Rate-Splitting for Multi-User Multi-Antenna Systems: Bridging the Extremes

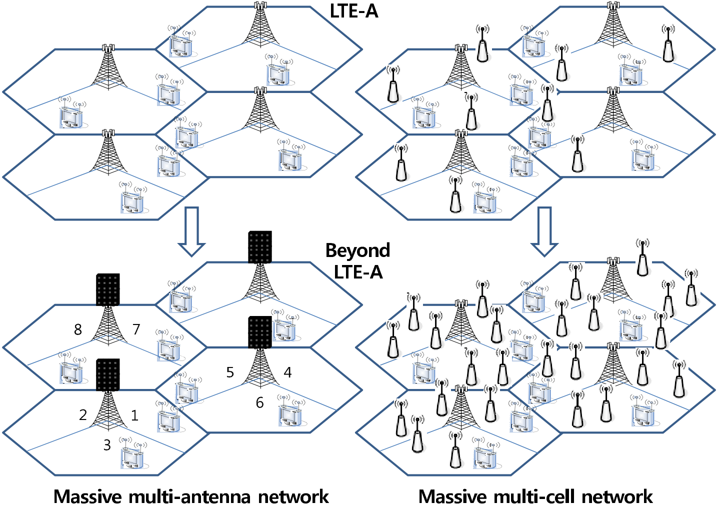
Bruno Clerckx

Dept. of Electrical and Electronic Engineering
Imperial College London

Queen Mary University London, Nov 2017

Contents

MIMO Networks: Single-user, Multi-user, Multi-cell, Massive, Network, Cooperative, Coordinated,...



Motivation 1 for a New Physical Layer

- Big loss as the CSIT accuracy decreases.
- High CSIT accuracy has become increasingly difficult to satisfy
  - Dense HetNet, Massive MIMO
- So far, techniques designed for perfect CSIT applied to imperfect CSIT scenarios.
- Imperfect CSIT hardly avoidable.
- Wiser to design wireless networks from scratch accounting for imperfect CSIT?

Motivation 1 for a New Physical Layer

Information theoretic channel (e.g. MISO BC)

$\Downarrow$

Information theoretic limits (Capacity region)

$\Downarrow$

Communication scheme (e.g. DPC)

$\Downarrow$

Suboptimal scheme (Linear precoding)

$\Downarrow$

Signal processing (Precoder optimization)

$\Downarrow$

Imperfect CSIT (Robust optimization)

For example, robust optimization of $\mathbf{p}_1, \ldots, \mathbf{p}_K$ in

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{p}_k s_k.$$

BUT !!! The design is motivated by perfect CSIT to start with.

A Bottom-up Approach

Information theoretic channel (e.g. MISO BC with Imperfect CSIT)

$\Downarrow$

Information theoretic limits (Capacity region - unkown)

$\Downarrow$

Alternative information theoretic limits (DoF region)

$\Downarrow$

Communication scheme (Based on Rate-Splitting)

$\Downarrow$

Suboptimal scheme (Linear precoding)

$\Downarrow$

Signal processing (Precoder optimization)

For example, optimizing $\mathbf{p}_c, \mathbf{p}_1, \ldots, \mathbf{p}_K$ in

$$\mathbf{x} = \mathbf{p}_c s_c + \sum_{k=1}^{K} \mathbf{p}_k s_k$$

where $\mathbf{p}_c s_c$ comes from Rate-Splitting.
Motivated by optimality in a DoF sense (multiplexing gain)

Motivation 2 for a New Physical Layer

- MIMO networks rely on two extreme interference management strategies: **fully decode interference** and **treat interference as noise**
  - NOMA based on superposition coding with successive interference cancellation relies on strong users to fully decode and cancel interference created by weaker users
  - MU-MIMO, CoMP, Massive MIMO, millimetre wave MIMO based on linear precoding rely on fully treating any multi-user interference as noise
- Rate-Splitting as a **more general** and **more powerful** transmission framework: **partially** decode interference and **partially** treat interference as noise
  - Softly bridge and therefore reconcile the two extreme strategies
  - RS encompasses NOMA and MU-MIMO with linear precoding as special cases

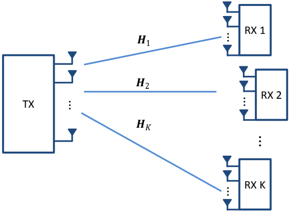
$$\mathbf{x} = \mathbf{p}_{\mathrm{c}} s_{\mathrm{c}} + \sum_{k=1}^{2} \mathbf{p}_k s_k$$

where $\mathbf{p}_{\mathrm{c}} s_{\mathrm{c}}$ comes from Rate-Splitting.

The MISO Broadcast Channel and Partial CSIT

1. Limitations of Current 4G and Emerging 5G Architecture
2. The MISO Broadcast Channel and Partial CSIT
    - System model
    - Perfect CSIT
    - Imperfect CSIT
3. Fundamentals of Rate Splitting
4. Precoder Optimization
5. Extensions of Rate-Splitting
6. Conclusions and Future Challenges

System model



$$y_k(t) = \mathbf{h}_k^H(t)\mathbf{x}(t) + n_k(t)$$

- M transmit antennas and K single-antenna users ($M \geq K$).
- Channel state (matrix): $\mathbf{H}(t) = [\mathbf{h}_1(t), \ldots, \mathbf{h}_K(t)]$.
- In each t, transmitter obtains the estimate $\widehat{\mathbf{H}}(t)$ (i.e. CSIT).

System model: Transmission and Linear precoding

Linear precoding signal model:

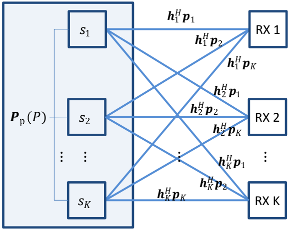
- Independent symbol streams: $W_1, \ldots, W_K \mapsto s_1, \ldots, s_K$.
- t is dropped for simplicity.
- Unity average power: $\mathbb{E}\{s_i s_k^*\} = 1$ if $i = k$, and 0 if $i \neq k$.
- Linear Precoding:

$$\mathbf{x} = \mathbf{p}_1 s_1 + \ldots + \mathbf{p}_K s_K.$$

- Average power constraint: $\sum_{k=1}^{K} \|\mathbf{p}_k\|^2 \leq P$.
- $\mathbf{P}_{\mathrm{p}} = [\mathbf{p}_1, \ldots, \mathbf{p}_K]$ can be adapted based on CSIT

$$\mathbf{P}_{\mathrm{p}}\big(\widehat{\mathbf{H}}(1)\big), \mathbf{P}_{\mathrm{p}}\big(\widehat{\mathbf{H}}(2)\big), \ldots, \mathbf{P}_{\mathrm{p}}\big(\widehat{\mathbf{H}}(T)\big).$$

System model: SINR and Rate

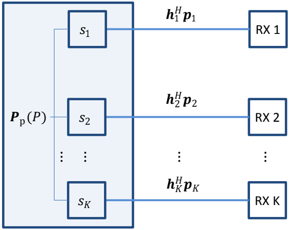


$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_k s_k}^{\text{desired signal}} + \overbrace{\mathbf{h}_k^H \sum_{i \neq k} \mathbf{p}_i s_i}^{\text{interference}} + \overbrace{n_k}^{\text{noise}}$$

- SINR (instantaneous): $\mathrm{SINR}_k = \frac{|\mathbf{h}_k^H \mathbf{p}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{p}_i|^2 + \sigma_{\mathrm{n}}^2}$.
- Rate (instantaneous): $R_k = \log_2\left(1 + \mathrm{SINR}_k\right)$.
- Ergodic Rate (for $T \gg 1$): $\mathbb{E}\{R_k\}$.

Perfect CSIT

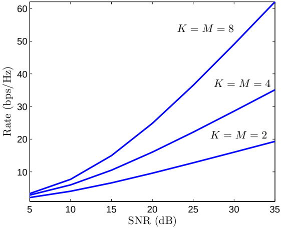
- Perfect CSIT: $\widehat{\mathbf{H}} = \mathbf{H}$.
- Zero-Forcing (ZF) precoding:
  - $\mathbf{P}_{\mathrm{p}} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{B}$ where $\mathbf{B}$ is diagonal.
  - This yields: $\mathbf{p}_k \in \mathrm{null}\left([\mathbf{h}_1, \ldots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \ldots, \mathbf{h}_K]^H\right)$.



$$y_k = \mathbf{h}_k^H \mathbf{p}_k s_k + n_k$$

- Each user receives an interference-free stream.
- In other words, each user gets one full DoF.

Perfect CSIT:Degrees of Freedom (DoF)



- DoF: fraction of an interference-free stream's capacity as $P \to \infty$.
- Considering the Ergodic rate:

$$d_k = \lim_{P \to \infty} \frac{\mathbb{E}\{R_k\}}{\log_2(P)}.$$

- For MISO, we have $d_k \leq 1$ due to single-antenna receivers.
- Under perfect CSIT, ZF and equal power allocation achieves full DoF:

$$\sum_{k=1}^{K} d_k = K.$$

Imperfect CSIT

What happens when CSIT is imperfect?

Imperfect CSIT model:

$$\mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}$$

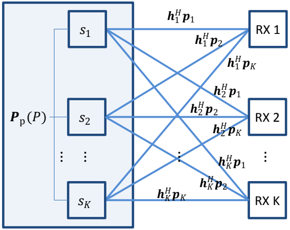
$$\mathbf{h}_k = \underbrace{\widehat{\mathbf{h}}_k}_{\text{estimate}} + \underbrace{\widetilde{\mathbf{h}}_k}_{\text{error}}$$

Estimate obtained through feedback or UL training [2].

- CSIT error power: $\mathbb{E}\left\{\|\widetilde{\mathbf{h}}_k\|^2\right\} = \sigma_{\mathrm{e},k}^2$.
- CSIT error scaling: $\alpha_k = \lim_{P\to\infty} -\frac{\log(\sigma_{\mathrm{e},k}^2)}{\log(P)}$
- It follows that: $\mathbb{E}\left\{\|\widetilde{\mathbf{h}}_k\|^2\right\} \sim P^{-\alpha_k}$.
- Assume: $\alpha_1, \ldots, \alpha_K = \alpha$.
  - $\alpha > 0$: CSIT improves with P (e.g. increasing number of feedback bit).
  - $\alpha = 0$: CSIT fixed with P (e.g. fixed number of feedback bit).
  - $\alpha = 1$: CSIT perfect in a DoF sense (as we see next).

Imperfect CSIT: Zero-Forcing

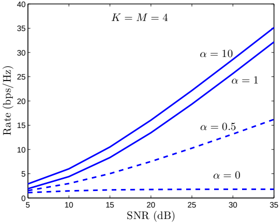
- ZF over the imperfect channel estimate:
    - $\mathbf{P}_{\mathrm{p}} = \widehat{\mathbf{H}}(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}})^{-1}\mathbf{B}$.
    - This yields: $\mathbf{p}_k \in \mathrm{null}\left(\left[\widehat{\mathbf{h}}_1, \ldots, \widehat{\mathbf{h}}_{k-1}, \widehat{\mathbf{h}}_{k+1}, \ldots, \widehat{\mathbf{h}}_K\right]^H\right)$.



$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_k s_k}^{\text{desired signal}} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{i\neq k} \mathbf{p}_i s_i}^{\text{residual interference}} + \overbrace{n_k}^{\text{noise}}$$

- Each user cannot enjoy an interference-free stream anymore.
- What happens to the DoF?

Imperfect CSIT: DoF loss



- ZF and equal power allocation: $\|\mathbf{p}_1\|^2 = \ldots = \|\mathbf{p}_K\|^2 = \frac{P}{K}$.

$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_k s_k}^{\text{desired signal } \sim P} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{i \neq k} \mathbf{p}_i s_i}^{\text{residual interference } \sim P^{1-\alpha}} + \overbrace{n_k}^{\text{noise} \sim P^0}$$

- Assume $\alpha \in [0, 1]$.
- $\mathrm{SINR}_k \sim P^\alpha$ from which $\mathbb{E}\{R_k\} = \log_2(P^\alpha) + O(1)$.
- $d_k = \alpha$ from which the sum DoF [1, 2]:

$$\sum_{k=1}^{K} d_k = K\alpha.$$

Imperfect CSIT: Interference

Perfect CSIT:

- Inter-user interference can be fully eliminated.
- Full DoF is achieved.

Partial CSIT with $\alpha \geq 1$:

- Inter-user interference can be reduced to the level of noise.
- No DoF loss.

Partial CSIT with $\alpha < 1$:

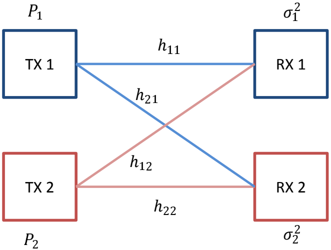
- Inter-user interference cannot be reduced to the level of noise.
- Treating interference as noise causes DoF loss.

If interference cannot be eliminated or reduced to noise level, why not decode it and remove it from the received signal (fully or in part)?

Let us first take a step back, and look at the 2-user Interference Channel (IC).

Fundamentals of Rate Splitting

Two-User Interference Channel (IC)



$$y_k = h_{k1}x_1 + h_{k2}x_2 + n_k$$

- Message W_k from TX-k to RX-k.
- Encoding: $W_k \mapsto x_k$.
- Decoding: $y_k \mapsto \widehat{W}_k$.

Symmetric setup:

- $|h_{11}|^2 = |h_{22}|^2 = |h_{\mathrm{d}}|^2$ and $|h_{12}|^2 = |h_{21}|^2 = |h_{\mathrm{c}}|^2$
- $P_1 = P_2 = P$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Two-User IC: Rate-Splitting

Weak interference $|h_{\mathrm{c}}|^2 < |h_{\mathrm{d}}|^2$ **(or general case)**:

- Not strong enough to **decode**, or weak enough to **treat as noise**.
- **Rate-Splitting**: part **decoded** by other and part **treated as noise**.
  - Split messages: $W_k \mapsto W_{k0}, W_{k1} \mapsto x_{k0}, x_{k1}$.
  - Split power: $P_k \mapsto P_{k0}, P_{k1}$.
  - RX-1 decodes x_{20} and x_1 (composed of x_{10}, x_{11}).
  - RX-2 decodes x_{10} and x_2 (composed of x_{20}, x_{21}).
- Reduces to **treat as noise** when $P_{10} = P_{20} = 0$.
  - i.e. $|W_{10}| = |W_{20}| = 0$.
  - $W_k \mapsto x_{k1}$.
- Reduces to **decode** interference when $P_{11} = P_{21} = 0$.
  - i.e. $|W_{11}| = |W_{21}| = 0$.
  - $W_k \mapsto x_{k0}$.
- **Bridges** the two in general [3].

The MISO-BC with imperfect CSIT revisited

Rate-Splitting for MISO-BC[4]:

- The general idea is to split messages.
- One part decoded by all, while the other treated as noise.

But!

- In what proportion are messages split?
- How much power to allocate?
- How to transmit each part?

Strategy:

- Private messages:
  - Parts which are treated as noise.
  - Received at the level of noise
- Common message(s):
  - Parts which are decoded by all.
  - Transmitted in a public manner.

MISO-BC: Parts to treat as noise (private messages)

Interference reduction through power control:

- Reduce allocated power to P^α.
- Note that $P^\alpha \leq P$ for $\alpha \in [0,1]$.
- Equal power allocation: $\|\mathbf{p}_1\|^2 = \ldots = \|\mathbf{p}_K\|^2 = \frac{P^\alpha}{K}$.

$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_k s_k}^{\text{desired signal } \sim P^\alpha} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{i \neq k} \mathbf{p}_i s_i}^{\text{residual interference } \sim P^{\alpha-\alpha} = P^0} + \overbrace{n_k}^{\text{noise} \sim P^0}$$

- Interference is reduced to noise level $\sim P^0$.
- This also limits desired power $\sim P^\alpha$.
- DoF is maintained: $d_k = \alpha$ and $\sum_{k=1}^K d_k = K\alpha$.
- Only power levels (scalings) from 0 to α are occupied.
- The remaining power levels (α to 1) are freed for the other parts.

MISO-BC: Parts to decode (common message)

Superpose $W_c \mapsto s_c$ (with precoder $\mathbf{p}_c$) to be decoded by all users.

$$\mathbf{x} = \mathbf{p}_c s_c + \sum_{k=1}^{K} \mathbf{p}_k s_k$$

where $\|\mathbf{p}_c\|^2 = P - P^\alpha \sim P$ and $\|\mathbf{p}_1\|^2 = \ldots = \|\mathbf{p}_K\|^2 = \frac{P^\alpha}{K} \sim P^\alpha$

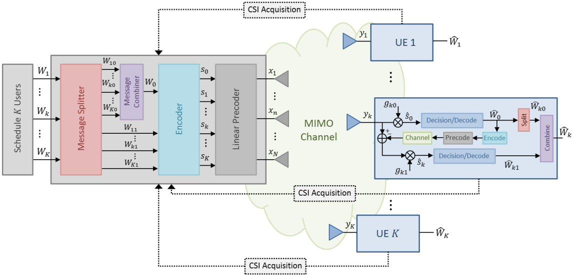
$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_c s_c}^{\sim P} + \overbrace{\mathbf{h}_k^H \mathbf{p}_k s_k}^{\sim P^\alpha} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{i \neq k} \mathbf{p}_i s_i}^{\sim P^0} + \overbrace{n_k}^{\sim P^0}$$

- $\mathrm{SINR}_{c,k} \sim P^{1-\alpha}$ from which $\mathbb{E}\{R_{c,k}\} = \log_2(P^{1-\alpha}) + O(1)$.
- DoF of common message: $d_c = 1 - \alpha$.
- SIC is used to remove s_c, as it is decoded by all.
- DoF of private messages is maintained: $d_k = \alpha$.
- Sum DoF is boosted: $d_c + \sum_{k=1}^{K} d_k = (1-\alpha) + K\alpha$ [11].

What remains is to load both parts (private and common) with user data.

MISO-BC: Rate-Splitting

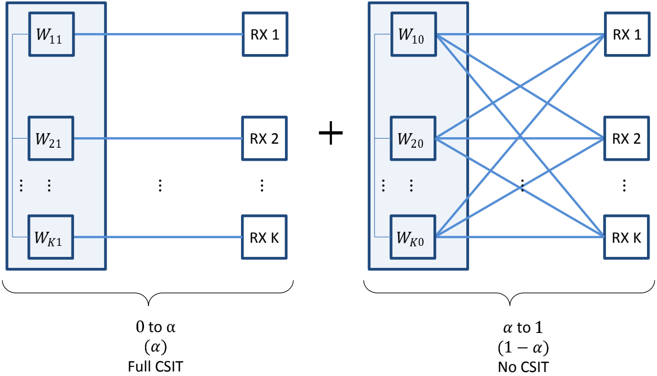
Instead of a new common message, s_c is loaded with part of user messages.



- Split message of user-1 : $W_1 \mapsto W_{10}, W_{11}$.
- Common part: $W_{10} \mapsto s_c$, decoded by all users but intended to users-1.
- Private part: $W_{11} \mapsto s_1$ decoded by user-1.
- $W_2, \dots, W_K \mapsto s_2, \dots, s_K$ decoded by corresponding users.

Splitting can be done for other (or all) users as in figure [25].

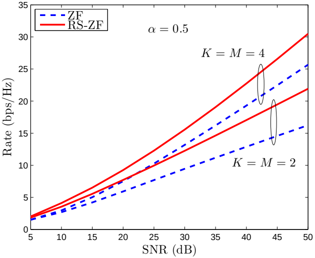
MISO-BC: Weighted sum interpretation



Decomposed into a weighted superposition of two networks [19]

- Perfect CSIT.
  - Achieves sum DoF of K.
  - Weighted by α.
- No CSIT
  - Achieves sum DoF of 1.
  - Weighted by $1-\alpha$.

MISO-BC: DoF with RS



Proposition

In the K user MISO-BC with partial CSIT, sum DoF achieved by ZF is given by

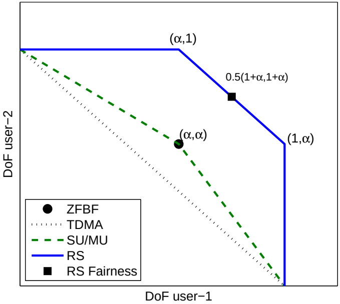
$$d_{\Sigma}^{\mathrm{ZF}} = K\alpha$$

while the sum DoF achieved by RS-ZF is given by

$$d_{\Sigma}^{\mathrm{RS}} = 1 + (K-1)\alpha.$$

Optimality of RS to achieve the entire DoF region of the K-user MISO BC shown in [28]. Converse based on [5].

MISO-BC: Two-User DoF region

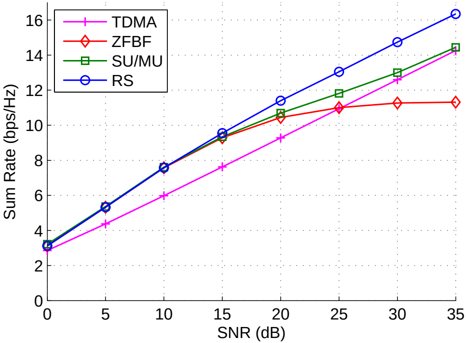


- Assume splitting for user-1
  - user-1 DoF: $d_c + d_1 = (1 - \alpha) + \alpha = 1$.
  - user-2 DoF: $d_2 = \alpha$.
- Time-sharing between splitting for user-1 and user-2.
- Compared to time-sharing between ZF and TDMA.

Sum-Rate enhancement and Feedback reduction

From DoF to rate analysis [6]:

- So far we have looked at the DoF gains of RS $(P \rightarrow \infty)$.
- Sum-rate enhancement (slope gain and/or SNR gain) over ZF, TDMA, switching between TDMA/ZF (SU/MU)[6].
- $M = 4$ antennas, $K = 2$ users, and $B = 15$ bits.



Precoder Optimization

1. Limitations of Current 4G and Emerging 5G Architecture
2. The MISO Broadcast Channel and Partial CSIT
3. Fundamentals of Rate Splitting
4. Precoder Optimization
   - Ergodic Sum-Rate Maximization
   - Robust Max-Min Fairness
5. Extensions of Rate-Splitting
6. Conclusions and Future Challenges

Precoder Optimization

Recall that the RS (linearly precoded) signal model is:

$$\mathbf{x} = \mathbf{p}_{\mathrm{c}} s_{\mathrm{c}} + \sum_{k=1}^{K} \mathbf{p}_k s_k$$

- Precoding matrix: $\mathbf{P} = [\mathbf{p}_{\mathrm{c}}, \mathbf{p}_1, \ldots, \mathbf{p}_K]$.
- Power constraint: $\mathrm{tr}\big(\mathbf{P}\mathbf{P}^H\big) \leq P$.
- So far we considered simple barely optimized designs (ZF, random).
- The choice of $\mathbf{P}$ influences $R_{\mathrm{c}}, R_1, \ldots, R_K$.

Challenges

- Transmitter only knows $\widehat{\mathbf{H}}$ and not $\mathbf{H}$.
- Instantaneous $R_{\mathrm{c}}, R_1, \ldots, R_K$ not known by the transmitter.
- Transmission should be carried out at reliable (decodable) rates.

Ergodic Sum-Rate Maximization

RS problem [11]:

$$\mathcal{R}_{\mathrm{RS}}(P): \begin{cases} \max\limits_{\bar{R}_c, \mathbf{P}} & \bar{R}_{\mathrm{c}} + \sum_{k=1}^{K} \bar{R}_k \\ \text{s.t.} & \bar{R}_{\mathrm{c},k} \geq \bar{R}_{\mathrm{c}}, \ \forall k \in \mathcal{K} \\ & \mathrm{tr}\big(\mathbf{P}\mathbf{P}^H\big) \leq P \end{cases}$$

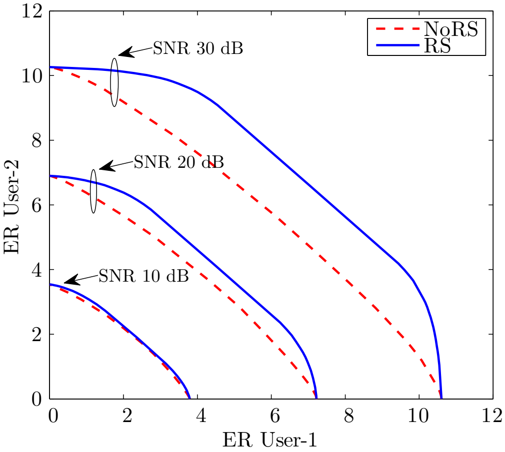
as opposed to the conventional (NoRS) formulation

$$\mathcal{R}(P): \begin{cases} \max\limits_{\mathbf{P}_{\mathrm{p}}} & \sum_{k=1}^{K} \bar{R}_k \\ \text{s.t.} & \mathrm{tr}\big(\mathbf{P}_{\mathrm{p}}\mathbf{P}_{\mathrm{p}}^H\big) \leq P. \end{cases}$$

- Stochastic optimization problem (due to expectations inside the ARs).
- Even a deterministic version is non-convex and very difficult.
- WMMSE approach can efficiently handle sum rate problems.

Ergodic Sum-Rate Maximization: Two-user ER region

- More generally, we can solve the Weighted ESR problem [11].



- Shows the ER trade-offs between the two users.

Robust Max-Min Fairness

Non-Ergodic transmission over $T = 1$ random state $\{\mathbf{H}, \widehat{\mathbf{H}}\}$.

- For kth user, CSIT errors bounded by sphere with radius δ_k:

$$\mathbb{H}_k = \left\{ \mathbf{h}_k \mid \mathbf{h}_k = \widehat{\mathbf{h}}_k + \widetilde{\mathbf{h}}_k, \|\widetilde{\mathbf{h}}_k\| \leq \delta_k \right\}$$

- For any $\mathbf{P}$, worst-case rates defined as:

$$\bar{R}_{\mathrm{c},k} = \min_{\mathbf{h}_k \in \mathbb{H}_k} R_{\mathrm{c},k}(\mathbf{h}_k) \quad \text{and} \quad \bar{R}_k = \min_{\mathbf{h}_k \in \mathbb{H}_k} R_k(\mathbf{h}_k).$$

- For given $\widehat{\mathbf{H}}$, transmission at worst-case rates is reliable (robust).

Rate-Splitting revisited [12]: Sharing the common message

- $W_k \mapsto W_{k0}, W_{k1}$ for all $k \in \{1, \ldots, K\}$.
- $W_{10}, \ldots, W_{K0} \mapsto s_{\mathrm{c}}$.
- $W_{11}, \ldots, W_{K1} \mapsto s_1, \ldots, s_K$.

Robust Max-Min Fairness

$$\mathcal{R}_{\mathrm{RS}}(P): \begin{cases} \max\limits_{\bar{\mathbf{c}},\mathbf{P}} & \min\limits_{k\in\mathcal{K}} (\bar{R}_k + \bar{C}_k) \\ \text{s.t.} & \bar{R}_{\mathrm{c},k} \geq \sum_{i=1}^{K} \bar{C}_i, \ \forall k \in \mathcal{K} \\ & \bar{C}_k \geq 0, \ \forall k \in \mathcal{K} \\ & \mathrm{tr}\big(\mathbf{P}\mathbf{P}^H\big) \leq P. \end{cases}$$

where $\bar{\mathbf{c}} = [\bar{C}_1, \ldots, \bar{C}_M]$.

- Portion of the common message rate given to user k: $\bar{C}_k$.
- Sum of all portions: $\sum_{k=1}^{K} \bar{C}_k = \bar{R}_{\mathrm{c}} = \min_i \bar{R}_{\mathrm{c},i}$.
- Rate of user k: $\bar{R}_k + \bar{C}_k$ (private and common portions).

Classical (NoRS) problem formulated as:

$$\mathcal{R}(P): \begin{cases} \max\limits_{\mathbf{P}_{\mathrm{p}}} & \min\limits_{k\in\mathcal{K}} \bar{R}_k \\ \text{s.t.} & \mathrm{tr}\big(\mathbf{P}_{\mathrm{p}}\mathbf{P}_{\mathrm{p}}^H\big) \leq P. \end{cases}$$

Robust Max-Min Fairness: Simulation results

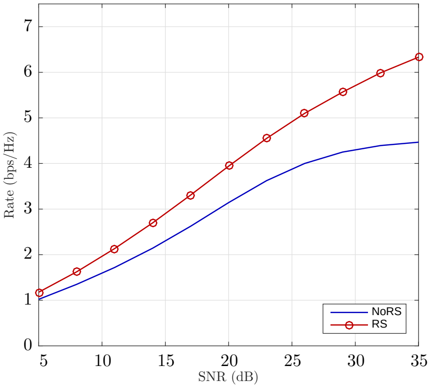


Figure: $K = M = 3$ and $\delta_1, \delta_2, \delta_3 = 0.1$.

- NoRS saturates due to non-scaling CSIT errors.
- RS avoids saturation and performs better across all SNRs [12].

Extensions of Rate-Splitting

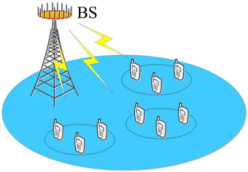
1. Limitations of Current 4G and Emerging 5G Architecture
2. The MISO Broadcast Channel and Partial CSIT
3. Fundamentals of Rate Splitting
4. Precoder Optimization
5. Extensions of Rate-Splitting
    - Massive MISO
    - Multi-Cell Coordination
    - Overloaded systems
    - Multigroup multicast beamforming
    - Multiuser Millimeter Wave Beamforming
    - Bridging NOMA and MU-MIMO
    - RF Impairments
6. Conclusions and Future Challenges

Massive MISO

Massive MIMO challenge: the huge demand for accurate CSIT.

The use of Rate-Splitting[10]:

- The constraint: $R_{\mathrm{c}} = \min_k \{R_{\mathrm{c},k}\}$.
- This highly reduces the gain when K is large.
- Channel statistics $\mathbf{R}_k$ can be further exploited.
- Large training and feedback overhead.



User grouping based on spatial correlation:

- Two-tier precoding [15, 16, 17]

$$\mathbf{x} = \sqrt{\frac{P}{K}} \sum_{g=1}^{G} \mathbf{B}_g \mathbf{W}_g \mathbf{s}_g,$$

- Users in g-th group share the same channel statistics: $\mathbf{R}_g$.
- $\mathbf{B}_g$: outer-precoding matrix based on channel statistics.
- $\mathbf{W}_g$: inner-precoding matrix designed based on short-term effective channel estimates: $\widehat{\overline{\mathbf{H}}}_g = \mathbf{B}_g^H \widehat{\mathbf{H}}_g$.

Massive MISO: Hierarchical Rate-Splitting (HRS)

- Overlap between the eigen-subspaces ⇒ inter-group interference.
- Imperfect CSIT ⇒ intra-group interference.

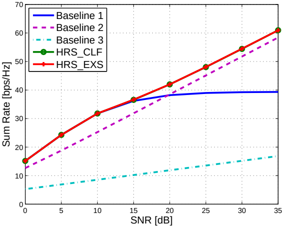
- **Hierarchical Rate-Splitting**[10]: a hierarchy of common messages to combat the inter-group and intra-group interference in Massive MIMO

$$\mathbf{x} = \overbrace{\sqrt{P_{\mathrm{sc}}}\mathbf{w}_{\mathrm{sc}}\, s_{\mathrm{sc}}}^{\text{system common msg.}} + \sum_{g=1}^{G} \mathbf{B}_g \left(\overbrace{\sqrt{P_{cg}}\mathbf{w}_{cg}\, s_{cg}}^{\text{group common msg.}} + \overbrace{\sqrt{P_{gk}}\mathbf{W}_g\, \mathbf{s}_g}^{\text{private msgs.}} \right)$$

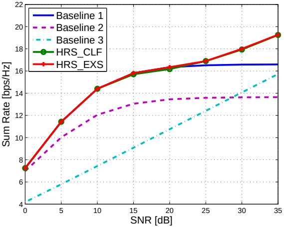
- System common msg. decoded by all users: for inter-group interference.
- Group common msg. decoded by group: for intra-group interference

Massive MISO: Simulation results

- HRS under imperfect CSIT, $M = 100$, $K = 12, \tau^2 = 0.4$



(a) disjoint eigen-subspace

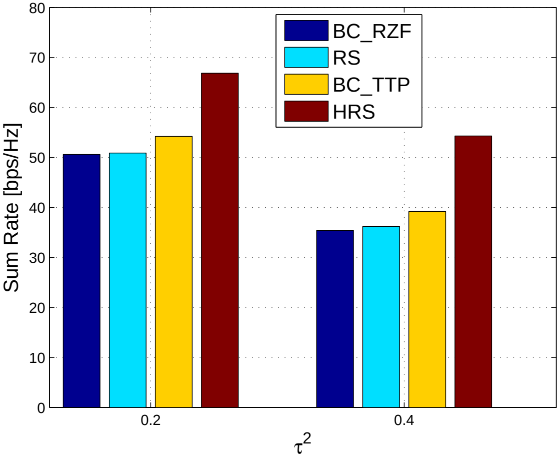


(b) overlapping eigen-subspace

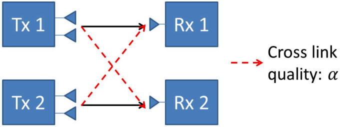
- HRS behaves as two-tier BC at low to medium SNR.
- HRS achieves a non-saturating sum rate.
- HRS decreases the complexity of precoder design and scheduling.
- HRS increases the complexity of the encoders and decoders.

Massive MISO: Simulation results

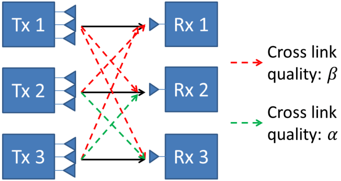
- $M = 100$, $K = 12, \tau^2 = 0.4$, $SNR = 30dB$, disjoint eigen-subspaces



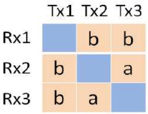
Multi-Cell Coordination: Topological Rate-Splitting (TRS)



(c) two-cell scenario [18]

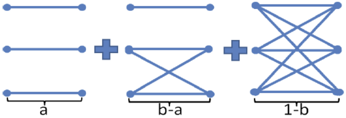


(d) three-cell scenario [19]



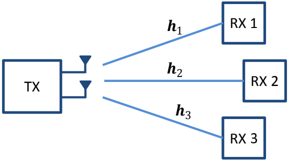
	Tx1	Tx2	Tx3
Rx1		b	b
Rx2	b		a
Rx3	b	a	

(e) CSIT pattern



(f) Weighted-sum interpretation [19]

Overloaded systems



- Overloaded scenarios: $K > M$.
- Scheduling over orthogonal resource blocks (time/frequency).
- Serve at most M users at a time.
- Reduces to conventional MISO BC in each block.
- With perfect CSIT, achieves DoF M in each block.

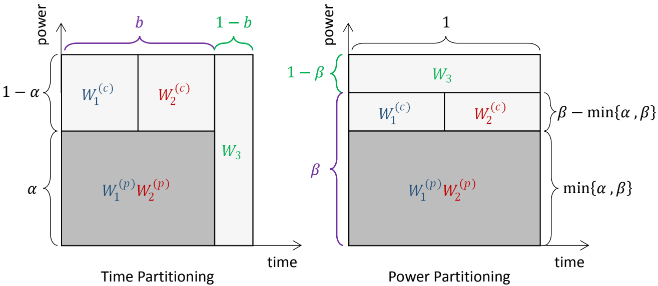
Consider a scenario where some user have little or no CSIT:

- IoT with many devices.
- Low-power sensor-like receivers.
- Can be served using the common message in the RS scheme [20].

Overloaded systems: Three-User example

- System: $M = 2$ antennas and $K = 3$ users.
- CSIT: $\alpha_1 = \alpha_2 = \alpha$ and $\alpha_3 = 0$.

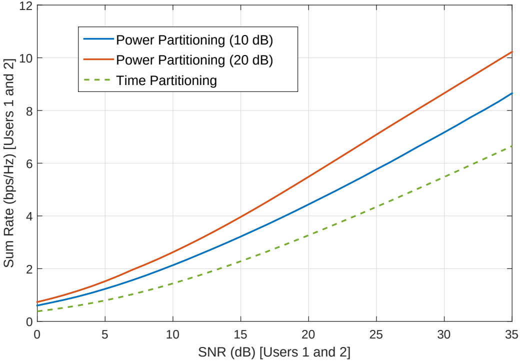
Scheduling approach



Power partitioning

- A **superposition of non-orthogonal layers and an orthogonal layer**
- Power partitioning achieves the optimum DoF region [20]

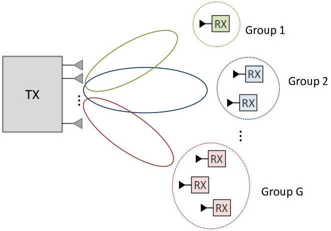
Overloaded systems: Numerical Results



- sum rate of RX-1 and RX-2 while maintaining the same rate for RX-3.
- Long-term SNR for RX-3 is 10 dB and 20 dB lower.
- Parameters: quality $\alpha = 0.5$, resource allocation $b = 0.5$.

Multigroup multicast beamforming

Users clustered into groups depending on content demand.



- K users grouped into $\mathcal{G}_1, \dots, \mathcal{G}_G$.
- One message for each group: $W_1, \dots, W_G$.
- Classical beamforming:

$$\mathbf{x} = \sum_{g=1}^{G} \mathbf{p}_g s_g.$$

Achieving max-min fairness (perfect CSIT):

$$\mathcal{R}(P) : \begin{cases} \max\limits_{\mathbf{P}_{\mathrm{p}}} & \min\limits_{g \in \{1,\dots,G\}} \min\limits_{i \in \mathcal{G}_g} R_i \\ \text{s.t.} & \sum\limits_{g=1}^{G} \|\mathbf{p}_g\|^2 \leq P. \end{cases}$$

- **Overloaded scenarios**: M is not enough for interference nulling [14].
- Rate saturation (even with perfect CSIT) due to inter-group interference.

Multigroup multicast beamforming: Simulation results

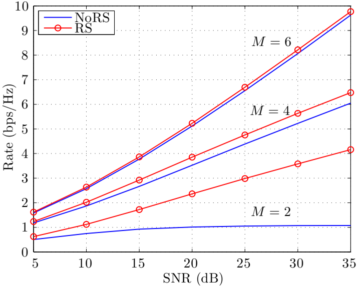


Figure: $K = 6$ users, $G = 3$ groups, $|\mathcal{G}_1| = 1$, $|\mathcal{G}_2| = 2$ and $|\mathcal{G}_3| = 3$.

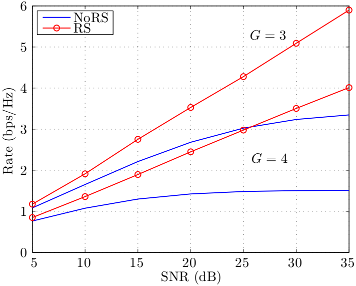
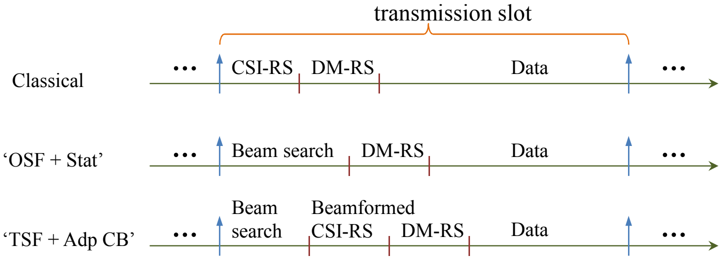


Figure: $M = 4$ antennas, $|\mathcal{G}_g| = 2$ users per group, $G = 3$ and 4.

RS to mitigate inter-group interference in overloaded scenarios [13, 14].

Multiuser Millimeter Wave Beamforming

Signalling and feedback procedure [26]



Multiuser Millimeter Wave Beamforming

RS to save second-stage channel training and feedback [26].

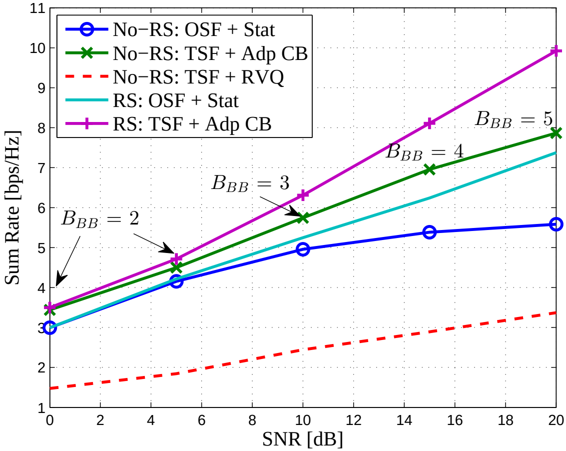


Figure: No-RS with extra feedback versus RS. $M = 32$, $K = 4$, $B_{RF} = 4$.

Bridging NOMA and MU-MIMO

RS: A general and powerful transmission framework [31]

- W_1, W_2 split into $\{W_1^{12}, W_1^1\}$ for user-1 and $\{W_2^{12}, W_2^2\}$ for user-2
- W_1^{12}, W_2^{12} are encoded together into a common stream s_{12}
- W_1^1 and W_2^2 encoded into private stream s_1 for user-1 and s_2 for user-2
- Data streams are linear precoded $\mathbf{x} = \mathbf{p}_{12}s_{12} + \mathbf{p}_1 s_1 + \mathbf{p}_2 s_2$
- Both users firstly decode s_{12} by treating s_1 and s_2 as noise.

Conventional MU-MISO with Linear Precoding

- Simply allocate no power to s_{12} and treat multi-user interference as noise.

NOMA based on SC-SIC

- Forcing user-1 to fully decode the message of user-2
- Allocate no power to s_2, encode W_1 into s_1 and encode W_2 into s_{12}
  $\mathbf{x} = \mathbf{p}_{12}s_{12} + \mathbf{p}_1 s_1$
- User-1 and user-2 decode s_{12} by treating s_1 as noise and user-1 decodes s_1 after canceling s_{12}

Bridging NOMA and MU-MIMO: Perfect CSIT

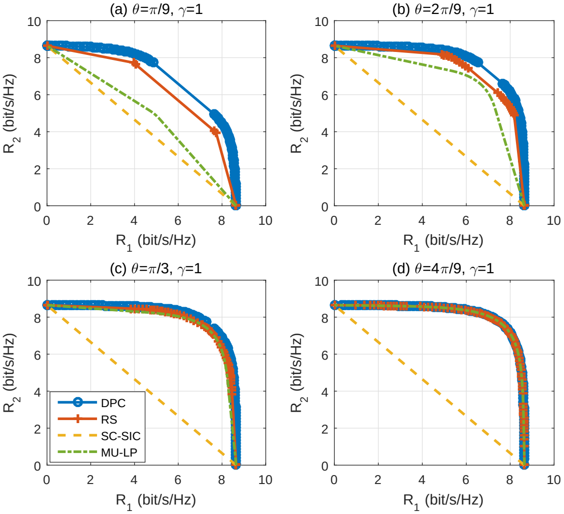


Figure: Achievable rate region of different strategies when $\gamma = 1$, SNR=20 dB [31].

Bridging NOMA and MU-MIMO: Perfect CSIT

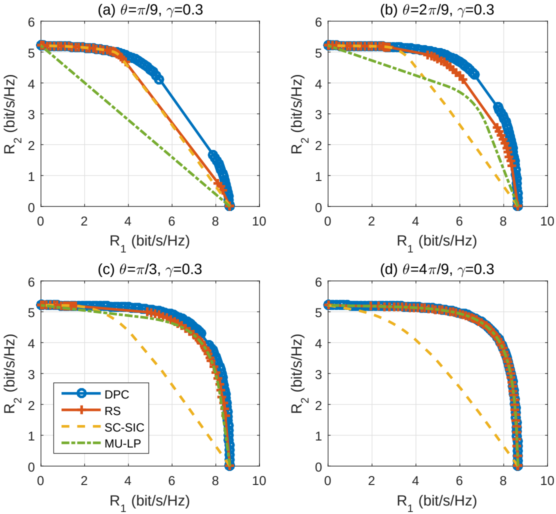


Figure: Achievable rate region with different strategies when $\gamma = 0.3$, SNR=20 dB [31].

Bridging NOMA and MU-MIMO: Imperfect CSIT

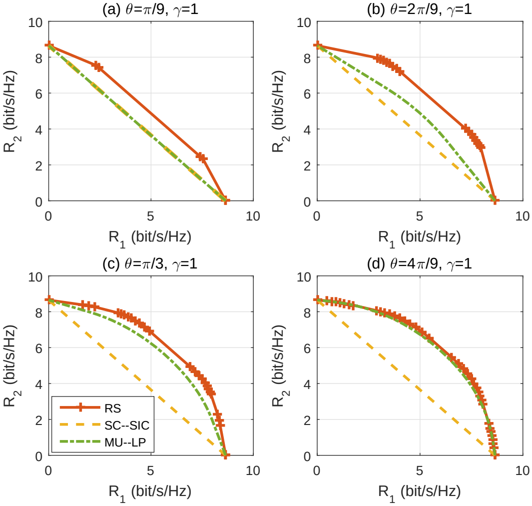


Figure: Achievable rate region of different strategies when $\gamma = 1$, SNR=20 dB [31].

Bridging NOMA and MU-MIMO: Imperfect CSIT

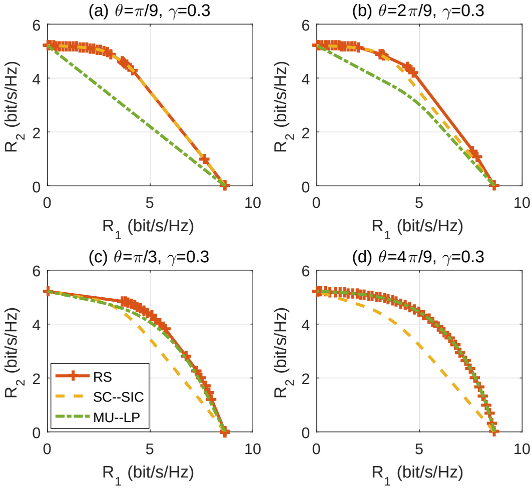
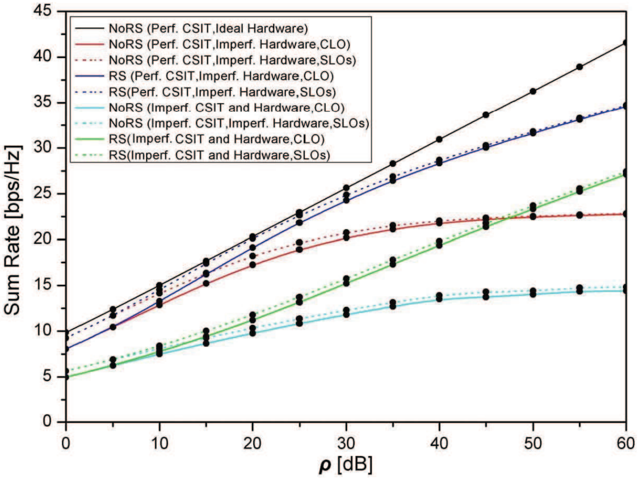


Figure: Achievable rate region with different strategies when $\gamma = 0.3$, SNR=20 dB [31].

RF Impairments

RS to mitigate phase noise impairments [**?**].



Conclusions and Future Challenges

Conclusions and Future Challenges

- 4G and current 5G candidates (MU-MIMO, CoMP, Massive MIMO, millimetre wave MIMO) rely on *private message transmissions*
  - *Treat interference as noise*
  - Such a strategy is only motivated in the presence of perfect CSIT
  - *Apply techniques designed for perfect CSIT to imperfect CSIT*

- NOMA forces strong users to *fully decode* and cancel interference created by weaker users:
  - Works only for degraded channels (SISO BC or MISO BC with aligned channels)

- RS **partially** decodes interference and **partially** treats interference as noise
  - Superposed transmission of common and private messages
  - Motivated by information theory for the realistic scenario of imperfect CSIT
  - A more general and powerful transmission framework
  - Benefits: unified framework, spectral/energy efficiencies, reliability, CSI feedback overhead reduction

- RS leads to fundamental changes in the design of PHY and Lower MAC
  - A gold mine of research problems for academia and industry

- The standardization of rate-splitting can leverage 3GPP current study/work items

Future Challenges: A gold mine of research problems

Introduction

- Overview, open problems, impact on standard specifications and operational challenges [25].

Fundamental Limits

- DoF region for K-user MISO BC with imperfect CSIT [5, 28].
- Capacity region of K-user MISO BC with imperfect CSIT: DPC + RS?
- DoF region for MIMO BC with imperfect CSIT [7, 8, 9].
- DoF region of overloaded MISO BC with imperfect CSIT [20].
- DoF region for MISO IC with imperfect CSIT [19]. TRS?
- DoF region for MIMO IC with imperfect CSIT [8]. RS + IA?
- Interplay between RS and coded caching [21, 30].

Optimization

- Ergodic sum-rate maximization for BC [11].
- Robust Max-Min Fairness for BC [12].
- RS beamforming optimization for other types of channels.

PHY challenges

- Finite SNR rate analysis [6].
- Energy efficiency of RS-based transmission.
- Space-time/frequency RS [23, 24, 6].
- RS with multi-carrier transmissions.

Future Challenges

PHY challenges (continued)

- RS with non-linear precoding.
- Diversity (and BER) performance of RS-based strategies.
- RS for Multigroup Multicast [13, 14].
- RS/HRS for Massive MIMO [10].
- RS as a way to combat pilot contamination.
- RS to mitigate hardware impairments [27].
- RS in higher frequency bands operation (e.g. millimeter-wave) [26].
- RS-based network MIMO.
- Coordination/cooperation among distributed antennas in homogeneous and heterogeneous network deployments.
- RS in half-duplex relay.
- RS in full duplex.
- RS in overloaded systems [20].
- RS and NOMA/MUST [31].
- RS and superposition of multicast and unicast messages.
- RS and physical layer security.
- RS in D2D and cognitive radio [29].

Future Challenges

PHY/MAC challenges

- User pairing and scheduling of common and private messages.
- RS design with Quality of Experience (QoE) and traffic constraints.

Performance Analysis

- Performance analysis of RS using stochastic geometry.

Standardization

- Link and system-level evaluations of RS.
- MIMO receiver implementation.
- Transmission schemes/mode.
- CSI feedback mechanisms.
- Downlink and uplink signaling.

References I

[1] N. Jindal, “MIMO Broadcast Channels With Finite-Rate Feedback,” Trans. on Info Theory, vol. 52, no. 11, pp. 5045-5060, Nov. 2006.

[2] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, Multiuser MIMO achievable rates with downlink training and channel state feedback, IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 28452866, Jun. 2010.

[3] A. El Gamal and Y.-H. Kim, *Network information theory*, Cambridge university press, 2011.

[4] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT, IEEE Trans. Inf. Theory, vol. 59, no. 1, pp. 315328, Jan. 2013.

[5] A.G. Davoodi, S.A. Jafar, “Aligned Image Sets under Channel Uncertainty: Settling a Conjecture by Lapidoth, Shamai and Wigger on the Collapse of Degrees of Freedom under Finite Precision CSIT,” arXiv:1403.1541.

[6] C. Hao, Y. Wu and B. Clerckx, “Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach,” IEEE Trans. on Comm., vol. 63, no. 9, pp. 3232-3246, Sept. 2015.

[7] J. Chen and P. Elia, Symmetric two-user MIMO BC with evolving feedback, in Inf. Theory and Applications Workshop (ITA), 2014, San Diego, CA, Feb 2014, pp. 15.

[8] C. Hao, B. Rassouli and B. Clerckx, “Achievable DoF Regions of MIMO Networks with Imperfect CSIT,” IEEE Trans. on Info Theory, vol. 63, no. 10, pp. 6587-6606, Oct 2017.

[9] B. Yuan, S.A. Jafar, “Elevated Multiplexing and Signal Space Partitioning in the 2 User MIMO IC with Partial CSIT,” IEEE SPAWC 2016.

[10] M. Dai, B. Clerckx, D. Gesbert and G. Caire, “A Rate Splitting Strategy for Massive MIMO with Imperfect CSIT,” IEEE IEEE Trans. on Wireless Comm., vol. 15, no. 7, pp. 4611-4624, July 2016.

[11] H. Joudeh and B. Clerckx, “Sum-Rate Maximization for Linearly Precoded Downlink Multiuser MISO Systems with Partial CSIT: A Rate-Splitting Approach,” IEEE Trans. on Comm., IEEE Trans. on Comm. vol. 64, no. 11, pp. 4847-4861, Nov 2016.

[12] H. Joudeh and B. Clerckx, “Robust Transmission in Downlink Multiuser MISO Systems: A Rate-Splitting Approach,” IEEE Trans. on Sig. Proc. Vol. 64, No. 23, pp. 6227-6242 , Dec 2016.

[13] H. Joudeh and B. Clerckx, “A Rate-Splitting Strategy for Max-Min Fair Multigroup Multicasting,” IEEE SPAWC 2016.

[14] H. Joudeh and B. Clerckx, Rate-Splitting for Max-Min Fair Multigroup Multicast Beamforming in Overloaded Systems, IEEE Trans. on Wireless Commun., vol 16, no 11, pp 7276-7289, Nov 2017.

References II

[15] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, “Recent trend of multiuser MIMO in LTE-advanced,” IEEE Commun. Mag., vol. 51, no. 3, pp. 127135, March 2013.

[16] A.Adhikary, J.Nam, J.Ahn and G.Caire. “Joint spatial division and multiplexing - the large-scale array regime,” IEEE TIT, Oct. 2013.

[17] J. Chen and V. Lau, “Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 12301238, June 2014.

[18] X. Yi, D. Gesbert, S. Yang, and M. Kobayashi, On the dof of the multiple-antenna time correlated interference channel with delayed csit, in 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Nov 2012, pp. 15661570.

[19] C. Hao and B. Clerckx, “MISO Networks with Imperfect CSIT: A Topological Rate Splitting Approach,” IEEE Trans. on Comm., Vol 65, No 5, pp. 2164-2179, May 2017

[20] E. Piovano, H. Joudeh and B. Clerckx, “Overloaded MU-MISO Transmission with Imperfect CSIT,” 2016 Asilomar Conference on Signals, Systems, and Computers.

[21] J. Zhang, F. Engelmann and P. Elia, “Coded caching for reducing CSIT-feedback in wireless communications,” Allerton, Monticello, IL, 2015, pp. 1099-1105.

[22] B. Rassouli, C. Hao and B. Clerckx, “DoF Analysis of the MIMO Broadcast Channel with Alternating/Hybrid CSIT,” IEEE Trans. on Info Theory, vol. 62, no. 3, pp. 1312-1325, Mar. 2016.

[23] R. Tandon, S. Jafar, S. S. Shitz, and H. Poor, On the synergistic benefits of alternating CSIT for the MISO broadcast channel, IEEE Trans. Inf. Theory., vol. 59, no. 7, pp. 41064128, Jul. 2013.

[24] J. Chen and P. Elia, Optimal DoF region of the two-user MISO-BC with general alternating CSIT, in Proc. Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, USA, Nov. 2013, pp. 18601864.

[25] B. Clerckx, H. Joudeh, C. Hao, M. Dai and B. Rassouli, “Rate Splitting for MIMO Wireless Networks: A Promising PHY-Layer Strategy for LTE Evolution,” IEEE Comm. Mag, pp 98-105, May 2016.

[26] M. Dai and B. Clerckx, “Multiuser Millimeter Wave Beamforming Strategies with Quantized and Statistical CSIT,” IEEE Trans. on Wireless Comm., vol 16, no 11, pp 7025-7038, Nov 2017.

[27] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, “Rate-Splitting to Mitigate Residual Transceiver Hardware Impairments in Massive MIMO Systems,” IEEE Trans. on Veh. Tech., vol 66, no 9, pp 8196-8211, Sept 2017.

References III

[28] E. Piovano and B. Clerckx, “Optimal DoF region of the K-User MISO BC with Partial CSIT,” IEEE Commun. Letters, vol 21, no 11, pp 2368-2371, Nov 2017.

[29] A. Zappone, B. Matthiesen, E. Jorswieck, “Energy Efficiency in MIMO Underlay and Overlay Device-to-Device Communications and Cognitive Radio Systems,” IEEE Transactions on Signal Processing, 2016.

[30] E. Piovano, H. Joudeh and B. Clerckx, “On Coded Caching in the Overloaded MISO Broadcast Channel,” IEEE ISIT 2017.

[31] Y. Mao, B. Clerckx and V.O.K. Li, “Rate-Splitting for Downlink Multi-User MISO Systems: Bridging NOMA and Conventional Linear Precoding,” in submission