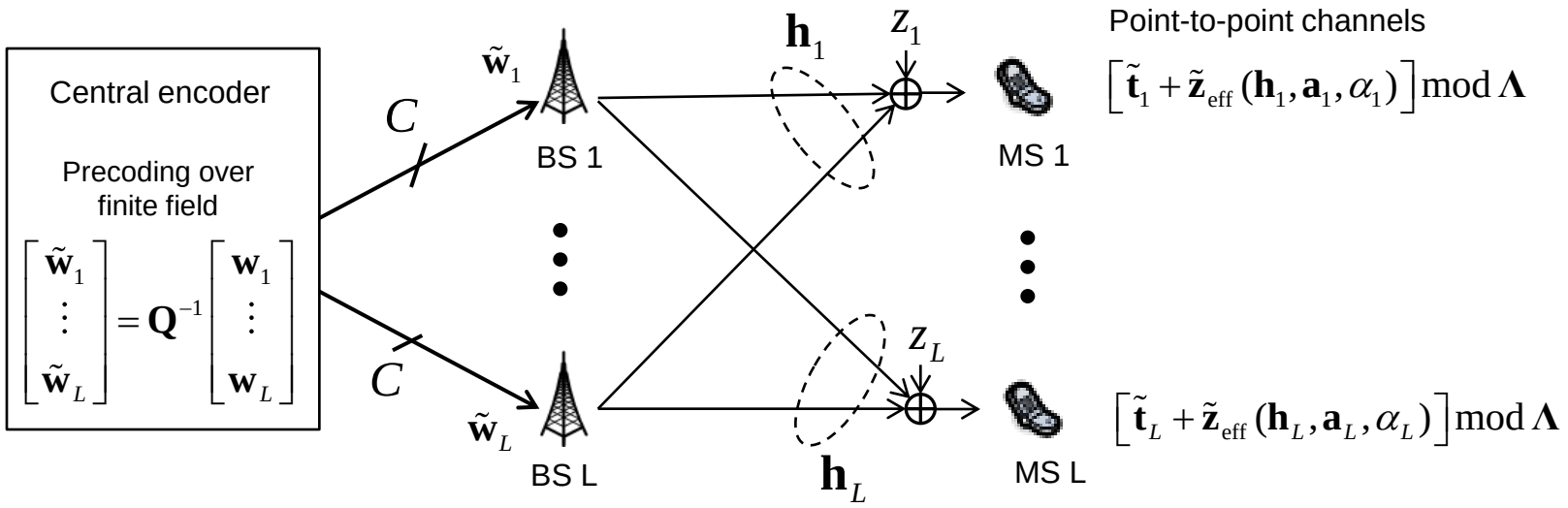
Structured Coding

- 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 $i \in \mathcal{L} = \{1, \cdots, L\}$.



Structured Coding

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



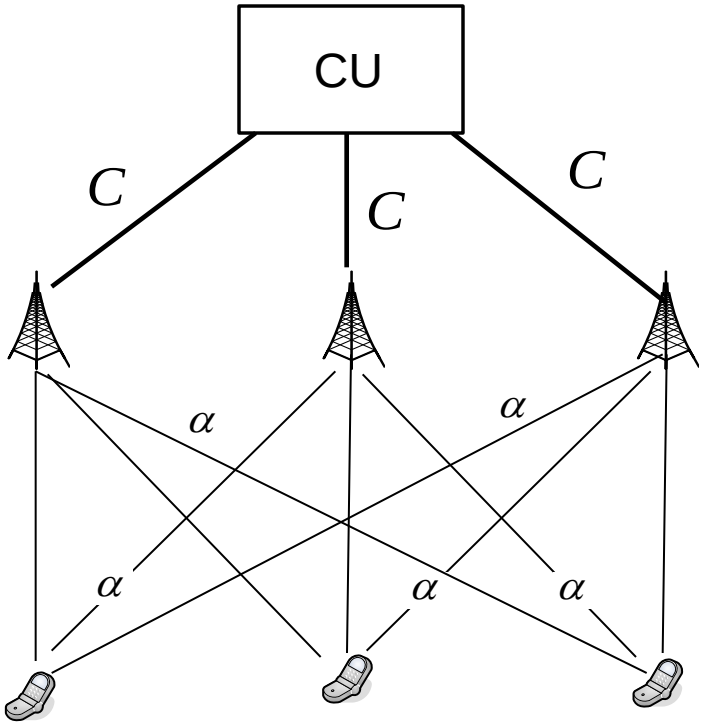
- The same lattice code is used by each BS.
- Each MS k estimates a function $\hat{\mathbf{w}}_k = \sum_{j=1}^{L} a_{k,j}\tilde{\mathbf{w}}_j$ by decoding on the lattice code.
- Achievable rate per MS is given by

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

Integer penalty

Numerical Example

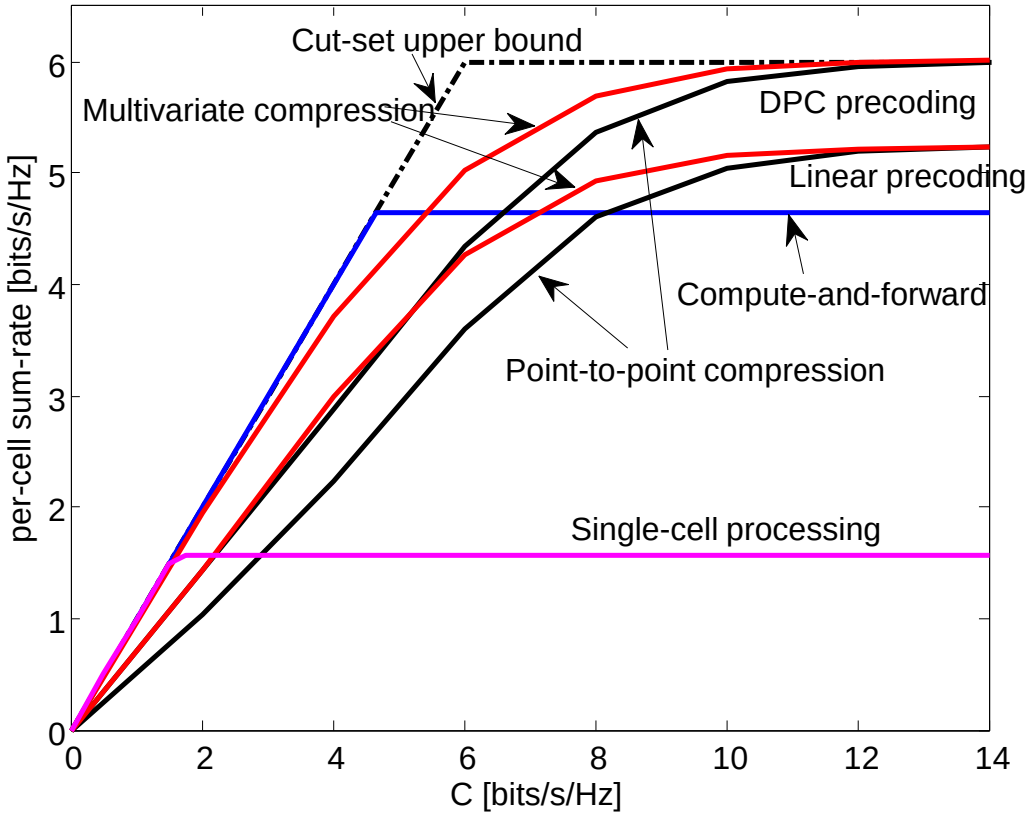
- Three-cell SISO circular Wyner model



- 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

$P = 20$ dB and α=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.

Technion
Israel Institute of Technology

Outline

I. Introduction

II. Uplink
   A. Distributed fronthaul compression
   B. Structured coding
   C. Multi-hop fronthaul topology

III. Downlink
   A. Multivariate fronthaul compression
   B. Structured coding
   C. Inter-cluster multivariate fronthaul compression

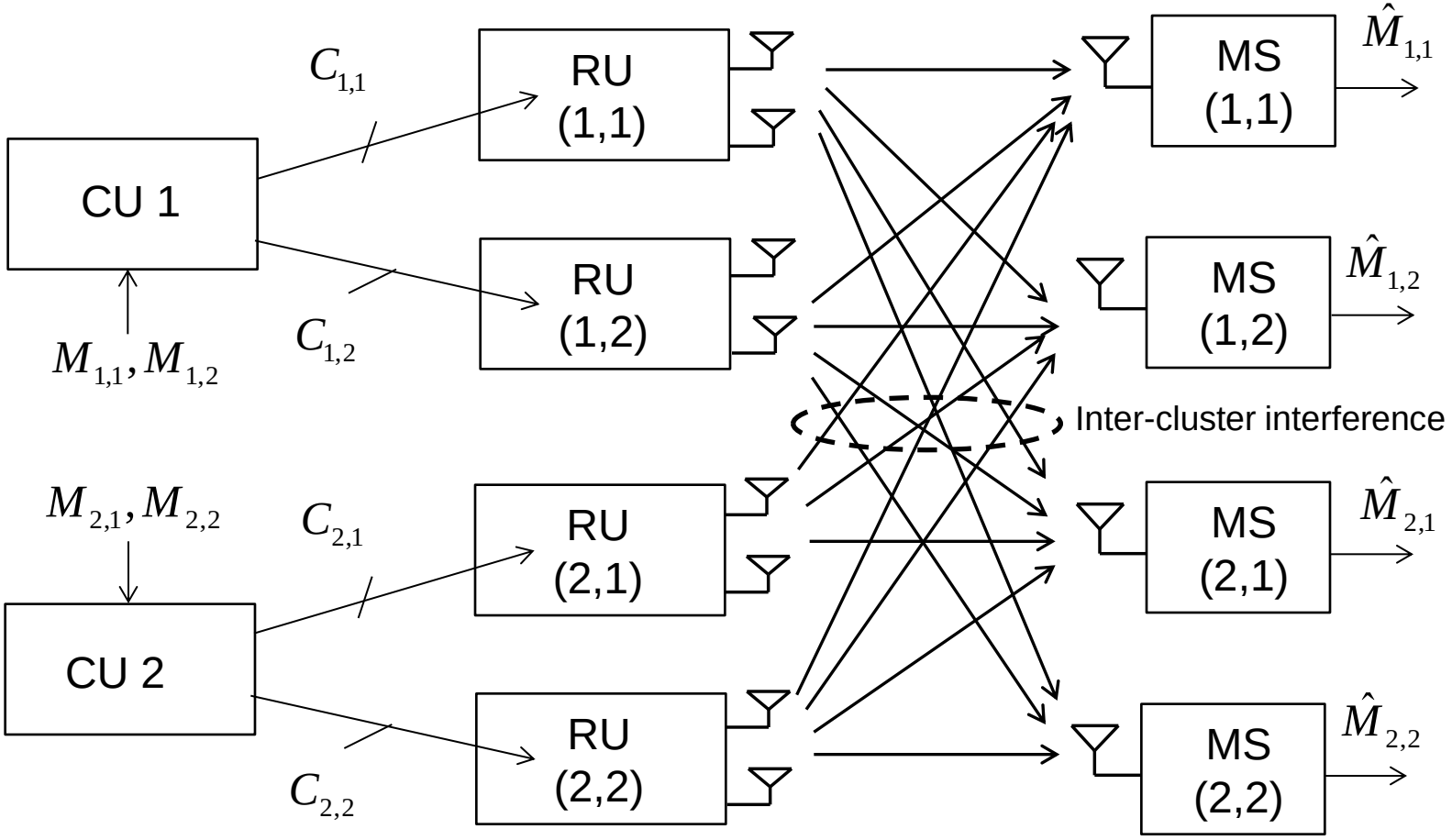
IV. Fog Radio Access Networks
   A. Edge caching and F-RAN
   B. Information-theoretic analysis
   C. Optimization of signal processing strategies

V. Outlook

Inter-Cluster Multivariate Fronthaul Design

[Park et al WCL'14]

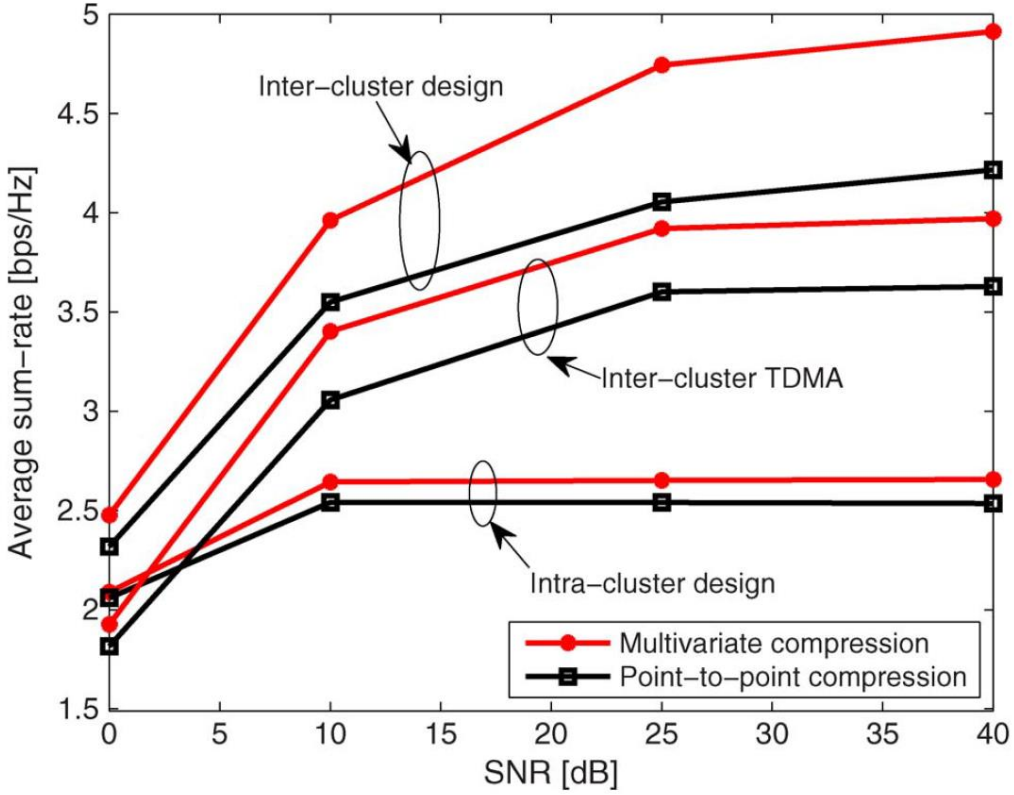
- An illustration of the downlink of multi-cluster cloud radio access network



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.

Technion
Israel Institute of Technology

Outline

I. Introduction

II. Uplink
- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology

III. Downlink
- A. Multivariate fronthaul compression
- B. Structured coding
- C. Inter-cluster multivariate fronthaul compression

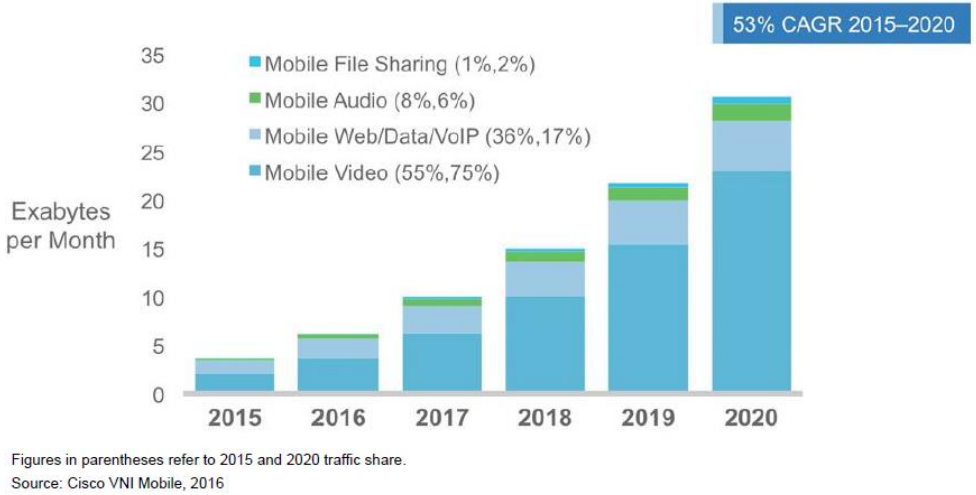
IV. Fog Radio Access Networks
- A. Edge caching and F-RAN
- B. Information-theoretic analysis
- C. Optimization of signal processing strategies

V. Outlook

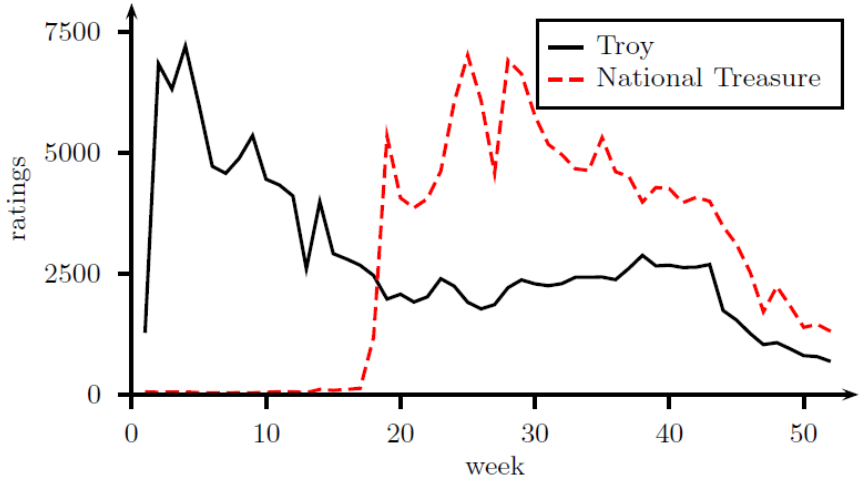
Technion
Israel Institute of Technology

Edge Caching

- Video-on-demand is driving wireless traffic growth.
- Predicted to account for almost ¾ 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

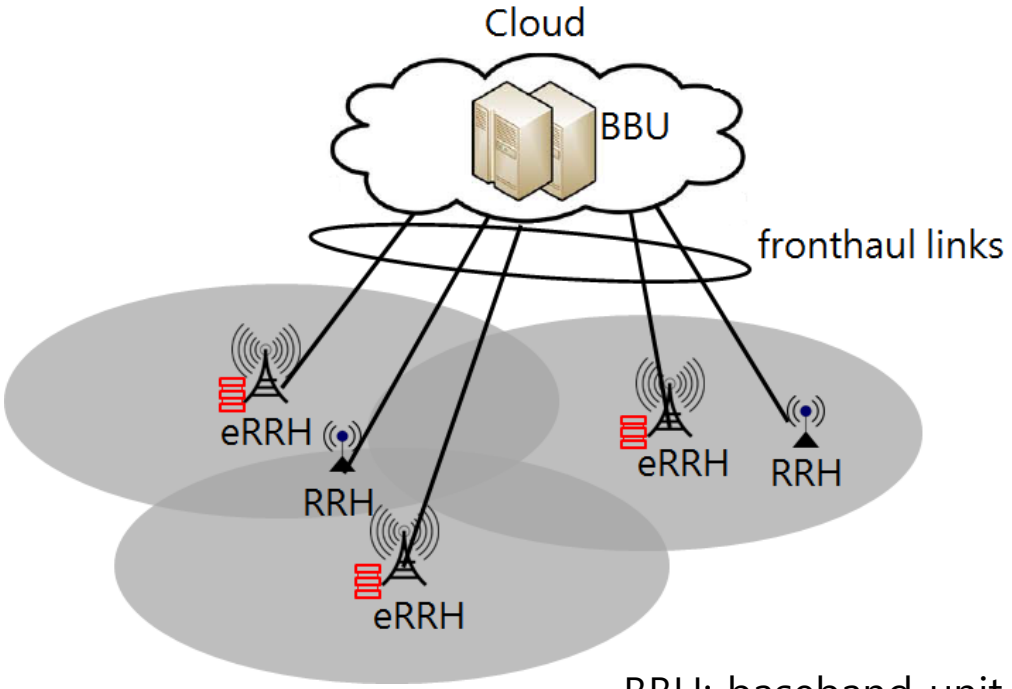


Technion Israel Institute of Technology

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].

- Advantages over C-RAN
  - Centralized interference management
  - Reduced fronthaul overhead
  - Low latency
  - Higher spectral efficiency



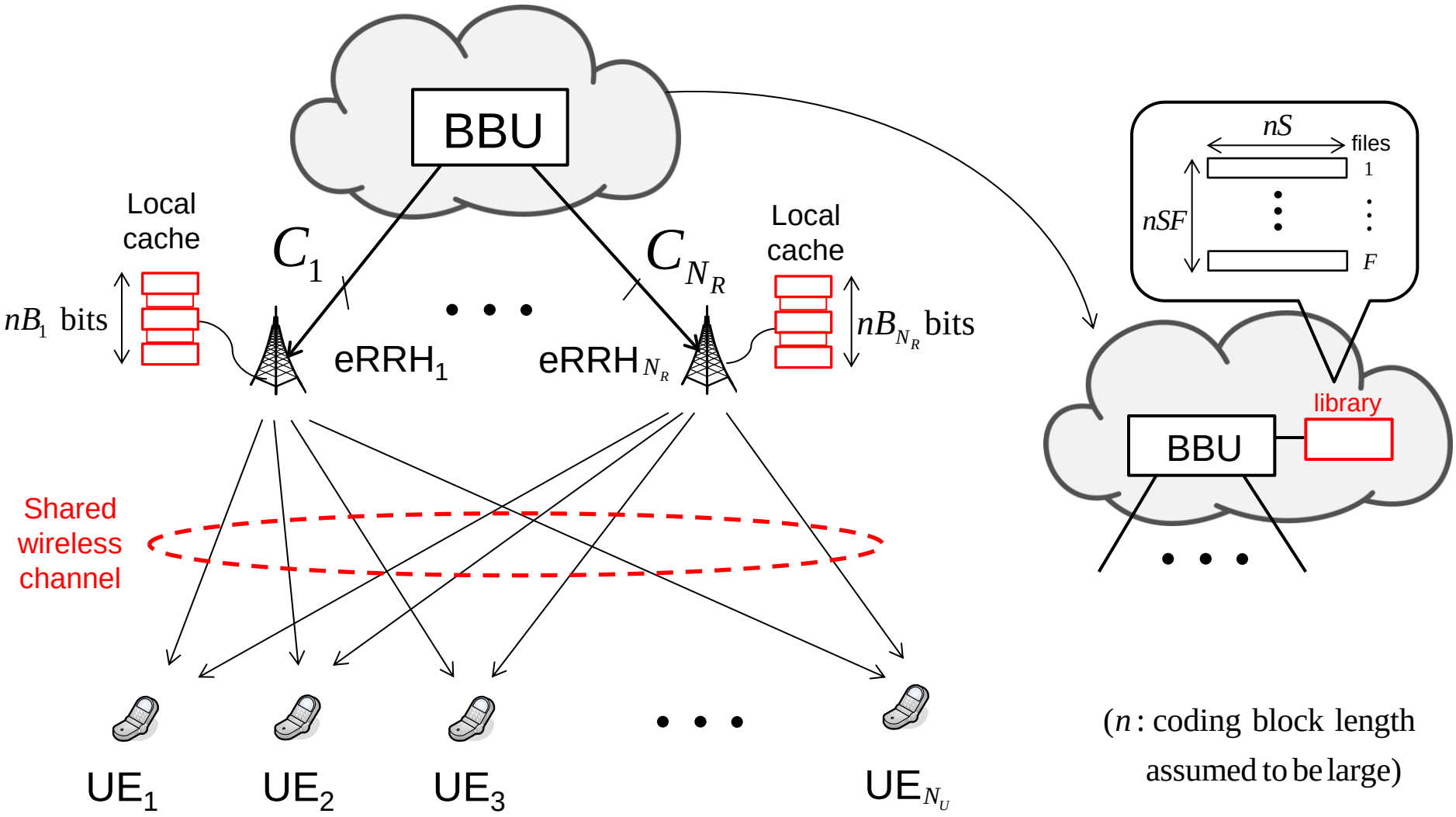
BBU: baseband unit

eRRH: enhanced RRH

: cache

Technion
Israel Institute of Technology

System Model



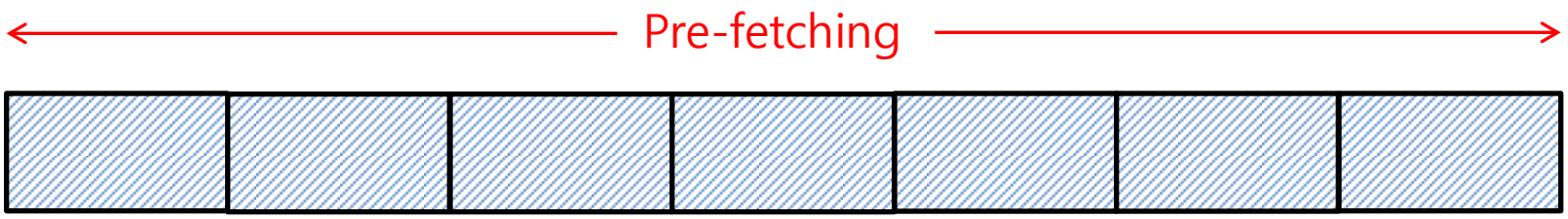
(n: coding block length assumed to be large)

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

Technion
Israel Institute of Technology

Design Space

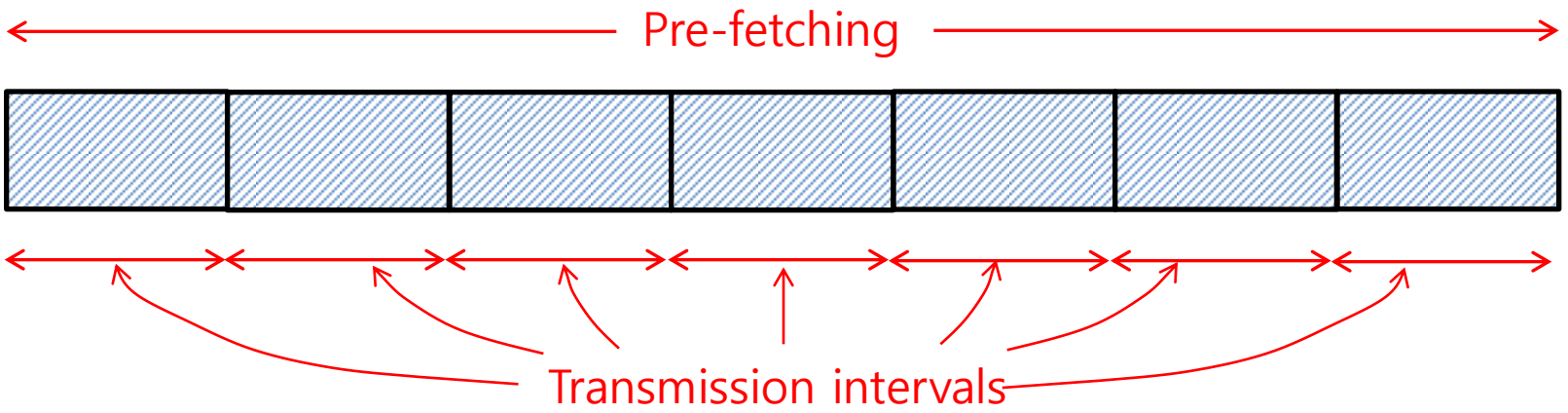
- **Phase I.** Pre-fetching phase



  - Pre-fetching strategy: What to cache

Design Space

- **Phase II.** Delivery phase



  - Pre-fetching strategy: What to cache
  - Fronthauling mode: What to transmit on the fronthaul links
    - Hard-/Soft-transfer mode
  - Edge transmission strategy: How to transmit on the wireless channel

State-of-the-Art

- Some pioneering works
  - Receiver-end caching [MAli-Niesen '14][MAli-Niesen '15]
  - Edge-caching (a.k.a. femto-caching) [Golrezaei et al '13]
- Information-theoretic analysis
  - DoF analysis of cache-aided IA [Naderializadeh et al '16]
  - Latency tradeoffs in cache-aided wireless networks [Sengupta et al '16]
- Pre-fetching phase design
  - Coded caching [Ugur et al '15]
  - Fronthaul-aware caching [XPeng et al '15], mobility-aware caching [Wang et al '16]
  - Stochastic geometry-based analysis of hybrid caching [Chen et al '16]
- Delivery phase design
  - Joint design of beamforming and eRRH clustering [Tao et al '16]
  - Optimization under hard/soft-transfer fronthauling modes [Park-Simeone-Shamai arXiv'16][Park-Simeone-Shamai SPAWC'16]

Technion
Israel Institute of Technology

Outline

I. Introduction

II. Uplink

- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology

III. Downlink

- A. Multivariate fronthaul compression
- B. Structured coding
- C. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks

- A. Edge caching and F-RAN
- B. Information-theoretic analysis
- C. Optimization of signal processing strategies

V. Outlook

Cache-aided Interference Alignment

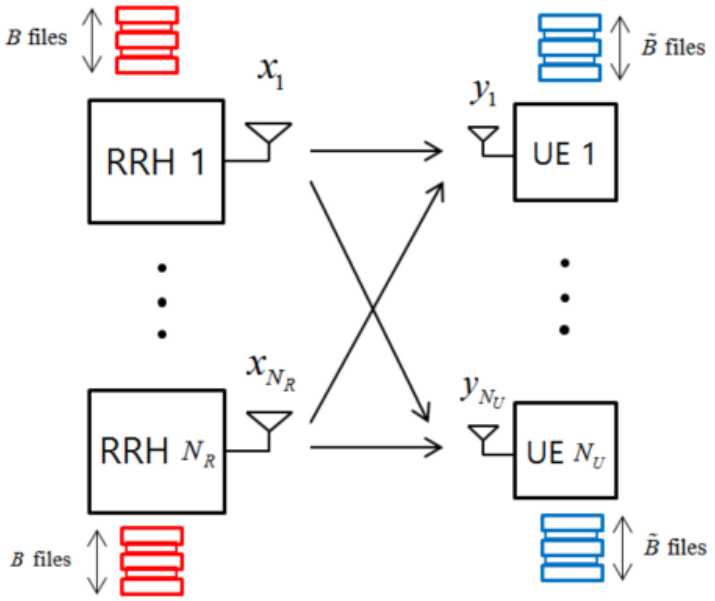
[Naderializadeh et al '16]

- DoF analysis under the assumption of
  - Single-antenna at each enhanced remote radio head (eRRH) and user equipment (UE), i.e., $n_{R,1} = \ldots = n_{R,N_R} = n_{U,1} = \ldots = n_{U,N_U} = 1$
  - No fronthaul connection between baseband unit (BBU) and eRRHs, i.e., $C_1 = \ldots = C_{N_R} = 0$
  - In addition to edge caching, *on-device caching* at UEs is also considered.
    - Each UE k can pre-fetch $\tilde{B}$ files.

- Lower/upper bounds on sum-DoF

$$\min\left\{\frac{N_R B + N_U \tilde{B}}{F},\ N_U\right\} \le \text{DoF} \le \min\left\{2\frac{N_R B + N_U \tilde{B}}{F},\ N_U\right\}$$

  - Within a factor of 2 characterization
  - Achievable scheme
    - Collaborative zero-forcing (ZF) at eRRHs
    - Known interference cancellation at UEs
  - Equal contribution of edge and on-device caching



: edge caching

: on-device caching

Latency Tradeoffs

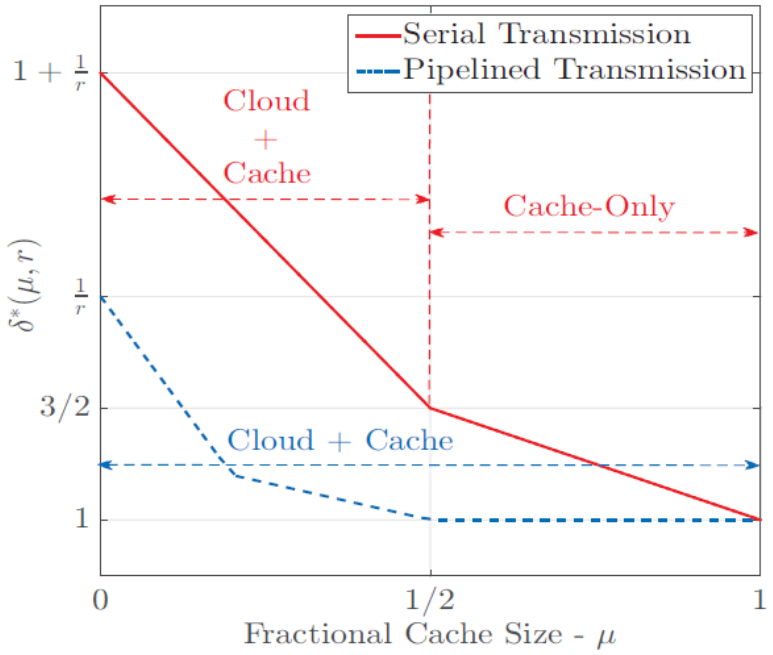
[Sengupta et al '16]

- Normalized Delivery Time (NDT) analysis under
  - Single-antenna at each eRRH and UE, i.e., $n_{R,1} = \ldots = n_{R,N_R} = n_{U,1} = \ldots = n_{U,N_U} = 1$
  - Fronthaul capacity scaling with SNR P, i.e., $C_1 = \ldots = C_{N_R} = r \log P$

where NDT is defined as $\dfrac{\text{Delivery time}}{\text{Ideal Delivery time (with full caching)}} \geq 1$

- Two approaches are considered for front-wireless transmission
  - Serial fronthaul-wireless transmission
  - Pipelined fronthaul-wireless transmission

- An example of $N_R = N_U = 2,\ r = 0.5$
  - Serial fronthaul/wireless
    - For small fronthaul capacity and large caching capacity, cloud processing at BBU is not useful.
  - Parallel fronthaul/wireless
    - Cloud processing is always useful.



Technion
Israel Institute of Technology

Outline

I. Introduction

II. Uplink

- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology

III. Downlink

- A. Multivariate fronthaul compression
- B. Structured coding
- C. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks

- A. Edge caching and F-RAN
- B. Information-theoretic analysis
- C. Optimization of signal processing strategies

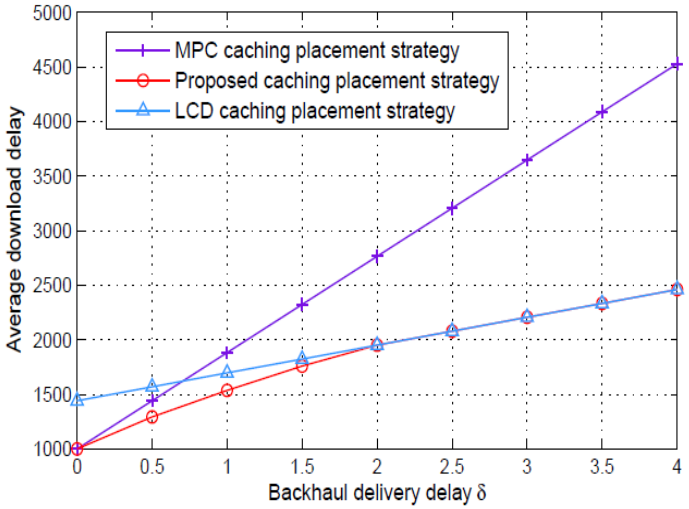
V. Outlook

Design of Pre-Fetching Policy

[XPeng et al '15][Wang et al '16]

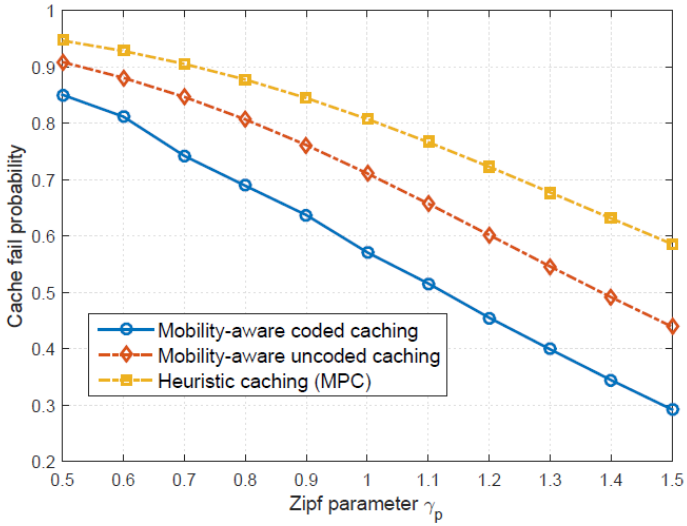
- Fronthaul-aware pre-fetching [XPeng et al '15]
  - Design goal
    - Minimization of *average download delay*
  - Numerical results
    - Proposed caching outperforms conventional pre-fetching policies

      (MPC: Most Popular Cache, LCD: Largest Cache Diversity).



[XPeng et al '15]

- Mobility-aware pre-fetching [Wang et al '16]
  - Design goal
    - Minimization of *cache failure probability*
    - Under coded/uncoded caching
  - Numerical results
    - Mobility-aware design can significantly reduce the cache failure probability as compared to conventional approach.



[Wang et al '16]

Design of Delivery Phase

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

- Joint design of cooperative beamforming and eRRH clustering [Tao et al '16]
  - With the goal of minimizing compound cost function:

$$C_{\text{total}} = C_{\text{fronthaul}} + \eta C_{\text{power}}$$

where $C_{\text{fronthaul}}$: Fronthaul cost;

C_{power}: eRRH transmission power;

η: Balancing factor.

  - Optimization process
    - **Step 1.** Smoothened l0-norm approximation and Semi-definite relaxation (SDR)
    - **Step 2.** Concave-convex procedure (CCCP) based algorithm

Design of Delivery Phase

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

- Joint design of cloud and edge processing
  - First studied in [Park-Simeone-Shamai arXiv'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 arXiv'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 arXiv'16][Park-Simeone-Shamai SPAWC'16]

- Joint design of cloud and edge processing

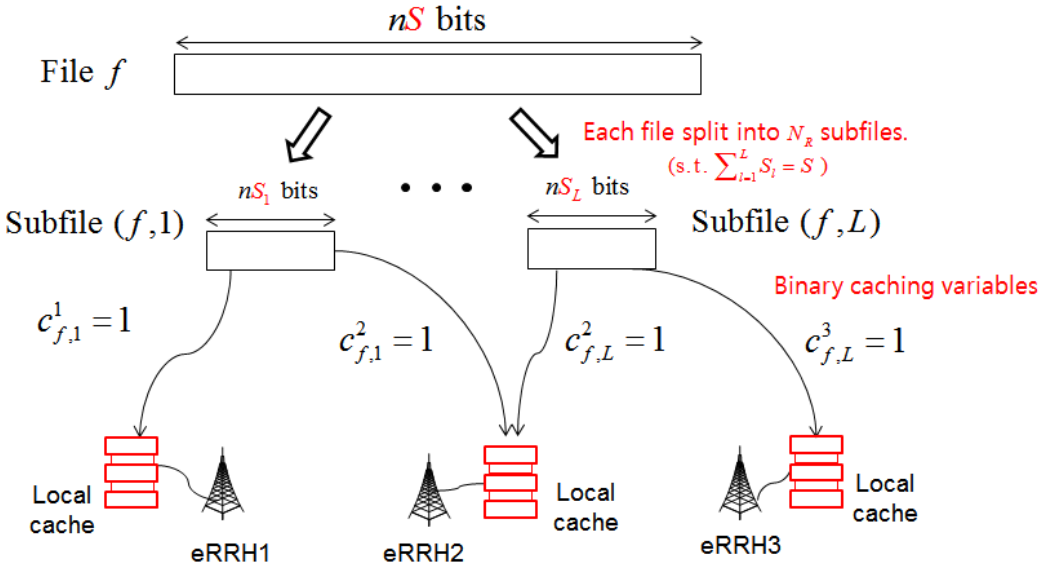
  [Park-Simeone-Shamai arXiv'16] [Park-Simeone-Shamai SPAWC'16]

  - Arbitrary fixed pre-fetching strategies are assumed.
  - Modeling of pre-fetching strategy

$$c_{f,l}^{i} = \begin{cases} 1, & \text{subfile } (f,l) \text{ cached by eRRH } i \\ 0, & \text{otherwise} \end{cases}$$

Caching capacity constraint:

$$\sum_{f \in \mathcal{F}} \sum_{l \in \mathcal{L}} c_{f,l}^{i} S_l \le B_i.$$



Design of Delivery Phase

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

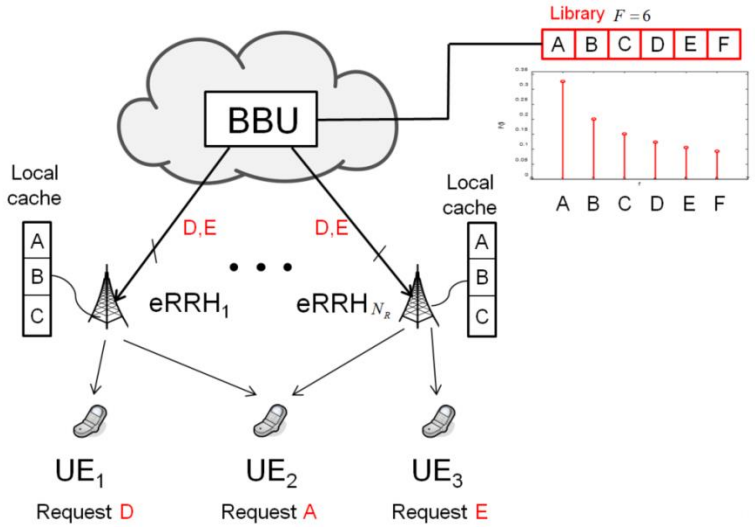
- Joint design of cloud and edge processing

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

  – Two different fronthauling modes are considered:

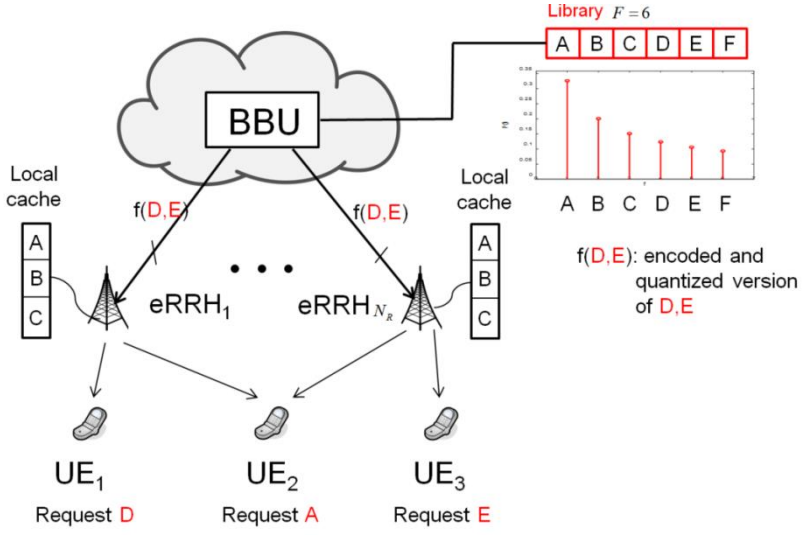
i) Hard-transfer fronthauling

– 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.



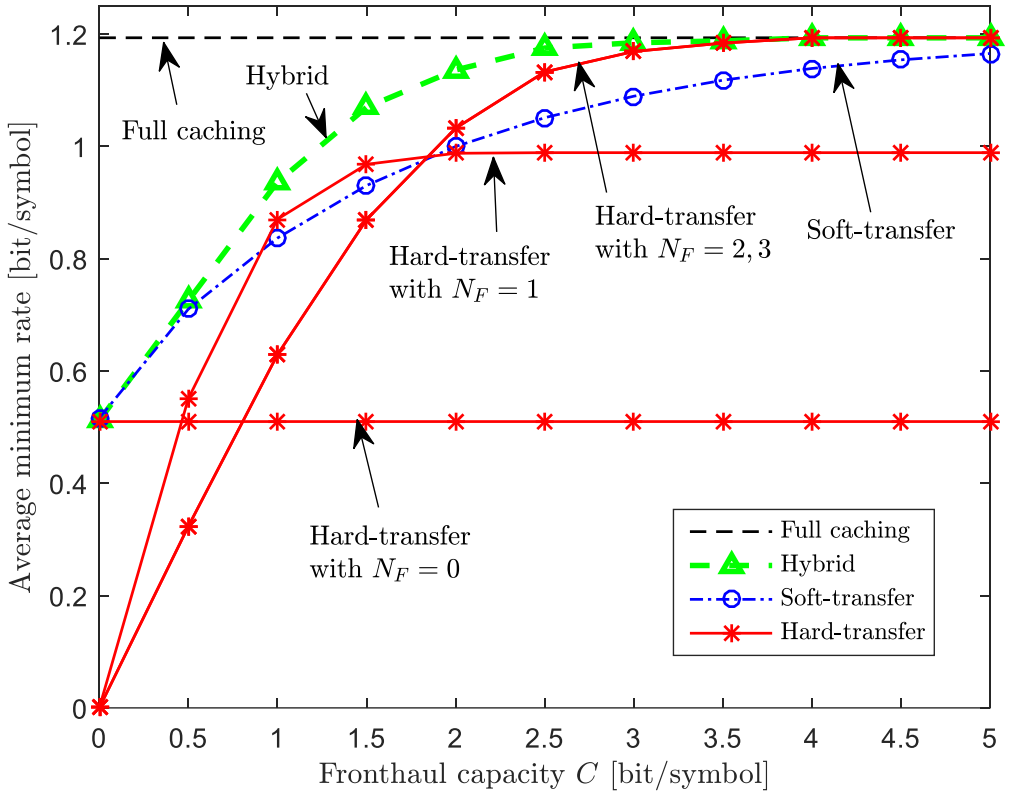
Design of Delivery Phase

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

- Joint design of cloud and edge processing

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

– Numerical results (delivery rate maximization) [Park-Simeone-Shamai arXiv'16]



(N_F: Maximum cluster size for each file)

- A smaller caching capacity can be compensated by a larger fronthaul capacity C.
- For small C, it is desirable to keep the cluster size of hard-transfer mode small.
- Soft-transfer mode outperforms the hard-transfer mode when the fronthaul capacity C is small.
- Hybrid fronthauling scheme shows relative gains as compared to the hard/soft-transfer schemes.

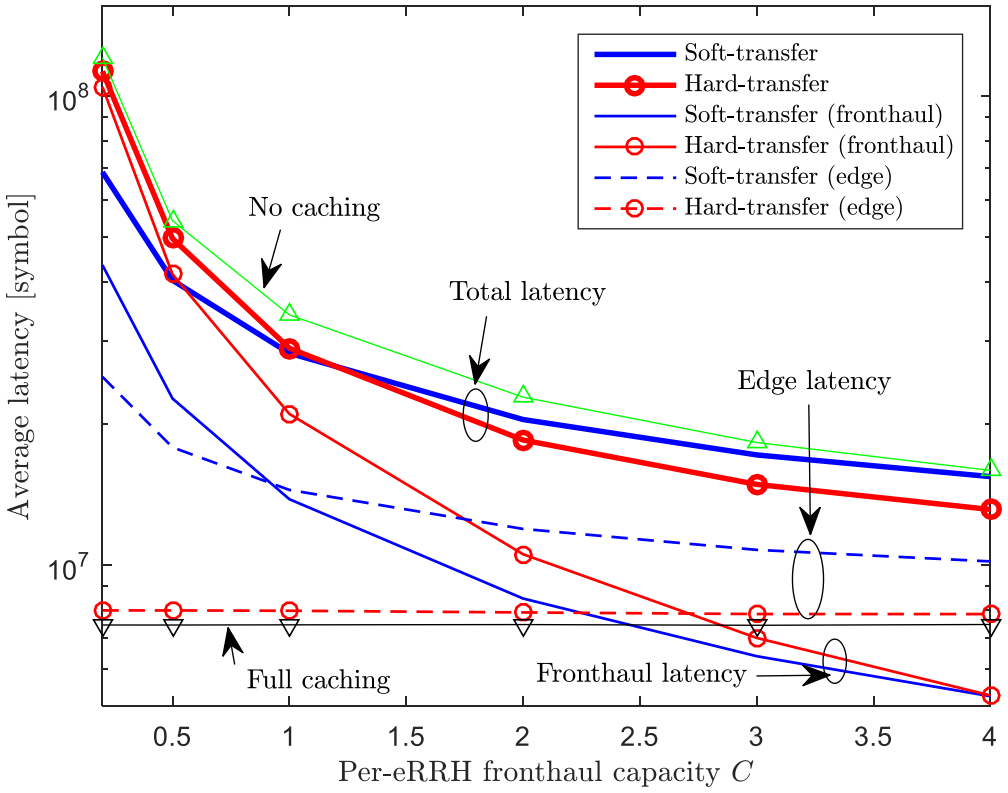
Design of Delivery Phase

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

- Joint design of cloud and edge processing

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

  – Numerical results (delivery latency minimization) [Park-Simeone-Shamai SPAWC'16]



- Again, the soft-transfer mode outperforms the hard-transfer mode when C is small.
- As C decreases, the contribution to the latency due to fronthaul transmission grows more rapidly than the edge latency.
  - Soft-transfer mode seems more efficient in the use of fronthaul resources
    - by means of baseband compression

Technion
Israel Institute of Technology

Outline

I. Introduction
II. Uplink
  A. Distributed fronthaul compression
  B. Structured coding
  C. Multi-hop fronthaul topology
III. Downlink
  A. Multivariate fronthaul compression
  B. Structured coding
  C. Inter-cluster multivariate fronthaul compression
IV. Fog Radio Access Networks
  A. Edge caching and F-RAN
  B. Information-theoretic analysis
  C. Optimization of signal processing strategies
V. Outlook

Technion Israel Institute of Technology

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 $\log(\mathrm{snr})$, oblivious coding approaches are advantageous over CoF based structured codes.

Technion Israel Institute of Technology

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-trasnfer mode when the fronthaul links have small capacity due to the more efficient use of fronthaul resources.

Technion Israel Institute of Technology

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]
  - 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]).
  - 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]

Technion Israel Institute of Technology

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?

Technion Israel Institute of Technology

Concluding Remarks

- Interesting open problems (ctd')
  - Fog-RAN
    Caching strategies in the eRRHs:
    [XPeng et al '15][Ugur et al '15][Wang et al '16][Chen et al '16]
    Robust cloud and edge processing with imperfect CSI
    [Dhifallah et al arXiv'16]
    Cloud and edge processing with pipelined fronthaul-wireless transmission
    [Sengupta et al arXiv'16]

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] I. E. Aguerri and A. Zaidi, “Lossy compression for compute-and-forward in limited backhaul uplink multicell processing," arXiv:1602.08714, Feb. 2016.

[5] 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.

[6] 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.

[7] 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.

Summarizing/Tutorial References

[8] 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].

[9] 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.

[10] 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.

[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," arXiv:1608.05768, Aug. 2016.

[12] 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.

[13] A. Zaidi and I. E. Aguerri, "Tutorial: Fronthaul Compression for Cloud Radio Access Networks," *The Thirteenth Internationl Symposium on Wireless Communication Systems* (*ISWCS '16*), Sep. 20-23, 2016, in Poznań, Poland.

Summarizing/Tutorial References

[14] 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.

Technion
Israel Institute of Technology

Thank you!

Technion
Israel Institute of Technology

References

References

[Alc-Luc] J. Segel and M. Weldon, "Lightradio portfolio-technical overview," Technology White Paper 1, Alcatel-Lucent.

[China] China Mobile, "C-RAN: the road towards green RAN," White Paper, ver. 2.5, China Mobile Research Institute, Oct. 2011.

[CPRI] Ericsson AB, Huawei Technologies, NEC Corporation, Alcatel Lucent and Nokia Siemens Networks, "Common public radio interface (CPRI): interface specification," CPRI specification v5.0, Sep. 2011.

[IDC] Integrated Device Technol. Inc., "Front-haul compression for emerging C-RAN and small cell networks," Apr. 2013.

[Costa '83] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[Tse-Hanly '98] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels-Part I: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796-2815, Nov. 1998.

[Viswanath et al '02] P. Viswanath, D. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277-1294, Jun. 2002.

References

[Shamai-Steiner '03] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617-2635, Oct. 2003.

[Koetter-Medard '03] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782-795, Oct. 2003.

[Sanderovich et al '08] A. Sanderovich, S. Shamai (Shitz), Y. Steinberg and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008-3023, Jul. 2008.

[Simeone et al '09] O. Simeone, O. Somekh, H. V. Poor and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Adv. Sig. Proc.*, 2009.

[3GPP-TR-136942] 3GPP TR 36.942 ver. 8.1.0, Rel. 8, Jan. 2009.

[dCoso-Ibars] A. d. Coso and C. Ibars, "Achievable rates for the AWGN channel with multiple parallel relays," *IEEE Trans. Wireless Comm.*, vol. 8, no. 5, pp. 2524-2534, May 2009.

[Nazer et al '09] B. Nazer, A. Sanderovich, M. Gastpar and S. Shamai (Shitz), "Structured superposition for backhaul constrained cellular uplink," in *Proc. IEEE ISIT '09*, Seoul, Korea, Jun. 2009.

References

[Sanderovich et al '09] A. Sanderovich, O. Somekh, H. V. Poor and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457-3478, Aug. 2009.

[dCoso-Simoens '09] A.d.Coso and S.Simoens, "Distributed compression for MIMO coordinated networks with a backhaul constraint," *IEEE Trans. Wireless Comm.*, vol. 8, no. 9, pp. 4698-4709, Sep. 2009.

[Marsch-Fettweis '09] P. Marsch and G. Fettweis, "On downlink network MIMO under an constrained backhaul and imperfect channel knowledge," in *proc. IEEE Globecom '09*, Honolulu, Hawaii, Nov. 2009.

[Verdu-Shamai '10] S. Verdu and S. Shamai (Shitz), "Variable-rate channel capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2651-2667, Jun. 2010.

[Huh et al '10] H. Huh, H. C. Papadopoulos and G. Caire, "Multiuser MISO transmitter optimization for intercell interference mitigation," *IEEE Trans. Sig. Processing*, vol. 58, no. 8, pp. 4272-4285, Aug. 2010.

[Irmer et al '11] R. Irmer, H. Droste, P. Marsch, M. Gastpar, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Comm. Mag.*, vol. 49, no. 2, pp. 102-111, Feb. 2011.

References

[Lim et al '11] S. H. Lim, Y.-H. Kim, A. E. Gamal and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132-3152, May 2011.

[Yassaee-Aref '11] M. H. Yassaee and M. R. Aref, "Slepian-Wolf coding over cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3462-3482, Jun. 2011.

[Hoydis et al '11] J. Hoydis, M. Kobayashi and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Sig. Proc.*, vol. 59, no. 6, pp. 2824-2833, Jun. 2011.

[Hong-Caire '11] S.-N. Hong and G. Caire, "Quantized compute and forward: A low-complexity architecture for distributed antenna systems," in *Proc. IEEE ITW '11*, Paraty, Brazil, Oct. 2011.

[Beck-Teboulle '11] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery problems," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, eds., pp. 42-88, Cambridge University Press. 2010.

[ElGamal-Kim '11] A. E. Gamal and Y.-H. Kim, "Network information theory," Cambridge University Press, 2011.

References

[Simeone et al '12] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor and S. Shamai (Shitz), "Cooperative wireless cellular systems: An information-theoretic view," *Foundations and Trends in Communications and Information Theory*, vol. 8, nos. 1-2, pp. 1-177, 2012.

[Wainwright '12] M. Wainwright, "Graphical models and message-passing: Some introductory lectures," tutorial available at www.eecs.berkeley.edu/wainwrig/kyoto12.

[Park et al TVT'13] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Tech.*, vol. 62, no. 2, pp. 692-703, Feb. 2013.

[Park et al SPL'13] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Joint decompression and decoding for cloud radio access networks," *IEEE Sig. Processing Letters*, vol. 20, no. 5, pp. 503-506, May 2013.

[Hong-Caire '13] S.-N. Hong and G. Caire, "Compute-and-forward strategy for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227-5243, Sep. 2013.

[Hong-Caire '13] S.-N. Hong and G. Caire, "Compute-and-forward strategy for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227-5243, Sep. 2013.

References

[Ni et al '13] W. Ni, R. P. Liu, I. B. Collings and X. Wang, "Indoor cooperative small cells over Ethernet," *IEEE Comm. Mag.*, vol. 51, no. 9, pp. 100-107, Sep. 2013.

[Zhou-Yu '13] L. Zhou and W. Yu, "Uplink multicell processing with limited backhaul via per-base-station successive interference cancellation," *IEEE Journ. Sel. Areas Comm.*, vol. 31, no. 10, pp. 1981-1993, Oct. 2013.

[Park et al TSP'13] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Sig. Processing*, vol. 61, no. 22, pp. 5646-5658, Nov. 2013.

[Karasik et al '13] R. Karasik, O. Simeone and S. Shamai (Shitz), "Robust uplink communications over fading channels with variable backhaul connectivity," *IEEE Trans. Wireless Comm.*, vol. 12, no. 11, pp. 5788-5799, Nov. 2013.

[Pedarsani et al '13] R. Pedarsani, M. A. Maddah-Ali and U. Niesen, "Online coded caching," arXiv:1311.3646, Nov. 2013.

[Liao et al '13] W.-C. Liao, M. Hong, H. Farmanbar, X. Li, Z.-Q. Luo and H. Zhang, "Min flow rate maximization for software defined radio access networks," arXiv:1312.5345, Dec. 2013.

References

[Golrezaei et al '13] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402-8413, Dec. 2013.

[Patil-Yu '14] P. Patil and W. Yu, "Hybrid compression and message-sharing strategy for the downlink cloud radio-access network," in *Proc. Inf. Theory and Application Workshop (ITA) 2014*, San Diego, CA, Feb. 2014.

[Kang et al TWC'14] J. Kang, O. Simeone, J. Kang and S. Shamai (Shitz), "Joint signal and channel state information compression for the backhaul of uplink network MIMO systems," *IEEE Trans. Wireless Comm.*, vol. 13, no. 3, pp. 1555-1567, Mar. 2014.

[Park et al CISS'14] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Performance evaluation of multiterminal backhaul compression for cloud radio access networks," in *Proc. IEEE CISS '14*, Princeton, NJ, Mar. 2014.

[Rost et al '14] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella and D. Wubben, "Cloud technologies for flexible 5G radio access networks," *IEEE Comm. Mag.*, vol. 52, no. 5, pp. 68-76, May 2014.

[MAli-Niesen '14] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856-2867, May 2014.

References

[Dhillon-Caire '14] H. S. Dhillon and G. Caire, "Wireless backhaul networks: capacity bound, scalability analysis and design guidelines," arXiv:1406.2738, Jun. 2014

[Liu-Kang '14] N. Liu and W. Kang, "A new achievability scheme for downlink multicell processing with finite backhaul capacity," in *Proc. IEEE ISIT '14*, Honolulu, Hawaii, Jul. 2014.

[He-Nazer-Shamai '14] W. He, B. Nazer and S. Shamai (Shitz), "Uplink-downlink duality for integer-forcing," in *Proc. IEEE ISIT '14*, Honolulu, Hawaii, Jul. 2014.

[Park et al TVT'15] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Multihop backhaul compression for the uplink of cloud radio access networks," to appear in *IEEE Trans. Veh. Technol.*, 2015.

[Park et al WCL'14] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Inter-cluster design of precoding and fronthaul compression for cloud radio access networks," *IEEE Wireless Comm. Ltters*, vol. 3, no. 4, pp. 369-372, Aug. 2014.

[Andrews et al JSAC'14] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong and J. C. Zhang, "What will 5G be?" *IEEE Journ. Sel. Areas Comm.*, Special Issue on 5G Communication Systems, vol. 32, no. 6, pp. 1065-1082, Sep. 2014.

[Park et al SPM'14] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Fronthaul compression for cloud radio access networks," *IEEE Sig. Processing Mag.*, vol. 31, no. 6, pp. 69-79, Nov. 2014.

References

[Katz-Zaidel-Shamai ‘14] G. Katz, B. M. Zaidel and S. Shamai (Shitz), “On layered strategies for cooperative processing in clustered cellular architectures,” submitted to *IEEE Trans. Veh. Tech.*

[Jain-Kim-Giannakis ‘14] S. Jain, S.-J. Kim and G. B. Giannakis, “Backhaul-constrained multi-cell cooperation leveraging sparsity and spectral clustering,” arXiv:1409.8359, Sep. 2014.

[Kang et al arXiv‘14] J. Kang, O. Simeone, J. Kang and S. Shamai (Shitz), “Fronthaul compression and precoding design for C-RANs over Ergodic fading channels,” arXiv:1412.7713, Dec. 2014.

[MPeng et al WC‘15] 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.

[MPeng et al arXiv‘15] M. Peng, S. Yan, K. Zhang and C. Wang, “Fog computing based radio access networks: Issues and challenges,” arXiv:1506.04233, Jun. 2015.

[MAli-Niesen ‘15] M. A. Maddah-Ali and U. Niesen, “Decentralized coded caching attains order-optimal memory-rate tradeoff,” *IEEE/ACM Trans. Networking*, vol. 23, no. 4, pp. 1029-1040, Aug. 2015.

References

[Lundqvist, CTN-Sep. 2015] H. Lundqvist, Senior Research, Huawei Sweden, "Death by starvation?: backhaul and 5G," *IEEE ComSoc Technology News*, CTN Issue: Sep. 2015.

[XPeng et al '15] X. Peng, J.-C. Shen, J. Zhang and K. B. Letaief, "Backhaul-aware caching placement for wireless networks," arXiv:1509.00558, Sep. 2015.

[Hong-Caire '15] S.-N. Hong and G. Caire, "Beyond scaling laws: On the rate performance of dense device-to-device wireless networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4735-4750, Sep. 2015.

[Bi et al '15] S. Bi, R. Zhang, Z. Ding and S. Cui, "Wireless communications in the era of big data," arXiv:1508.06369, Aug. 2016.

[China '15] China Mobile, "Next generation fronthaul interface," White Paper, Oct. 2015.

[Ugur et al '15] Y. Ugur, Z. H. Awan and A. Sezgin, "Cloud radio access networks with coded caching," arXiv:1512.02385, Dec. 2015.

[Tao et al '15] M. Tao, E. Chen, H. Zhou and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," arXiv:1512.06938, Dec. 2015.

References

[Tandon-Simeone '16] R. Tandon and O. Simeone, "Fog radio access networks: Fundamental latency trade-offs," in *Proc. IEEE Inf. Theory and Applications Workshop (ITA) 2016*, La Jolla, CA, Jan. 2016.

[Park-Simeone-Shamai arXiv'16] S.-H. Park, O. Simeone and S. Shamai (Shitz), "Joint optimization of cloud and edge processing for fog radio access networks," arXiv:1601.02460, Jan. 2016.

[Dai-Yu '16] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," arXiv:1601.01070, Jan. 2016.

[Chen et al '16] Z. Chen, J. Lee, T. Q. S. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," arXiv:1601.00321, Jan. 2016.

[Naderializadeh et al '16] N. Naderializadeh, M. A. Maddah-Ali and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," arXiv:1602.04207, Feb. 2016.

[Cisco '16] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020," White paper, Feb. 2016.

References

[Wang et al '16] R. Wang, J. Zhang and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," arXiv:1605.03709, May 2016.

[Dhifallah et al '16] O. Dhifallah, H. Dahrouj, T. Y. Al-Naffouri and M.-S. Alouini, "Robust beamforming in cache-enabled cloud radio access networks," arXiv:1605.09321, May 2016.

[Liu-Patil-Yu '16] L. Liu, P. Patil and W. Yu, "An uplink-downlink duality for cloud radio access network," in *Proc. IEEE Intern. Symp. Inf. Theory (ISIT) 2016*, Barcelona, Spain, Jul. 2016.

[Park-Simeone-Shamai SPAWC'16] S.-H. Park, O. Simeone and S. Shamai (Shitz), "Joint cloud and edge processing for latency minimization in fog radio access networks," in *Proc. IEEE SPAWC 2016*, Edingburgh, UK, Jul. 2016.

[Agiwal et al '16] 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.

[Yang et al '16] T. Yang, N. Liu, W. Kang and S. Shamai, "An upper bound on the sum capacity of the downlink multicell processing with finite backhaul capacity," arXiv:1609.00833, Sep. 2016.

References

[Aguerri et al ‘17] I. E. Aguerri, A. Zaidi, G. Caire and S. Shamai (Shitz), "On the Capacity of Cloud Radio Access Networks with Oblivious Relaying,” arXiv:1701.07237, Jan. 2017.

[Wu et al ‘17] X. Wu, L. P. Barnes and A. Ozgur, “The capacity of the relay channel,” arXiv:1701.02043, Jan. 2017.

Abstract

Fronthaul Constrained Cloud and Fog Radio Access Networks: An Information Theoretic View

---

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 (NJIT), and O. Sahin (InterDigital)
Recent work of S. Shamai has been supported by the European Union's Horizon 2020, Research And Innovation Program: 694630.