Cooperative MIMO Precoding with Distributed CSI: A Hierarchical Approach

Italo Atzeni David Gesbert

EURECOM
Sophia Antipolis
Communication Systems Department

italo.atzeni@eurecom.fr

SPAWC 2018 - Kalamata, Greece

Introduction

- Network MIMO System: distributed TXs sharing user data symbols and channel state information (CSI) cooperatively serve several RXs $\longrightarrow$ **cooperative precoding design**.
- Distributed CSI (D-CSI): CSI is known **imperfectly** and **differently** across the TXs due to limited and uneven feedback.
- Team decision problem: multiple decentralized decision makers aim at coordinating their strategies while not being able to accurately predict the actions taken by the others.
- Hierarchical D-CSI: enforced by a suitable information exchange mechanism between TXs at a certain signaling/power cost $\longrightarrow$ can be leveraged to yield implementable and efficient distributed precoding solutions.

Downlink Network MIMO System Model

- N TXs, nth TX equipped with M_n antennas ($M \triangleq \sum_{n=1}^{N} M_n$), K single-antenna RXs.
- Channels: $\mathbf{h}_{kn} \in \mathbb{C}^{M_n \times 1}$ between TX n and RX k, $\mathbf{h}_k \triangleq [\mathbf{h}_{k1}^{\mathrm{T}} \ldots \mathbf{h}_{kN}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M \times 1}$ between the N TXs and RX k, and $\mathbf{H} \triangleq [\mathbf{h}_1 \ldots \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ between the N TXs and the K RXs, with $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_k)$.
- Multiuser precoding matrix:

$$\mathbf{W} \triangleq [\mathbf{w}_1 \ldots \mathbf{w}_K] = \begin{bmatrix} \mathbf{W}^{(1)} \\ \vdots \\ \mathbf{W}^{(N)} \end{bmatrix}$$

with $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ beamforming vector used by the N TXs to serve RX k and $\mathbf{W}^{(n)} \in \mathbb{C}^{M_n \times K}$ precoding submatrix used by TX n to serve the K RXs.

- Receive signal at RX k: $y_k \triangleq \mathbf{h}_k^{\mathrm{H}} \mathbf{x} + z_k$, with $\mathbf{x} \triangleq \mathbf{W}\mathbf{s} \in \mathbb{C}^{M \times 1}$ obtained by precoding the user data symbol vector $\mathbf{s} \in \mathbb{C}^{K \times 1}$ and $z_k \sim \mathcal{CN}(0, \sigma^2)$ noise at RX k.
- Sum rate:

$$R(\mathbf{H}, \mathbf{W}) \triangleq \sum_{k=1}^{K} \log_2 \left(1 + \frac{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_j|^2 + \sigma^2} \right).$$

Regularized Zero Forcing Precoding

Regularized zero forcing (RZF) precoding is adopted at each TX:

$$\mathbf{W}_{\mathrm{rzf}}^{(n)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}) \triangleq \sqrt{P_n} \times \frac{\boldsymbol{\Delta}_n^{\mathrm{T}} \hat{\mathbf{H}}^{(n)} \big((1-\alpha^{(n)}) (\hat{\mathbf{H}}^{(n)})^{\mathrm{H}} \hat{\mathbf{H}}^{(n)} + \alpha^{(n)} \mathbf{I}_K \big)^{-1}}{\big\| \boldsymbol{\Delta}_n^{\mathrm{T}} \hat{\mathbf{H}}^{(n)} \big((1-\alpha^{(n)}) (\hat{\mathbf{H}}^{(n)})^{\mathrm{H}} \hat{\mathbf{H}}^{(n)} + \alpha^{(n)} \mathbf{I}_K \big)^{-1} \big\|_{\mathrm{F}}}$$

with $\alpha^{(n)} \in [0,1]$ regularization factor and $\boldsymbol{\Delta}_n \triangleq [\mathbf{0}_{M_n \times \sum_{\ell=1}^{n-1} M_\ell} \ \mathbf{I}_{M_n} \ \mathbf{0}_{M_n \times \sum_{\ell=n+1}^{N} M_\ell}]^{\mathrm{T}} \in \mathbb{C}^{M \times M_n}$ block selection matrix.

Team Decision Precoding Problem

D-CSI:

$$\mathbf{W}_{\star}^{(n)} \triangleq \underset{\mathbf{W}^{(n)}}{\operatorname{argmax}} \, \mathbb{E}_{\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell \neq n} | \hat{\mathbf{H}}^{(n)}} \left[\max_{\{\mathbf{W}^{(\ell)}\}_{\ell \neq n}} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(n)}} \Big[R\Big(\mathbf{H}, \mathbf{W}^{(n)}(\hat{\mathbf{H}}^{(n)}), \{\mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)})\}_{\ell \neq n}\Big) \Big] \right]$$
$$\text{s.t.} \quad \big\|\mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)})\big\|_{\mathrm{F}}^2 \leq P_\ell, \qquad \ell = 1, \ldots, N$$

Hierarchical D-CSI:

$$\mathbf{W}_{\star,\mathrm{h}}^{(n)} \triangleq \underset{\mathbf{W}^{(n)}}{\operatorname{argmax}} \, \mathbb{E}_{\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=n+1}^{N} | \hat{\mathbf{H}}^{(n)}} \left[\max_{\{\mathbf{W}^{(\ell)}\}_{\ell=n+1}^{N}} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(n)}} \Big[R\Big(\mathbf{H}, \{\mathbf{W}_{\star,\mathrm{h}}^{(\ell)}\}_{\ell=1}^{n-1}, \mathbf{W}^{(n)}(\hat{\mathbf{H}}^{(n)}), \{\mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)})\}_{\ell=n+1}^{N}\Big) \Big] \right]$$
$$\text{s.t.} \quad \big\|\mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)})\big\|_{\mathrm{F}}^2 \leq P_\ell, \qquad \ell = n, \ldots, N$$

D-CSI Model

Each TX n has a different estimate of $\mathbf{H}$, denoted by $\hat{\mathbf{H}}^{(n)} \triangleq [\hat{\mathbf{h}}_1^{(n)} \ldots \hat{\mathbf{h}}_K^{(n)}] \in \mathbb{C}^{M \times K}$, given by

$$\hat{\mathbf{H}}^{(n)} = \sqrt{1-\epsilon_n^2}\,\mathbf{H} + \epsilon_n \mathbf{E}^{(n)}$$

where $\epsilon_n \in [0,1]$ and $\mathbf{E}^{(n)} \triangleq [\mathbf{e}_1^{(n)} \ldots \mathbf{e}_K^{(n)}]$, with $\mathbf{e}_k^{(n)} \sim \mathcal{CN}(0, \boldsymbol{\Upsilon}^{(n)})$, $\forall k = 1, \ldots, K$.

Cond. distributions of $\mathbf{H}|\hat{\mathbf{H}}^{(n)}$ and $\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell \neq n}|\hat{\mathbf{H}}^{(n)}$:

- $\mathbf{h}_k|\hat{\mathbf{h}}_k^{(n)} \sim \mathcal{CN}(\boldsymbol{\mu}_k^{(n)}, \boldsymbol{\Sigma}_k^{(n)})$, with

$$\boldsymbol{\mu}_k^{(n)} \triangleq \sqrt{1-\epsilon_n^2}\,\boldsymbol{\Sigma}_k \big((1-\epsilon_n^2) \boldsymbol{\Sigma}_k + \epsilon_n^2 \boldsymbol{\Upsilon}^{(n)} \big)^{-1} \hat{\mathbf{h}}_k^{(n)},$$
$$\boldsymbol{\Sigma}_k^{(n)} \triangleq \boldsymbol{\Sigma}_k - (1-\epsilon_n^2) \boldsymbol{\Sigma}_k \big((1-\epsilon_n^2) \boldsymbol{\Sigma}_k + \epsilon_n^2 \boldsymbol{\Upsilon}^{(n)} \big)^{-1} \boldsymbol{\Sigma}_k.$$

- $\hat{\mathbf{h}}_k^{(\ell)}|\hat{\mathbf{h}}_k^{(n)} \sim \mathcal{CN}(\boldsymbol{\mu}_k^{(\ell|n)}, \boldsymbol{\Sigma}_k^{(\ell|n)})$, with

$$\boldsymbol{\mu}_k^{(\ell|n)} \triangleq \sqrt{1-\epsilon_\ell^2}\,\boldsymbol{\mu}_k^{(n)},$$
$$\boldsymbol{\Sigma}_k^{(\ell|n)} \triangleq (1-\epsilon_\ell^2) \boldsymbol{\Sigma}_k^{(n)} + \epsilon_\ell^2 \boldsymbol{\Upsilon}^{(\ell)}.$$

Algorithms for the Case of 2 TXs

- Optimal approach:

$$\alpha_{\star,\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\operatorname{argmax}} \, \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(1)}} \left[\max_{\alpha^{(2)} \in [0,1]} R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big) \right],$$
$$\alpha_{\star,\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\operatorname{argmax}} \, R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\star,\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big).$$

- Naive approach: local CSI is assumed **perfect** and **shared** by more informed TXs

$$\alpha_{\mathsf{NA},\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\operatorname{argmax}} \, R\big(\hat{\mathbf{H}}^{(1)}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})\big),$$
$$\alpha_{\mathsf{NA},\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\operatorname{argmax}} \, R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\mathsf{NA},\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big).$$

- Locally robust approach: local CSI is assumed **imperfect** and **shared** by more informed TXs

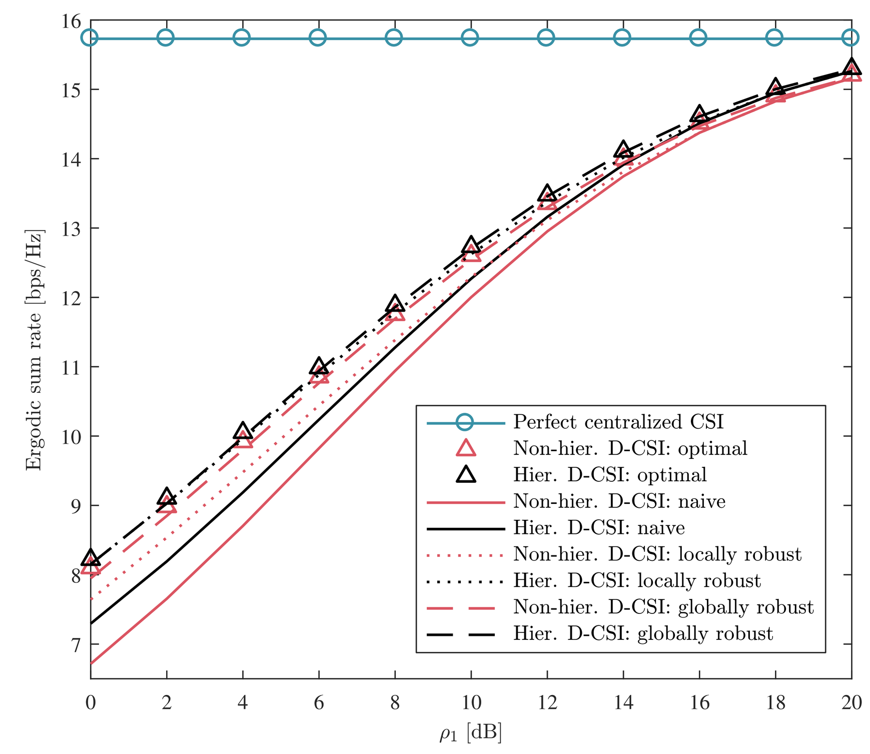
$$\alpha_{\mathsf{LR},\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\operatorname{argmax}} \, \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(1)}} \Big[R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})\big) \Big],$$
$$\alpha_{\mathsf{LR},\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\operatorname{argmax}} \, R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\mathsf{LR},\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big).$$

- Globally robust approach: local CSI is assumed **imperfect** and **not shared** by more informed TXs (neglecting the possibly different regularization factors adopted by the latter)

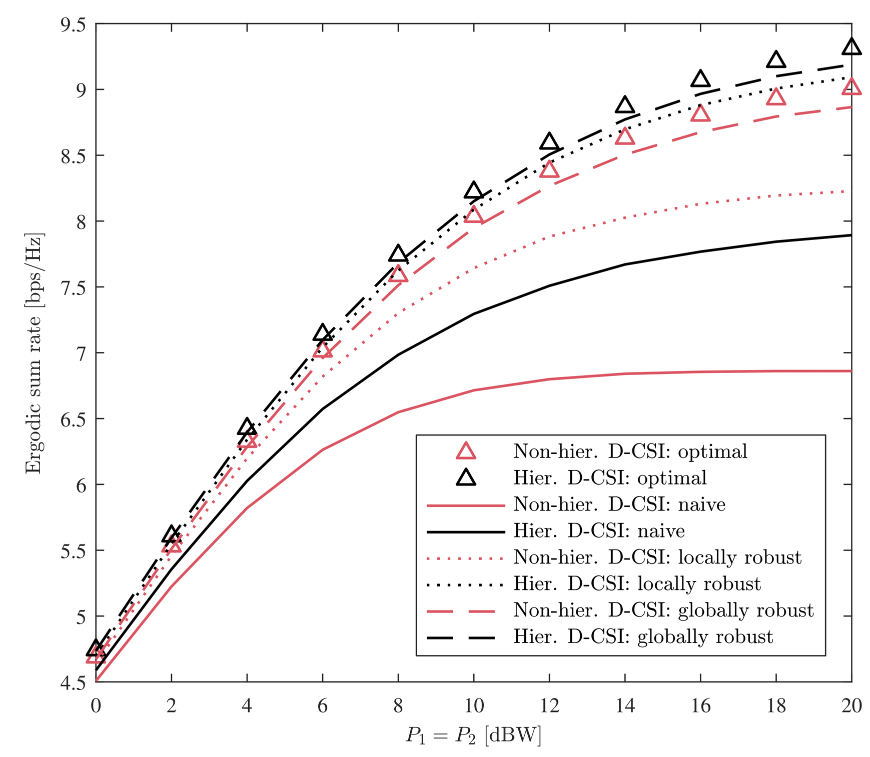
$$\alpha_{\mathsf{GR},\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\operatorname{argmax}} \, \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(1)}} \Big[R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(1)})\big) \Big],$$
$$\alpha_{\mathsf{GR},\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\operatorname{argmax}} \, R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\mathsf{GR},\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big).$$

Numerical Results

- $N = 2$ TXs (with 4 transmit antennas) facing each other at distance $d = 40$ m, $K = 5$ angularly equispaced RXs in $[\pi/4, 3\pi/4]$ between the TXs; ULAs at the TXs (uniform distribution of the AoDs with angle spread $\Delta\theta = \pi/8$); error covariance matrices $\{\boldsymbol{\Upsilon}^{(n)} = \mathbf{I}\}_{n=1}^{N}$, pathloss exponent $\eta = 2$, noise power $\sigma^2 = 0$ dBm.



Ergodic sum rate VS feedback SNR of TX 1 $\rho_1 \triangleq (1-\epsilon_1^2)/\epsilon_1^2$, with $P_1 = P_2 = 10$ dBW.



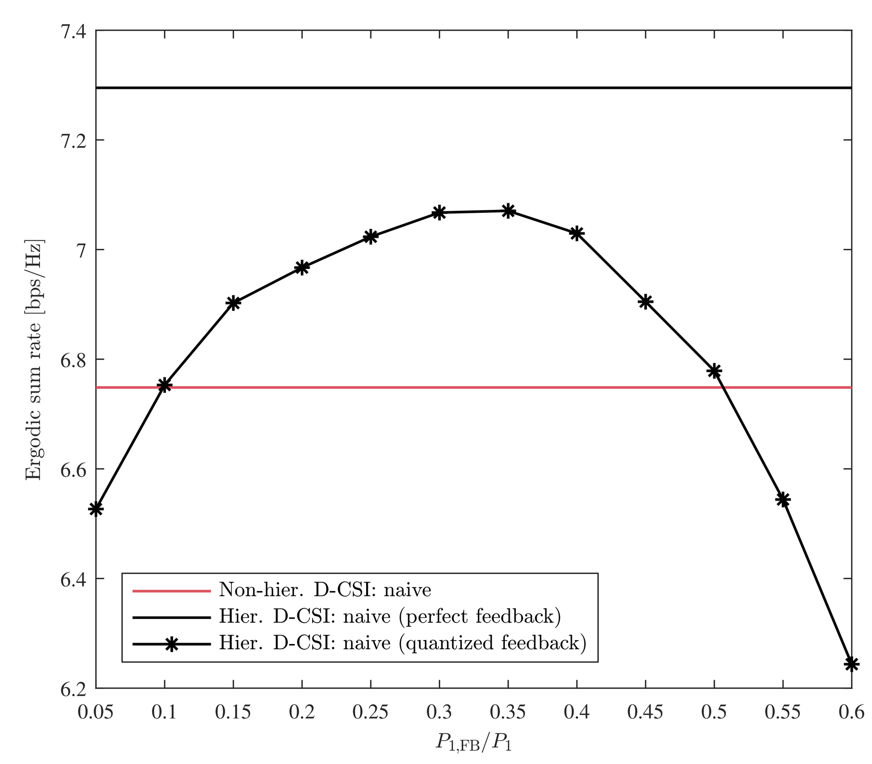
Ergodic sum rate VS per-TX power constraint, with $\rho_1 = 0$ dB.

- Information exchange VS cooperation gain: Feedback from TX 1 to TX 2, with $P_1 = P_{1,\mathrm{fb}} + P_{1,\mathrm{tx}}$ and number of feedback bits

$$\xi \triangleq \left\lfloor BT \log_2 \left(1 + d^{-\eta} \frac{P_{1,\mathrm{fb}}}{\sigma^2} \right) \right\rfloor.$$

Given the common codebook $\mathcal{W} \triangleq \{\hat{\mathbf{W}}_q^{(1)}\}_{q=1}^{2^\xi}$, TX 1 computes $\mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})$ and sends the index

$$\hat{q} \triangleq \underset{q \in [1, 2^\xi]}{\operatorname{argmin}} \, \|\hat{\mathbf{W}}_q^{(1)} - \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})\|_{\mathrm{F}}.$$



Ergodic sum rate VS feedback power, with $P_1 = P_2 = 10$ dBW, $P_{1,\mathrm{tx}} = 10\log_{10}(P_1 - P_{1,\mathrm{fb}})$ dBW, $\rho_1 = 0$ dB, $B = 1$ kHz, and $T = 5$ ms.