Cloud Radio Access Networks, Distributed Information Bottleneck, and more: A Unified Information Theoretic View

Shlomo Shamai

Technion—Israel Institute of Technology

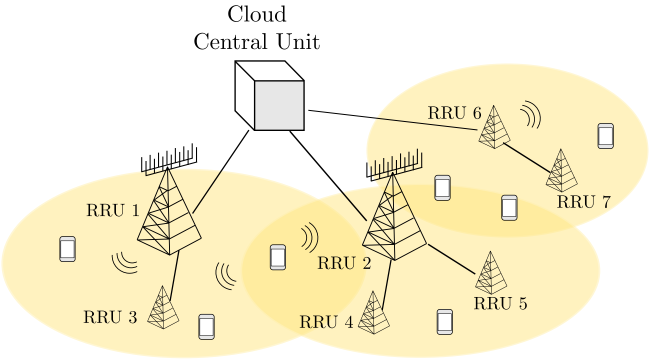
Joint work with: Inaki Estella Aguerri (Paris Research Center, Huawei France) Abdellatif Zaidi (Université Paris-Est, Paris) and Giuseppe Caire (TUB, Berlin)

sshlomo@ee.technion.ac.il

Columbia University, EE Dept. Seminar, 19 February, 2019

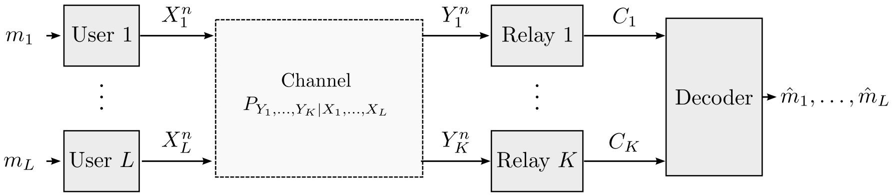
Outline

Introduction



- Cloud radio access network (C-RAN) architecture:
  - Heterogeneous dense networks;
  - Base stations (BSs), macro, pico, femto, operate as radio units (RUs);
  - Baseband processing takes place in the “cloud” or a central unit (CU).

Uplink Cloud RAN



- Multiple access relay channel in which L users communicate with a common destination through K relay nodes.

- Decoder interested in $\hat{m}_1, \ldots, \hat{m}_L$ such that, for n large enough,

$$\Pr\{(m_1, \ldots, m_L) \neq (\hat{m}_1, \ldots, \hat{m}_L)\} \to 0$$

- The capacity region of this model is still to be found
  - problem open even in seemingly simpler cases, e.g., one user and two relays (the diamond channel), parallel Gaussian relay channel *[Schein-Gallager '00]*.

Relay Operations

- Main difficulty is in characterizing the optimal relay operation:
  - **Decode-and-Forward (DF):** *[Cover-ElGamal'97], [Kramer-Gastpar'05] ...*
  - **Compute-and-Forward (CompF):** *[Nazer-Gastpar'11], [Nazer et al'12], [Hong-Caire'13]...*
  - **Compress-and-Forward (CoF):** *[Sanderovich et al'09], [Park et al'13], [Zhou et at'13]...*
  - **Noisy Network Coding (NNC):** *[Lim et al'11]...*
  - **Others:** Amplify and Forward, Partial-Decode-Compress-and-Forward *[Cover-ElGamal'97]*, Compute-Quantize-and-Forward *[Estella-Zaidi'16]*.
- Relaying operations can be divided into:
  - Non-oblivious: relays aware of the users' codebooks (modulation, coding...) at all time, e.g., DF, CompF.
  - Oblivious (or Nomadic): *[Sanderovich et al'08]* relays operate without knowledge of the users' codebooks, e.g., CoF, NNC.
- Oblivious processing motivated mainly by practical constraints.
- Formally, obliviousness of the relays to actual codebooks is modeled through randomized encoding *[Sanderovich et al'08], [Lapidoth-Narayan'98]*.

Randomized Encoding as a Model for Obliviousness

- Encoding function at transmitter

$$\phi^n : [1, |\mathcal{X}|^{n2^{nR}}] \times [1, 2^{nR}] \to \mathcal{X}^n$$

which maps:

  - a codebook index $F \in [1, |\mathcal{X}|^{n2^{nR}}]$ and
  - a message $M \in [1, 2^{nR}]$

into a codeword

$$X^n(F, M) = \phi^n(F, M).$$

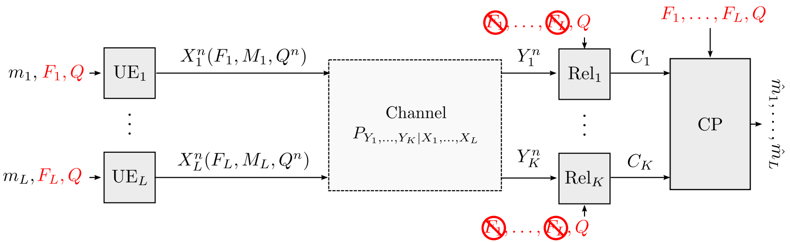
- The pair (p_F, ϕ^n) must satisfy

$$\text{Prob}[X^n(F, M) = x^n] = \prod_{i=1}^{n} p_X(x_i)$$

for some $p_X(x)$, $x \in \mathcal{X}$, where $\text{Prob}[\cdot]$ is calculated with respect to

$$p_{F,M}(f, m) = p_F(f) \cdot 2^{-nR}.$$

Oblivious Relay Processing with Enabled Resource-sharing



- Resource-sharing random variable Q^n available at all terminals *[Simeone et al'11]*.
- Q^n way easier to share, (e.g., on/off activity).
- Memoryless Channel: $P_{Y_1,\ldots,Y_K|X_1,\ldots,X_1}$
- User $l \in \{1,\ldots,L\}$: $\phi_l^n : [1, |\mathcal{X}_l|^{n2^{nR_l}}] \times [1, 2^{nR_l}] \times \mathcal{Q}^n \to \mathcal{X}_l^n$
- Relay $k \in \{1,\ldots,K\}$: $g_k^n : {\mathcal{Y}_k}^n \times \mathcal{Q}^n \to [1, 2^{nC_k}]$
- Decoder:

$$\psi^n : [1, |\mathcal{X}_1|^{n2^{nR_1}}] \times \cdots \times [1, 2^{nC_K}] \times \mathcal{Q}^n \to [1, 2^{nR_1}] \times \ldots \times [1, 2^{nR_L}]$$

Main Capacity Results

Single-letter characterizations of:

1) *Capacity Region* of the Class of DM CRAN channels satisfying

$$Y_k^n -\!\!\circ\!\!- X_{\mathcal{L}}^n -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}^n,$$

2) *Capacity Region of Gaussian MIMO Channels with Gaussian Inputs*
   - In particular, we show that Gaussian auxiliaries are optimal.
   - And, time (frequency) sharing is in general needed.

3) *Inner and Outer Bounds* for General DM Model (i.e., Without the Markov Chain).

Capacity Region of a Class of CRAN Channels

Theorem

For the class of discrete memoryless channels satisfying

$$Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}$$

with oblivious relay processing and enabled resource-sharing, a rate tuple $(R_1, \ldots, R_L)$ *is achievable if and only if for all* $\mathcal{T} \subseteq \mathcal{L}$ *and for all* $\mathcal{S} \subseteq \mathcal{K}$*,*

$$\sum_{t\in\mathcal{T}} R_t \leq \sum_{s\in\mathcal{S}} [C_s - I(Y_s; U_s | X_{\mathcal{L}}, Q)] + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q),$$

for some joint measure of the form

$$P_Q \prod_{l=1}^{L} P_{X_l|Q} \prod_{k=1}^{K} P_{Y_k|X_{\mathcal{L}}} \prod_{k=1}^{K} P_{U_k|Y_k,Q},$$

with the cardinality of Q *bounded as* $|\mathcal{Q}| \leq K + 2$*.*

Direct Part

Capacity region achievable with

- **Compress-and-Forward with Joint-Decompression-Decoding**
  - Generalization of scheme from *[Sanderovich et al'09]* to L users.
  - Based on compress-and-forward à la Cover-El Gamal with joint decoding and decompression **(JDD)** at the CP.
  - Gaussian inputs are not optimal for finite capacity fronthauls.
- Separate Decompression-Decoding not optimal in general.
- **Noisy Network Coding**
  - Particular case of *[Theorem 1, Lim et al'11]*.

Sum-rate achievable also with

- **Compress-and-Forward with Separate Decompression-Decoding (SDD)**
  - The CP decodes explicitly the compression indices first and then decodes the users' transmitted messages.

Outline of Converse Part

- Define $U_{i,k} := (J_k, Y_k^{i-1})$ and $\bar{Q}_i := (X_{\mathcal{L}}^{i-1}, X_{\mathcal{L},i+1}^n, \tilde{Q})$.

  Fano's Inequality: $H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \leq \epsilon_n$ for $\mathcal{T} \subseteq \mathcal{L}$,

- **Upper bound on entropy term:** For $\mathcal{T} \subseteq \mathcal{L} := \{1, \ldots, L\}$,

$$H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, Q^n) \leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - n\sum_{t\in\mathcal{T}} R_t := n\Gamma_{\mathcal{T}}$$

  - Follows from

$$\begin{aligned} n\sum_{t\in\mathcal{T}} R_t = H(m_{\mathcal{T}}) &= I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) + H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \\ &\leq I(m_{\mathcal{T}}, F_{\mathcal{T}}; J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \\ &\leq H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, \tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q}) + n\epsilon_n \end{aligned}$$

- Reminiscent of log-loss penalty criterion in multi-terminal source coding *[Courtade-Weissman'14]*:

$$H(X^n|J_{\mathcal{K}}) \leq \mathrm{E}[d_{\log}(X^n; \hat{X}^n)] \simeq n(H(X) - I(X; \hat{X}))$$

Outline of Converse Part (Cont.)

- **Bound on users' rates:** For $\mathcal{T} \subseteq \mathcal{L}$

$$n \sum_{t \in \mathcal{T}} R_t \leq I(m_{\mathcal{T}}, F_{\mathcal{T}}; J_{\mathcal{K}} | F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \leq \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{K},i} | X_{\mathcal{T}^c,i}, \bar{Q}_i) + n\epsilon_n$$

- **Bound on relays' rates:** For $\mathcal{S} \subseteq \mathcal{K} := \{1, \ldots, K\}$

$$\begin{aligned} n \sum_{k \in \mathcal{S}} C_k &\geq \sum_{k \in \mathcal{S}} H(J_k) \geq I(X_{\mathcal{T}}^n, Y_{\mathcal{S}}^n; J_{\mathcal{S}} | X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) \\ &\geq \sum_{i=1}^{n} H(X_{\mathcal{T},i} | X_{\mathcal{T}^c,i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) - n\Gamma_{\mathcal{T}} + I(Y_{\mathcal{S}}^n; J_{\mathcal{S}} | X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \\ &= nR_{\mathcal{T}} - \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{S}^c,i} | X_{\mathcal{T}^c,i}, \bar{Q}_i) + \sum_{k \in \mathcal{S}} \sum_{i=1}^{n} I(Y_{k,i}; U_{k,i} | X_{\mathcal{L},i}, \bar{Q}_i) \end{aligned}$$

where we used the upper bound on the entropy and the Markov chain

$$Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K} \setminus k}$$

Remarks

- Sum-rate achievable with CF with JDD given by

$$R^{\text{sum}}_{\text{JDD}} = \max \min_{\mathcal{S} \subseteq \mathcal{K}} \Big\{ \sum_{s \in \mathcal{S}} C_s - I(Y_S; U_S | X_{\mathcal{L}}, U_{S^c}, Q) + I(X_{\mathcal{L}}; U_{\mathcal{S}^c} | Q) \Big\}.$$

- Using properties of sub-modular functions, we show that CF with SDD (and even the low-complexity version of it, consisting in *sequential* decompression followed by *sequential* decoding, denoted as SWZ) achieve the same sum-rate as CF with JDD.
- Note, however, that time-sharing is generally needed for the three to achieve optimal sum-rate!

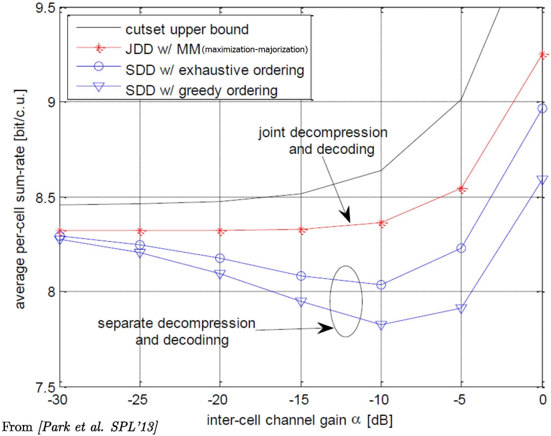
Theorem

For any $P_{Y_1,\ldots,Y_K|X_1,\ldots,X_L}$, not necessarily satisfying $Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K}\backslash k}$, we have

$$R^{\text{sum}}_{\text{JDD}} = R^{\text{sum}}_{\text{SDD}} = R^{\text{sum}}_{\text{SWZ}}$$

- In particular, for MIMO Gaussian channels recovers *[Zhou et al.'16]*.
- In terms of rate-region, CF with JDD generally outperforms CF with SDD.

Numerical example: 3 Cell Uplink Wyner Model



From *[Park et al. SPL'13]*

- For SWZ, optimizing over relay ordering improves performance, in general
- Without time (or resource)-sharing, as is in the figure, SDD may achieve smaller sum-rate than JDD.

Memoryless MIMO Gaussian Model

- The channel output at relay node k with M_k antennas:

$$\mathbf{Y}_k = \mathbf{H}_{k,\mathcal{L}} \, [\mathbf{X}_1^T, \ldots, \mathbf{X}_L^T]^T + \mathbf{N}_k,$$

  where

  - User l with N_l antennas transmits $\mathbf{X}_l$ with $\mathrm{E}[\|\mathbf{X}_l\|^2] \preceq \mathbf{K}_l$.
  - Relay k with M_k antennas.
  - $\mathbf{H}_{k,\mathcal{L}} = [\mathbf{H}_{k,1}, \ldots, \mathbf{H}_{k,L}]$, $\mathbf{H}_{k,l}$ channel between user l and relay k.
  - $\mathbf{N}_k \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$ is AWGN noise at relay k, assumed independent.

- Outputs satisfy $Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}$.

- Theorem 1 characterizes its **capacity region**. Finding the optimal $U_1, \ldots, U_K$ is generally not easy.

Capacity under Gaussian Signaling and Enabled Resource-Sharing

Theorem (Capacity Region under Gaussian Input with Enabled Resource-Sharing)

Let the input vectors use Gaussian Signaling with Enabled Resource-Sharing, i.e.,

$$\mathbf{X}_{l,q} \sim \mathcal{CN}(0, \mathbf{K}_{l,q}) \quad q \in \{1, \dots, |\mathcal{Q}|\} \qquad \sum_{q \in \mathcal{Q}} p_Q(q) \mathbf{K}_{l,q} \leq \mathbf{K}_l$$

The capacity region is given by the set of all rate tuples $(R_1, \dots, R_L)$ satisfying that for all $\mathcal{T} \subseteq \mathcal{L}$ and all $\mathcal{S} \subseteq \mathcal{K}$

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{k \in \mathcal{S}} \left[C_k - \mathrm{E}_Q \log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_{k,q}|} \right] + \mathrm{E}_Q \log \frac{|\sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_{k,q} \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},}^{-1}|}{|\mathbf{K}_{\mathcal{T},q}^{-1}|},$$

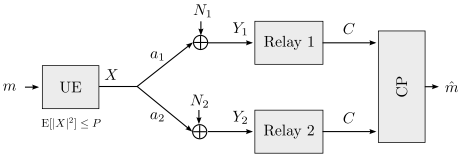
for some $\mathbf{0} \preceq \mathbf{B}_{k,q} \preceq \mathbf{\Sigma}_k^{-1}$, where $\mathbf{H}_{k,\mathcal{T}}$ is the channel between $\mathbf{X}_{\mathcal{T}}$ and $\mathbf{Y}_k$.

- Extends *[Theorem 5, Sanderovich et al'09]* to L users and MIMO.
- Achievable with $\mathbf{U}_{k,q} = \mathbf{Y}_{k,q} + \mathbf{Z}_{k,q}$, $\mathbf{Z}_{k,q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{B}_{k,q}^{-1} - \mathbf{\Sigma}_{k,q})$, $q \in \mathcal{Q}$.
- Gaussian signaling can be strictly suboptimal *[Sanderovich et al'09]*.

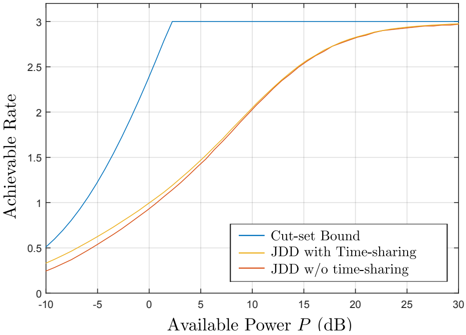
Converse Part

- For (X, U) arbitrarily correlated,
$$\log |(\pi e)\ \mathbf{J}^{-1}(X|U)| \leq h(X|U) \leq \log |(\pi e)\ \mathrm{mmse}(X|U)|$$
- For each $Q = q$,
$$I(\mathbf{Y}_k; \mathbf{U}_k | \mathbf{X}_{\mathcal{L}}, Q = q) = \log |(\pi e)\mathbf{\Sigma}_k| - h(\mathbf{Y}_s | \mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{s,q}, Q = q)$$
$$\geq \log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_{k,q}|},$$
where $\mathbf{0} \preceq \mathbf{B}_{k,q} \preceq \mathbf{\Sigma}_k^{-1}$ is chosen such that
$$\mathrm{mmse}(\mathbf{Y}_k | \mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{k,q}) = \mathbf{\Sigma}_k - \mathbf{\Sigma}_k \mathbf{B}_{k,q} \mathbf{\Sigma}_k$$
- Also,
$$I(\mathbf{X}_{\mathcal{T}}; \mathbf{U}_{S^c} | \mathbf{X}_{\mathcal{T}^c}, Q = q) = h(\mathbf{X}_{\mathcal{T},q} | \mathbf{X}_{\mathcal{T}^c}, q) - h(\mathbf{X}_{\mathcal{T}} | \mathbf{X}_{\mathcal{T}^c}, \mathbf{U}_{S^c,q}, Q = q)$$
$$\leq \log |\mathbf{K}_{\mathcal{T},q}| + \log \left| \sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_{k,q} \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},q}^{-1} \right|$$
by deBrujin Identity *[Palomar-Verdu'06],[Ekrem-Ulukuss'14], [Zhou et al'17]*
$$\mathbf{J}(\mathbf{X}_{\mathcal{T},q} | \mathbf{X}_{\mathcal{T}^c,q}, \mathbf{U}_{S^c,q}) = \sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_{k,q} \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},q}^{-1}.$$

Resource-sharing Enlarges Capacity Region



$Y_k = a_k X + N_k$, with $\mathrm{E}[X^2] \leq P$ and $N_k \sim \mathcal{N}(0,1)$, $k = 1, 2$



- JDD without resource (time)-sharing, i.e., $Q = \emptyset$
- JDD with resource (time)-sharing ($|\mathcal{Q}| = 2$. Recall that $|\mathcal{Q}| \leq K + 2 = 4$ here)
  - Phase I: UE transmits at P/α for αn samples. Relays compress at C_k/α.
  - Phase II: UE and Relays remain inactive for $(1-\alpha)n$ remaining samples.
- Intuition: For small P, the observations at the relays are too noisy; and, so, it is more advantageous to increase power and compression rate during shorter time.

Capacity under Constant Gaussian Signaling

Theorem (Capacity Region under Constant Gaussian Input)

If the input vectors use constant Gaussian Signaling, i.e.,

$$\mathbf{K}_{1,l} = \cdots = \mathbf{K}_{|\mathcal{Q}|,l} = \mathbf{K}_l, \qquad \mathbf{X}_l \sim \mathcal{CN}(0, \mathbf{K}_l),$$

the capacity region is given by the set of all rate tuples $(R_1, \ldots, R_L)$ *satisfying that for all* $\mathcal{T} \subseteq \mathcal{L}$ *and all* $\mathcal{S} \subseteq \mathcal{K}$

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{k \in \mathcal{S}} \left[C_k - \log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_k|} \right] + \log \frac{|\sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_k \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}|}{|\mathbf{K}_{\mathcal{T}}^{-1}|},$$

for some $\mathbf{0} \preceq \mathbf{B}_k \preceq \mathbf{\Sigma}_k^{-1}$, *where* $\mathbf{H}_{k,\mathcal{T}}$ *is the channel between* $\mathbf{X}_{\mathcal{T}}$ *and* $\mathbf{Y}_k$*.*

- Resource-sharing at the relays does not enlarge the capacity region under constant Gaussian Signaling.
- Proof follows from Jensen's Inequality and concavity of log-det.

Capacity under Gaussian Signaling in the High SNR Regime

- High SNR regime model:

$$\mathbf{\Sigma}_k = \epsilon \tilde{\mathbf{\Sigma}}_k; \quad \text{for some } \tilde{\mathbf{\Sigma}}_k \succeq \mathbf{0}, \quad \text{and } \epsilon \to 0.$$

- We have $\mathcal{R}_{\mathrm{GNS}}(C_{\mathcal{K}}) \subset \mathcal{R}_{\mathrm{GTS}}(C_{\mathcal{K}})$, where:
  - $\mathcal{R}_{\mathrm{GTS}}$: Capacity under Gaussian Input with enabled resource-sharing.
  - $\mathcal{R}_{\mathrm{GNS}}$: Capacity under Gaussian Input without resource-sharing ($Q = \emptyset$).

Theorem (Capacity Region under Gaussian Input in High SNR)

If $(R_1, \ldots, R_L) \in \mathcal{R}_{\mathrm{GTS}}(C_{\mathcal{K}})$, *then for any* $\epsilon > 0$, *for some* $\Delta_\epsilon \geq 0$,

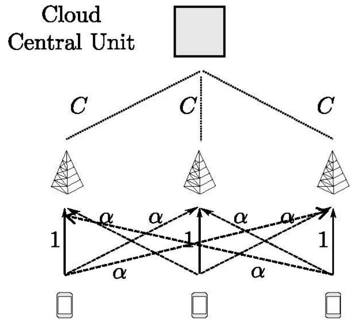
$$(R_1 - \Delta_\epsilon, \ldots, R_L - \Delta_\epsilon) \in \mathcal{R}_{\mathrm{GNS}}(C_{\mathcal{K}})$$

In addition, $\mathcal{R}_{\mathrm{GNS}}(C_{\mathcal{K}}) = \mathcal{R}_{\mathrm{GTS}}(C_{\mathcal{K}})$ *as* $\epsilon \to 0$, *since*

$$\lim_{\epsilon \to 0} \Delta_\epsilon = 0.$$

- For large SNR, the gains due to resource-sharing become limited.

Numerical example: Circular Wyner Model



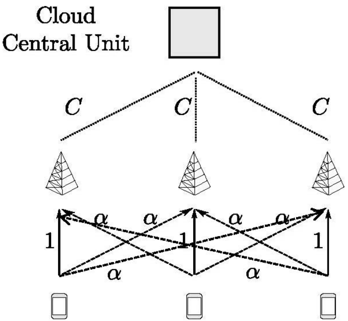
- Each cell contains a single-antenna and a single-antenna RU.
- Inter-cell interference takes place only between adjacent cells (circular).

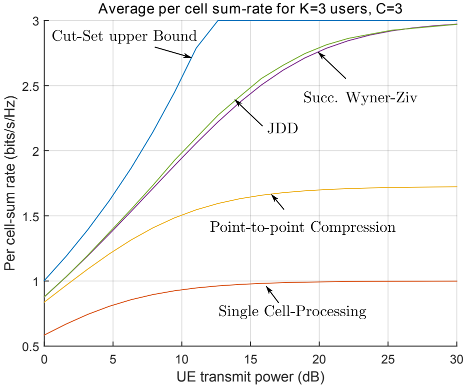
$$Y_k = \alpha X_{k-1} + X_k + \alpha X_{k+1} + N_k$$

where $N_k \sim \mathcal{CN}(0,1)$

- All RRUs have a fronthaul capacity of C.

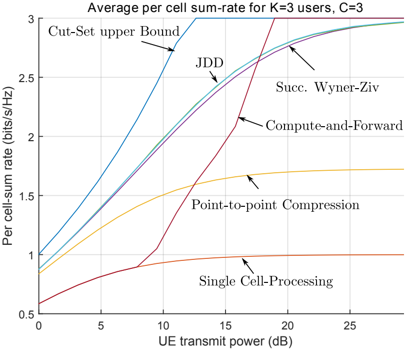
Numerical example: Circular Wyner Model (cont'd)

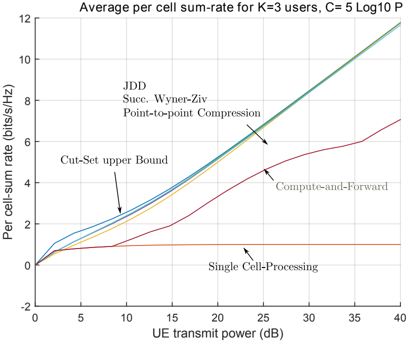




- JDD is capacity achieving under oblivious processing.
- For this simple network, JDD does not provide much gain compared to SDD and SWZ.
  - Here, the schemes SDD and SWZ do not employ resource-sharing.

Cost of Obliviousness





- **Optimal degrees-of-freedom:** when fronthaul capacity grows with SNR, e.g., $C = 5\log_{10}(\mathrm{snr})$. *[Sanderhovich et al'09].*
- Capacity under Gaussian signaling to within a **constant gap** of cut-set bound.
  - If $(R_1, \ldots, R_L)$ is within the cut-set bound, then

$$((R_1 - \Delta)^+, \ldots, (R_L - \Delta)^+), \qquad \Delta \le \begin{cases} \frac{N}{2}(2.45 + \log(\frac{KM}{N})) & \text{for } KM > 2N, \\ \frac{KM+N}{2} & \text{for } KM \le 2N \end{cases}$$

Inner and Outer Bounds for General CRAN Models

Theorem (Bounds)

For general DM CRAN channels with oblivious relay processing and enabled resource-sharing, a rate tuple $(R_1, \ldots, R_L)$ is achievable if (only if) for all $\mathcal{T} \subseteq \mathcal{L}$ and for all $\mathcal{S} \subseteq \mathcal{K}$,

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} C_s - I(Y_S; U_S | X_{\mathcal{L}}, U_{S^c}, Q) + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q),$$

Inner bound: *for some pmf* $P_Q \prod_{l=1}^{L} P_{X_l|Q} P_{Y_{\mathcal{K}}|X_{\mathcal{L}}} \prod_{k=1}^{K} P_{U_k|Y_k,Q}$.

Outer bound: *for some* $(Q, X_{\mathcal{L}}, Y_{\mathcal{K}}, U_{\mathcal{K}}, W)$

- *distributed according to* $P_Q \prod_{l=1}^{L} P_{X_l|Q} \; P_{Y_{\mathcal{K}}|X_{\mathcal{L}}} \; P_{W|Q}$
- $u_k = f_k(w, y_k, q)$ *for some random variable* W *and some deterministic functions* $\{f_k\}$, $k \in \mathcal{K}$.

- Problem is challenging, as it includes Korner-Marton modulo-sum problem *[Korner-Marton'79]* as a special case.

Information Bottleneck

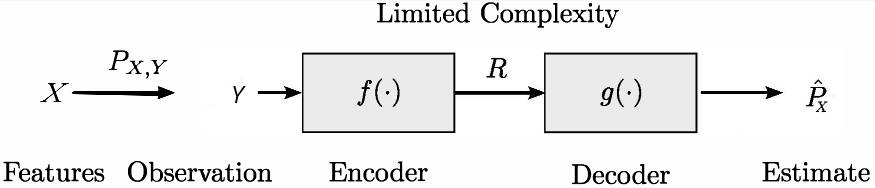
$$X \longrightarrow Y \longrightarrow U$$

- Efficiency of a given representation $U = f(Y)$ measured by the pair

  Rate (or *Complexity*): $I(U;Y)$ and Information (or *Relevance*): $I(U;X)$

- Information $I(X;U)$ can be achieved by OBLIVIOUS coding Y while with the logarithmic distortion with respect to X
- Single letter-wise, U is not necessarily a deterministic function of Y
- The non-oblivious bottleneck problem is immediate as the $\min(I(X;Y), R)$ is achievable by having the relay decoding the message transmitted by X
- The bottleneck problem connects to many timely aspects, such as 'deep learning' *[Tishby-Zaslavsky, ITW'15]*.

Digression: Learning via the Information Bottleneck Method



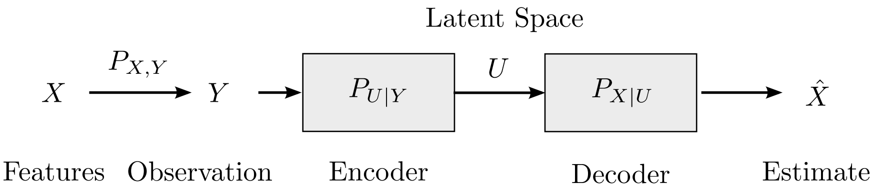
- Preserving all the information about X that is contained in Y, i.e., $I(X;Y)$, requires high *complexity* (in terms of *minimum description coding length*).
  - Other measures of complexity may be (Vapnik-Chervonenkis) VC-dimension, covering numbers, ..
- Efficiency of a given representation $U = f(Y)$ measured by the pair

  Complexity: $I(U;Y)$ and Relevance: $I(U;X)$
- Example:

$$\max_{p(u|x)} I(U;X) \quad \text{s.t.} \quad I(U;Y) \leq R, \quad \text{for} \quad 0 \leq R \leq H(Y)$$

$$\min_{p(u|x)} I(U;Y) \quad \text{s.t.} \quad I(U;X) \geq \Delta, \quad \text{for} \quad 0 \leq \Delta \leq I(X;Y)$$

Basically, a Remote Source Coding Problem !



- Reconstruction at decoder is under log-loss measure,

$$R(\Delta) = \min_{p(u|y)} I(U;Y)$$

where the minimization is over all conditional pmfs $p(u|y)$ such that

$$\mathbb{E}[\ell_{\text{log}}(X,U)] \leq H(X) - H(X|U) = H(X) - \Delta$$

- R. L. Dobrushin and B. S. Tsybakov, *"Information transmission with additional noise"*, IRE Tran. Info. Theory, Vol. IT-8, pp. 293-304, 1962.

- H. Witsenhausen, A. Wyner, *"A conditional entropy bound for a pair of discrete random variables"*, IEEE Trans. on Info. Theory, Vol. 21, pp. 493-501, 1975.

- Solution also coined as the Information Bottleneck Method *[Tishby'99]*

$$L_{\text{IB}}(\beta, P_{X,Y}) = \min_{p(u|y)} I(Y;U) - \beta I(X;U)$$

Other Connections

- **Common Reconstruction**. Because $U \multimap Y \multimap X$, we have

$$I(U;X) = I(U;Y) - I(U;Y|X)$$
$$\leq R - I(U;Y|X)$$

- Y. Steinberg, *"Coding and common reconstruction"*, IEEE Trans. on Info. Theory, vol. 55, no. 11, pp. 4995–5010 (X – side information is not used for the 'source' Y common reconstruction).

- **Information Combining**

$$I(Y;U,X) = I(U;Y) + I(X;Y) - I(U;X) \qquad \text{(since } U \multimap Y \multimap X\text{)}$$

Since $I(X;Y)$ is given and $I(Y;U) = R$, maximizing $I(U;X)$ is equivalent to minimizing $I(Y;U,X)$.

- I. Sutskover, S. Shamai and J. Ziv, "Extremes of Information Combining", IEEE Trans. Inform. Theory, vol. 51, no. 4, pp. 1313–1325, April 2005.

- I. Land and J. Huber, *"Information combining,"* Foundations and trends in Commun. and Inform. Theory, vol. 3, pp. 227–330, Nov. 2006.

Other Connections (Cont.)

- **Wyner-Ahlswede-Körner Problem**

  If X and Y are encoded at rates R_X and R_Y, respectively. For given $R_Y = R$, the minimum rate R_X that is needed to recover X losslessly is

$$R_X^\star(R) = \min_{p(u|y)\,:\,I(U;Y)\,\leq\,R} H(X|U)$$

  So, we get

$$\max_{p(u|y)\,:\,I(U;Y)\leq R} I(U;X) = H(X) - R_X^\star(R)$$

- R. F. Ahlswede and J. Korner, *"Source coding with side information and a converse for degraded broadcast channels"*, IEEE Trans. on Info. Theory, Vol. 21, pp. 629-637, 1975.

- A. D. Wyner, *"On source coding with side information at the decoder"*, IEEE Trans. on Info. Theory, Vol. 21, pp. 294-300, 1975.

Vector Gaussian Information Bottleneck

- $(\mathbf{X}, \mathbf{Y})$ jointly Gaussian, $\mathbf{X} \in \mathbb{R}^N$ and $\mathbf{Y} \in \mathbb{R}^M$
- Optimal encoding $P_{U|Y}$ is a noisy linear projection to a subspace whose dimensionality is determined by the bottleneck Lagrangian multiplier β *[Chechik et al. '05]*

$$\mathbf{U} = \mathbf{A}\mathbf{Y} + \mathbf{Z}, \qquad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

where

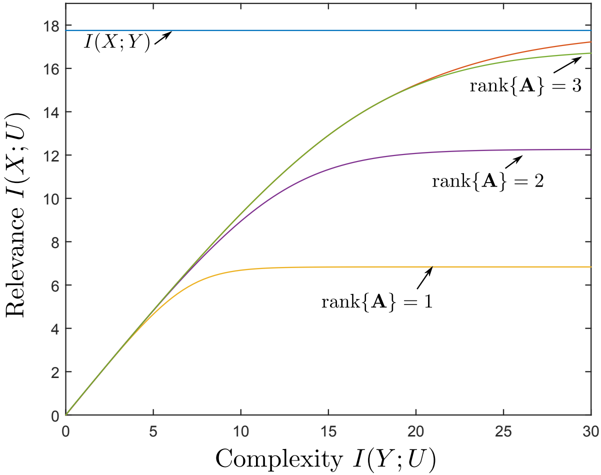
$$\mathbf{A} = \begin{cases} [\mathbf{0}^T; \ldots; \mathbf{0}^T], & \text{if } 0 \leq \beta \leq \beta_1^c \\ [\alpha_1 \mathbf{v}_1^T; \mathbf{0}^T; \ldots; \mathbf{0}^T], & \text{if } \beta_1^c \leq \beta \leq \beta_2^c \\ [\alpha_1 \mathbf{v}_1^T; \alpha_2 \mathbf{v}_2^T; \mathbf{0}^T; \ldots; \mathbf{0}^T], & \text{if } \beta_2^c \leq \beta \leq \beta_3^c \\ \vdots & \end{cases}$$

and $\{\mathbf{v}_1^T, \ldots, \mathbf{v}_N^T\}$ are the left eigenvectors of $\mathbf{\Sigma}_{y|x}\mathbf{\Sigma}_y^{-1}$, sorted by their ascending eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$; $\beta_i^c = 1/(1 - \lambda_i)$ are critical β values; $r_i = \mathbf{v}_i^T \mathbf{\Sigma}_y \mathbf{v}_i$ and

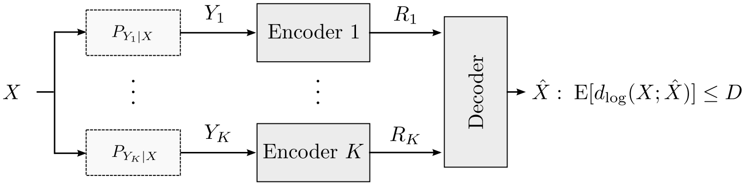
$$\alpha_i = \sqrt{\frac{\beta(1 - \lambda_i) - 1}{\lambda_i r_i}}$$

Rate-Information Trade-off Gaussian Vector Channel [Winkelbauer-Matz, ISIT'14].

Rate-Information Curve



CEO Source Coding Problem under Log-Loss



- CEO source coding problem under log-loss distortion:

$$d_{\log}(x,\hat{x}) := \log\left(\frac{1}{\hat{x}(x)}\right)$$

  where $\hat{x} \in \mathcal{P}(\mathcal{X})$ is a probability distribution on $\mathcal{X}$.

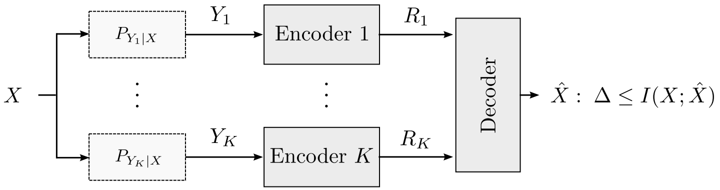
- Characterization of rate-distortion region in *[Courtade-Weissman'14]*
  - Key step: log-loss admits a lower bound in the form of conditional entropy of the source conditioned on the compression indices:

$$nD \geq \mathrm{E}[d_{\log}(X^n;\hat{X}^n)] \geq H(X^n|J_{\mathcal{K}}) = H(X^n) - I(X^n;J_{\mathcal{K}})$$

- Converse of Theorem 1 for Oblivious CRAN leverages on this relation applied to multiple channel inputs, which can be designed.

  Multiple description CEO problem-logloss distortion (Pichler-Piantanida-Matz, ISIT'17].

Distributed Information Bottleneck



- Information Bottleneck introduced by *[Tishby'99]* and *[Witsenhausen'80]* "Indirect Rate Distortion Problems", IT–26, no. 5, pp. 518–521, Sept. 1980.
- It is a CEO source-coding problem under log-loss!

Theorem (Distributed Information Bottleneck [Estella-Zaidi, IZS'18])

The D-IB region is the set of all tuples $(\Delta, R_1, \ldots, R_K)$ *which satisfy*

$$\Delta \leq \sum_{k \in \mathcal{S}} [R_k - I(Y_k; U_k | X, Q)] + I(X; U_{\mathcal{S}^c} | Q), \qquad \textit{for all } \mathcal{S} \subseteq \mathcal{K}$$

for some joint pmf $p(q)p(x) \prod_{k=1}^{K} p(y_k|x) \prod_{k=1}^{K} p(u_k|y_k, q)$.

Vector Gaussian Distributed Information Bottleneck

- $(\mathbf{Y}_1, \cdots, \mathbf{Y}_K, \mathbf{X})$ jointly Gaussian, $\mathbf{Y}_k \in \mathbb{R}^N$ and $\mathbf{X} \in \mathbb{R}^M$,

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{N}_k, \qquad \mathbf{N}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{n}_k})$$

- Optimal encoding $P^*_{U_k|Y_k}$ is Gaussian and $Q = \emptyset$ *[Estella-Zaidi'17]*

Theorem (Estella-Zaidi, IZS'18)

If $(\mathbf{X}, \mathbf{Y}_1, \ldots, \mathbf{Y}_K)$ are jointly Gaussian, the D-IB region is given by the set of all tuples $(\Delta, R_1, \ldots, R_L)$ satisfying that for all $\mathcal{S} \subseteq \mathcal{K}$

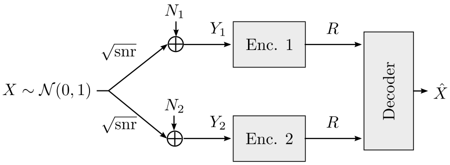
$$\Delta \leq \sum_{k \in \mathcal{S}} \left[R_k + \log |\mathbf{I} - \mathbf{B}_k| \right] + \log \left| \sum_{k \in \mathcal{S}^c} \bar{\mathbf{H}}_k^H \mathbf{B}_k \bar{\mathbf{H}}_k + \mathbf{I} \right|$$

for some $\mathbf{0} \preceq \mathbf{B}_k \preceq \mathbf{I}$, where $\bar{\mathbf{H}}_k = \mathbf{\Sigma}_{\mathbf{n}_k}^{-1/2} \mathbf{H}_k \mathbf{\Sigma}_{\mathbf{x}}^{1/2}$, and achievable with

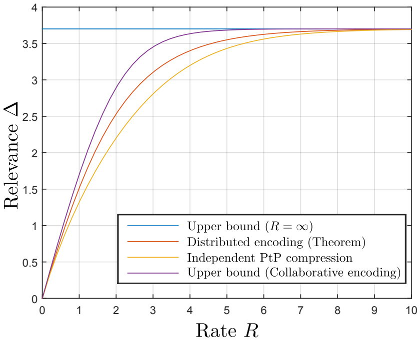
$$p^*(\mathbf{u}_k|\mathbf{y}_k, q) = \mathcal{CN}(\mathbf{y}_k, \Sigma_{\mathbf{n}_k}^{1/2}(\mathbf{B}_k - \mathbf{I})\mathbf{\Sigma}_{\mathbf{n}_k}^{1/2})$$

- Reminiscent of the sum-capacity in Gaussian Oblivious CRAN with Constant Gaussian Input constraint.

Example



$Y_k = \sqrt{\mathrm{snr}} X + N_k$, with $N_k \sim \mathcal{N}(0,1)$, $k = 1, 2$



- Optimal information (relevance):

$$\Delta^*(R, \mathrm{snr}) = \log_2 \left(1 + 2\,\mathrm{snr}\,2^{-2R}\left(2^{2R} + \mathrm{snr} - \sqrt{\mathrm{snr}^2 + (1 + 2\,\mathrm{snr})\,2^{2R}}\right)\right)$$

- Collaborative encoding upper bound: (Y_1, Y_2) encoded at rate $2R$

$$\Delta_{\mathrm{ub}}(R, \mathrm{sr}) = \log_2(1 + 2\,\mathrm{snr}) - \log_2(1 + 2\,\mathrm{snr}\,2^{-2R})$$

- Lower bound: Y_1 and Y_2 independently encoded

$$\Delta_i(R, \mathrm{snr}) = \log_2(1 + 2\,\mathrm{snr} - \mathrm{snr}\,2^{-R}) - \log_2(1 + \mathrm{snr}\,2^{-R})$$

The Distributed Information Bottleneck for Learning

- For simplicity, we look at the D-IB under sum-rate *[Estella-Zaidi'18]*

$$P^*_{U_k|Y_k} = \arg \min_{P_{U_k|Y_k}} I(X;U_{\mathcal{K}}) + \beta \sum_{k=1}^{K} [I(Y_k;U_k) - I(X;U_k)]$$

- The optimal encoders-decoder of the D-IB under sum-rate constraint satisfy the following self consistent equations,

$$p(u_k|y_k) = \frac{p(u_k)}{Z(\beta,u_k)} \exp\left(-\psi_s(u_k,y_k)\right),$$
$$p(x|u_k) = \sum_{y_k \in \mathcal{Y}_k} p(y_k|u_k)p(x|y_k)$$
$$p(x|u_1,\ldots,u_K) = \sum_{y_{\mathcal{K}} \in \mathcal{Y}_{\mathcal{K}}} p(y_{\mathcal{K}})p(u_{\mathcal{K}}|y_{\mathcal{K}})p(x|y_{\mathcal{K}})/p(u_{\mathcal{K}})$$

where

$$\psi_s(u_k,y_k) := D_{\mathrm{KL}}(P_{X|y_k}||Q_{X|u_k}) + \frac{1}{s}\mathrm{E}_{U_{\mathcal{K}\setminus k}|y_k}[D_{\mathrm{KL}}(P_{X|U_{\mathcal{K}\setminus k},y_k}||Q_{X|U_{\mathcal{K}\setminus k},u_k}))].$$

- Alternating iterations of these equations converge to a a solution for any initial $p(u_k|x_k)$, similarly to a Blahut-Arimoto algorithm.

D-IB for Vector Gaussian Sources: Iterative Optimization

- $(\mathbf{Y}_1, \cdots, \mathbf{Y}_K, \mathbf{X})$ jointly Gaussian, $\mathbf{Y}_k \in \mathbb{R}^N$ and $\mathbf{X} \in \mathbb{R}^M$,

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{N}_k, \qquad \mathbf{N}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- Optimal encoding $P^*_{U_k|Y_k}$ is Gaussian *[Estella-Zaidi'17]* and given by

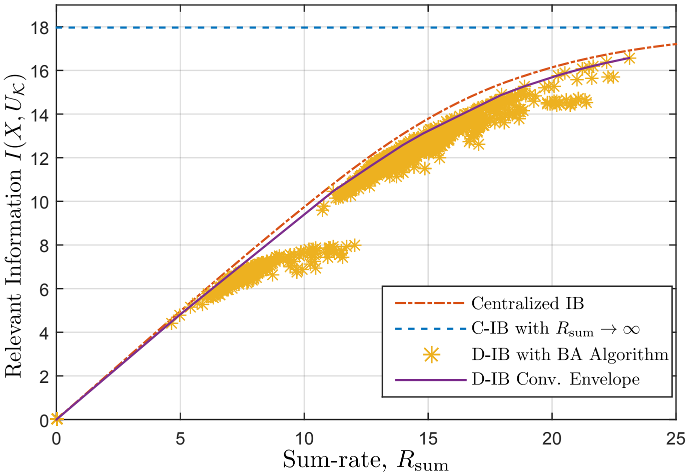
$$\mathbf{U}_k = \mathbf{A}_k \mathbf{Y}_k + \mathbf{Z}_k, \qquad \mathbf{Z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{z,k})$$

- For this class of distributions, the updates in the Blahut-Arimoto type algorithm simplify to:

$$\mathbf{\Sigma}_{\mathbf{z}_k^{t+1}} = \left(\left(1 + \frac{1}{\beta}\right) \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{x}} - \frac{1}{s} \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{u}^t_{\mathcal{K}\setminus k}} \right)^{-1},$$

$$\mathbf{A}_k^{t+1} = \mathbf{\Sigma}^{-1}_{\mathbf{z}_k^{t+1}} \left(\left(1 + \frac{1}{\beta}\right) \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{x}} \mathbf{A}_k^t (\mathbf{I} - \mathbf{\Sigma}_{\mathbf{y}_k|\mathbf{x}} \mathbf{\Sigma}^{-1}_{\mathbf{y}_k}) - \frac{1}{\beta} \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{u}^t_{\mathcal{K}\setminus k}} \mathbf{A}_k^t (\mathbf{I} - \mathbf{\Sigma}_{\mathbf{y}_k|\mathbf{u}^t_{\mathcal{K}\setminus k}} \mathbf{\Sigma}^{-1}_{\mathbf{y}_k}) \right).$$

D-IB for Vector Gaussian Sources (cont'd)



- Performance of distributed-IB is close to that of centralized IB

Wrap Up

- We have studied transmission over a CRAN under oblivious processing constraints at the relays and enabled resource-sharing.
    - i.e., relays are not allowed to know or acquire the users' codebooks.
- Our results shed light on the optimal relay operations:
    - NNC and CF with JDD optimal when the outputs at the relay nodes are conditionally independent on the users inputs.
    - Computed the Capacity Region under Gaussian Inputs in MIMO CRAN.
- Oblivious processing relevant from a practical viewpoint:
    - Bounded rate loss in comparison with the non-oblivious setting.
- Discussed relevant connections with CEO under logarithmic loss and Information Bottleneck Method.

Short Outlook

- Duality issues:
  - Downlink/uplink, e.g., Compute-forward v.s. reverse Compute-forward
  - Gaussian MAC/BC duality extends also for finite-capacity fronthauls $\{C_k\}$

    - See, e.g., Liu-Patil-Yu, *"An Uplink-Downlink Duality for Cloud Radio Access Network"*, ISIT'2016. More advanced downlink: Multi-Marton Coding: [Patil-Yu, 1801.00394]. Also "Channel Diagonalization for Cloud Radio Access", [Liu-Patil-Yu, arXiv:1802.01807]
  - Duality aspects via information bottleneck interpretations.
- Optimal input distributions under rate-constrained compression at relays.
  - Discrete signaling is already known to sometimes outperform Gaussian signaling for single-user Gaussian CRAN *[Sanderovich et al. '08]*.
  - It is conjectured that the optimal input distribution is discrete.
  - Improved upper bounds (over cut-set) for non-oblivious relay based schemes, to better evaluate the cost of oblivious processing (á la: Vu-Barnes-Ozgur, arXiv:1701.02043 Gaussian primitive relay).

Short Outlook cont.'

- Bounds on general information bottleneck problems [Painsky-Tishby, `arXiv:1711.02421`], [Eswaran-Gastpar, `arXiv:1805.06515`].
- A variety of related C-RAN & Distributed bottleneck problems:
  - Impact of block length n [C may not scale linearly with $n \Rightarrow$ Courtade conjecture $(C = 1)$] [Courtade-Kumar, IT'14], [Yang-Wesel, arXiv:1807.11289, July'18], The $C = n - 1$ case [Huleihel-Ordentlich, arXiv:1701.03119v2, May'17].
  - Bandlimited time-continuous models (Homri-Peleg-Shamai, arXiv:1510.08202).
  - Multi-layer Information Bottleneck Problem (Yang-Piantanida-Gündüz, arXiv:1711.05102).
  - Distributed Information-Theoretic Clustering (Pichler-Piantanida-Matz, arXiv:1602.04605, Dictator Functions, arXiv:1604.02109).

Short Outlook cont.'

- Entropy constraint bottleneck:

  $X - Y - U$
  $\max I(X;U)$ under the constraint $H(U) \leq C$
  practical applications: LZ distortionless compression.

  $\Rightarrow U = f(y)$ a deterministic function [Homri-Peleg-Shamai, Oblivious Processing in a Fronthaul Constrained Gaussian Channel, `arXiv:1510.08202`].

  - The deterministic bottleneck: advantages in complexity as compared to a classical bottleneck: [Strouse-Schwab, arXiv:1604.00268].

Some Related Tutorials

- 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.
- 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.
- Yuhan Zhou ; Yinfei Xu ; Wei Yu ; Jun Chen, “On the Optimal Fronthaul Compression and Decoding Strategies for Uplink Cloud Radio Access Networks,”IEEE Transactions on Information Theory, vol. 62, no. 12, Dec. 2016.
- 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 Pozna#, Poland.
- 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.
- 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.

Thank you!