UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Communication Rates for Fading Channels with Imperfect Channel-State Information

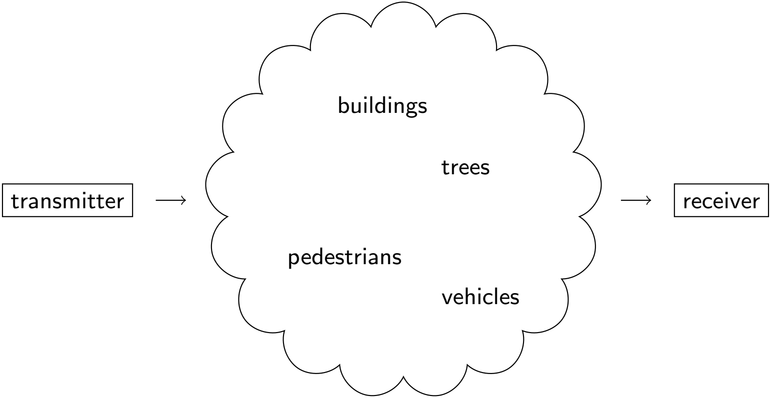
PhD Defense

Author: Adriano Pastore
Supervisor: Javier Rodríguez Fonollosa

Universitat Politecnica de Catalunya
Dep. of Signal Theory and Communications
Barcelona, Spain

June 27, 2014

What is fading?



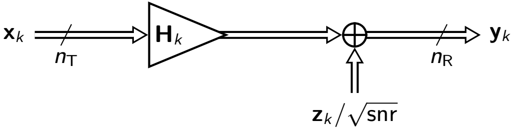
The communication suffers from:

- shadowing
- scattering
- multiple reflections

all of which **vary in time!**

Outline

MIMO system with perfect CSI

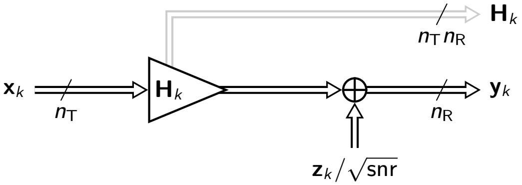


Discrete-time system equation

At time instant k, the output $\mathbf{y}_k$ corresponding to the input $\mathbf{x}_k$ is

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}_k \qquad (k \in \mathbb{Z})$$

MIMO system with perfect CSI

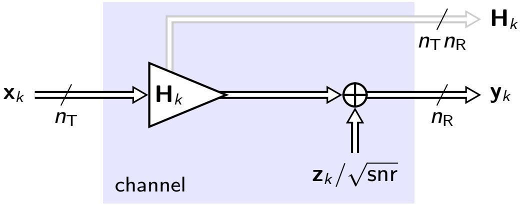


Discrete-time system equation

At time instant k, the output $\mathbf{y}_k$ corresponding to the input $\mathbf{x}_k$ is

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}_k \qquad (k \in \mathbb{Z})$$

MIMO system with perfect CSI

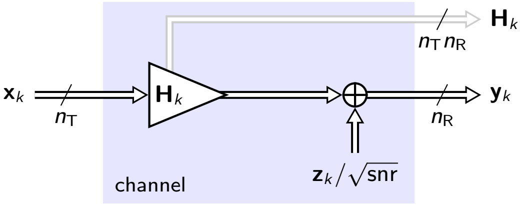


Discrete-time system equation

At time instant k, the output $\mathbf{y}_k$ corresponding to the input $\mathbf{x}_k$ is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \frac{1}{\sqrt{\mathrm{snr}}} \mathbf{z}_k \qquad (k \in \mathbb{Z})$$

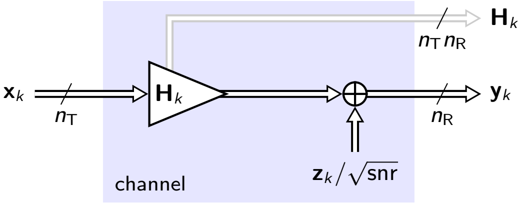
MIMO system with perfect CSI



Assumptions:

- $\{\mathbf{H}_k\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\{\mathbf{H}_k\}$ known at the receiver

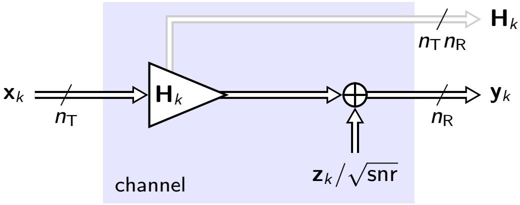
MIMO system with perfect CSI



Assumptions:

- $\{\mathbf{H}_k\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\{\mathbf{H}_k\}$ known at the receiver

MIMO system with perfect CSI



Assumptions:

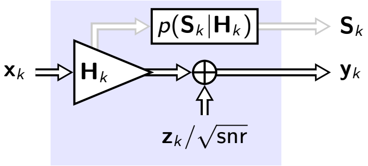
- $\{\mathbf{H}_k\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\{\mathbf{H}_k\}$ known at the receiver

Coding theorem

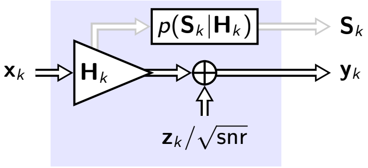
$$\lim_{n\to\infty} \frac{1}{2n+1} I\left(\mathbf{x}_{-n}^{n}; \mathbf{y}_{-n}^{n} \;\middle|\; \mathbf{H}_{-n}^{n}\right) = I(\mathbf{x}_0; \mathbf{y}_0 | \mathbf{H}_0) = I(\mathbf{x}; \mathbf{y} | \mathbf{H})$$

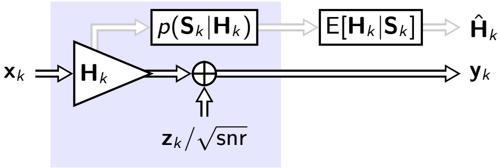
is an achievable rate.

MIMO system with imperfect CSI (1 of 2)



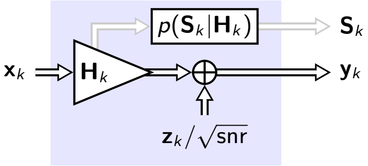
MIMO system with imperfect CSI (1 of 2)

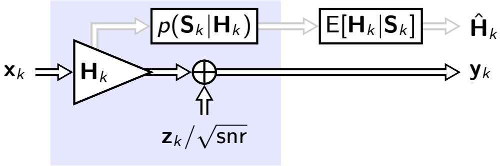




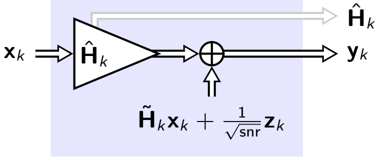
$\{\hat{\mathbf{H}}_k\}$ and $\{\mathbf{S}_k\}$ are equally informative!

MIMO system with imperfect CSI (1 of 2)



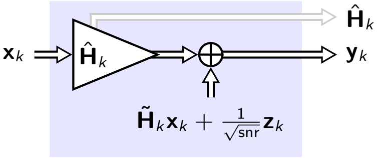


$\{\hat{\mathbf{H}}_k\}$ and $\{\mathbf{S}_k\}$
are equally informative!



Channel error:
$\tilde{\mathbf{H}}_k = \mathbf{H}_k - \hat{\mathbf{H}}_k$
$\mathrm{E}[\tilde{\mathbf{H}}_k|\hat{\mathbf{H}}_k] = \mathbf{0}$

MIMO system with imperfect CSI (2 of 2)



Discrete-time system equation

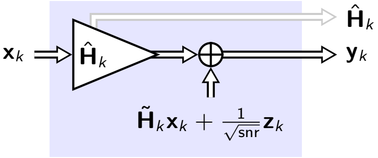
At time instant $k \in \mathbb{Z}$, the system output is

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k + \frac{1}{\sqrt{\mathsf{snr}}} \mathbf{z}_k \\ &= \left(\hat{\mathbf{H}}_k + \tilde{\mathbf{H}}_k \right) \mathbf{x}_k + \frac{1}{\sqrt{\mathsf{snr}}} \mathbf{z}_k \end{aligned}$$

Assumptions:

- $\{(\hat{\mathbf{H}}_k, \tilde{\mathbf{H}}_k)\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\hat{\mathbf{H}}$ known at the receiver, $\tilde{\mathbf{H}}$ unknown
- $\mathsf{E}[\tilde{\mathbf{H}} | \hat{\mathbf{H}}] = \mathbf{0}$ (for almost all $\hat{\mathbf{H}}$)

MIMO system with imperfect CSI (2 of 2)



Discrete-time system equation

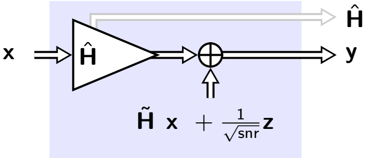
At time instant $k \in \mathbb{Z}$, the system output is

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k + \frac{1}{\sqrt{\mathsf{snr}}} \mathbf{z}_k \\ &= \left(\hat{\mathbf{H}}_k + \tilde{\mathbf{H}}_k \right) \mathbf{x}_k + \frac{1}{\sqrt{\mathsf{snr}}} \mathbf{z}_k \end{aligned}$$

Assumptions:

- $\{(\hat{\mathbf{H}}_k, \tilde{\mathbf{H}}_k)\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\hat{\mathbf{H}}$ known at the receiver, $\tilde{\mathbf{H}}$ unknown
- $\mathsf{E}[\tilde{\mathbf{H}} | \hat{\mathbf{H}}] = \mathbf{0}$ (for almost all $\hat{\mathbf{H}}$)

MIMO system with imperfect CSI (2 of 2)



Discrete-time system equation

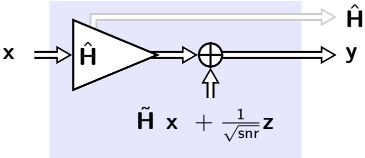
At time instant $k \in \mathbb{Z}$, the system output is

$$\begin{aligned} \mathbf{y} &= \mathbf{H} \mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \\ &= \left(\hat{\mathbf{H}} + \tilde{\mathbf{H}}\right)\mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \end{aligned}$$

Assumptions:

- $\{(\hat{\mathbf{H}}_k, \tilde{\mathbf{H}}_k)\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\hat{\mathbf{H}}$ known at the receiver, $\tilde{\mathbf{H}}$ unknown
- $\mathsf{E}[\tilde{\mathbf{H}}|\hat{\mathbf{H}}] = \mathbf{0}$ (for almost all $\hat{\mathbf{H}}$)

MIMO system with imperfect CSI (2 of 2)



Discrete-time system equation

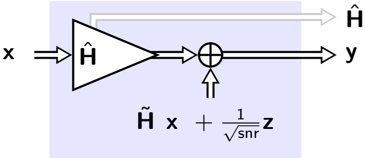
At time instant $k \in \mathbb{Z}$, the system output is

$$\begin{aligned}\mathbf{y} &= \mathbf{H}\,\mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \\ &= (\hat{\mathbf{H}} + \tilde{\mathbf{H}})\,\mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}\end{aligned}$$

Assumptions:

- $\{(\hat{\mathbf{H}}_k, \tilde{\mathbf{H}}_k)\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\hat{\mathbf{H}}$ known at the receiver, $\tilde{\mathbf{H}}$ unknown
- $\mathrm{E}[\tilde{\mathbf{H}}|\hat{\mathbf{H}}] = \mathbf{0}$ (for almost all $\hat{\mathbf{H}}$)

MIMO system with imperfect CSI (2 of 2)



Discrete-time system equation

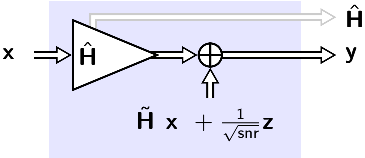
At time instant $k \in \mathbb{Z}$, the system output is

$$\begin{aligned} \mathbf{y} &= \mathbf{H}\,\mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \\ &= (\hat{\mathbf{H}} + \tilde{\mathbf{H}})\,\mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \end{aligned}$$

Assumptions:

- $\{(\hat{\mathbf{H}}_k, \tilde{\mathbf{H}}_k)\}$, $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are mutually independent i.i.d. sequences
- $\hat{\mathbf{H}}$ known at the receiver, $\tilde{\mathbf{H}}$ unknown
- $\mathrm{E}[\tilde{\mathbf{H}}|\hat{\mathbf{H}}] = \mathbf{0}$ (for almost all $\hat{\mathbf{H}}$)

MIMO system with imperfect CSI (2 of 2)



Coding theorem

$$\lim_{n\to\infty} \frac{1}{2n+1} I(\mathbf{x}_{-n}^{n}; \mathbf{y}_{-n}^{n} | \hat{\mathbf{H}}_{-n}^{n}) = I(\mathbf{x}_0; \mathbf{y}_0 | \hat{\mathbf{H}}_0) = I(\mathbf{x}; \mathbf{y} | \hat{\mathbf{H}})$$

is an achievable rate.

The worst-case noise lower bound

Assume a Gaussian input $\mathbf{x}_{\mathsf{G}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q})$

The worst-case noise lower bound

Assume a Gaussian input $\mathbf{x}_\mathsf{G} \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathbf{Q})$

Rewrite the system equation:

$$\begin{aligned}\mathbf{y} &= (\hat{\mathbf{H}} + \tilde{\mathbf{H}})\mathbf{x}_\mathsf{G} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \\ &= \hat{\mathbf{H}}\mathbf{x}_\mathsf{G} + \underbrace{\tilde{\mathbf{H}}\mathbf{x}_\mathsf{G} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}\end{aligned}$$

The worst-case noise lower bound

Assume a Gaussian input $\mathbf{x}_\mathrm{G} \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathbf{Q})$

Rewrite the system equation:

$$\begin{aligned}\mathbf{y} &= (\hat{\mathbf{H}} + \tilde{\mathbf{H}})\mathbf{x}_\mathrm{G} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z} \\ &= \hat{\mathbf{H}}\mathbf{x}_\mathrm{G} + \underbrace{\tilde{\mathbf{H}}\mathbf{x}_\mathrm{G} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}\end{aligned}$$

Worst-case noise bound
[Médard '00, Hassibi & Hochwald '03, Yoo & Goldsmith '06]

$$\begin{aligned} I(\mathbf{x}_\mathrm{G}; \mathbf{y} | \hat{\mathbf{H}}) &\geq \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\mathrm{E}\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger | \hat{\mathbf{H}}\right] + \tfrac{1}{\mathsf{snr}}\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]\right)}_{\text{effective noise covariance } \boldsymbol{\Gamma}_\mathrm{eff}}{}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right)\right] \\ &\triangleq R_\mathrm{WCN}\end{aligned}$$

Assume $\mathrm{E}[\mathbf{z}\mathbf{z}^\dagger] = \mathbf{I}_r$ without loss of generality

Covariance-constrained capacity

- Perfect CSIR [Foschini & Gans '98, Telatar '99]:

$$C_{\mathrm{coh}} = \sup_{\mathrm{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I(\mathbf{x};\mathbf{y}|\mathbf{H}) = I\big(\mathbf{x}_{\mathrm{G}};\mathbf{y}\big|\mathbf{H}\big) = \mathrm{E}_{\mathbf{H}}\Big[\log\det\big(\mathbf{I} + \boldsymbol{\Gamma}^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\big)\Big]$$

Covariance-constrained capacity

- Perfect CSIR [Foschini & Gans '98, Telatar '99]:

$$C_{\mathrm{coh}} = \sup_{\mathsf{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I(\mathbf{x};\mathbf{y}|\mathbf{H}) = I\big(\mathbf{x}_{\mathrm{G}};\mathbf{y}\big|\mathbf{H}\big) = \mathsf{E}_{\mathbf{H}}\Big[\log\det\big(\mathbf{I} + \boldsymbol{\Gamma}^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\big)\Big]$$

- Imperfect CSIR [Médard '00, Hassibi & Hochwald '03]:

$$C = \sup_{\mathsf{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I\big(\mathbf{x};\mathbf{y}\big|\hat{\mathbf{H}}\big) \geq I\big(\mathbf{x}_{\mathrm{G}};\mathbf{y}\big|\hat{\mathbf{H}}\big) \geq \mathsf{E}_{\hat{\mathbf{H}}}\Big[\log\det\big(\mathbf{I} + \boldsymbol{\Gamma}_{\mathrm{eff}}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\big)\Big]$$

Covariance-constrained capacity

- Perfect CSIR [Foschini & Gans '98, Telatar '99]:

$$C_{\text{coh}} = \sup_{\mathsf{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I(\mathbf{x};\mathbf{y}|\mathbf{H}) = I\left(\mathbf{x}_{\text{G}};\mathbf{y}\middle|\mathbf{H}\right) = \mathsf{E}_{\mathbf{H}}\left[\log\det\left(\mathbf{I} + \boldsymbol{\Gamma}^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right)\right]$$

- Imperfect CSIR [Médard '00, Hassibi & Hochwald '03]:

$$C = \sup_{\mathsf{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I\left(\mathbf{x};\mathbf{y}\middle|\hat{\mathbf{H}}\right) \geq I\left(\mathbf{x}_{\text{G}};\mathbf{y}\middle|\hat{\mathbf{H}}\right) \geq \mathsf{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \boldsymbol{\Gamma}_{\text{eff}}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right)\right]$$

Can we improve the lower bound?

$$\boxed{C \;\geq\; I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \quad\geq\quad \mathsf{E}\left[\log\det\left(\mathbf{I} + \boldsymbol{\Gamma}_{\text{eff}}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right)\right]}$$

Covariance-constrained capacity

- Perfect CSIR [Foschini & Gans '98, Telatar '99]:

$$C_{\text{coh}} = \sup_{\mathsf{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I(\mathbf{x};\mathbf{y}|\mathbf{H}) = I\left(\mathbf{x}_{\mathsf{G}};\mathbf{y}\middle|\mathbf{H}\right) = \mathsf{E}_{\mathbf{H}}\left[\log\det\left(\mathbf{I} + \boldsymbol{\Gamma}^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right)\right]$$

- Imperfect CSIR [Médard '00, Hassibi & Hochwald '03]:

$$C = \sup_{\mathsf{E}[\mathbf{x}\mathbf{x}^\dagger]=\mathbf{Q}} I\left(\mathbf{x};\mathbf{y}\middle|\hat{\mathbf{H}}\right) \geq I\left(\mathbf{x}_{\mathsf{G}};\mathbf{y}\middle|\hat{\mathbf{H}}\right) \geq \mathsf{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \boldsymbol{\Gamma}_{\text{eff}}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right)\right]$$

Can we improve the lower bound?

$$\boxed{C \;\geq\; I(\mathbf{x}_{\mathsf{G}};\mathbf{y}|\hat{\mathbf{H}}) \;\geq\; \mathbf{?} \;\geq\; \mathsf{E}\left[\log\det\left(\mathbf{I} + \boldsymbol{\Gamma}_{\text{eff}}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right)\right]}$$

Rate splitting: How does it help? (1 of 3)

Decompose $\mathbf{x}_\mathrm{G}$ into a sum of two independent Gaussian signals:

$$\underbrace{\mathbf{x}_\mathrm{G}}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q})} = \underbrace{\mathbf{x}_1}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q}_1)} + \underbrace{\mathbf{x}_2}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q}_2)} \qquad (\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2)$$

Rate splitting: How does it help? (1 of 3)

Decompose $\mathbf{x}_\mathrm{G}$ into a sum of two independent Gaussian signals:

$$\underbrace{\mathbf{x}_\mathrm{G}}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q})} = \underbrace{\mathbf{x}_1}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q}_1)} + \underbrace{\mathbf{x}_2}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q}_2)} \qquad (\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2)$$

$$\boxed{\mathbf{y} = \left(\hat{\mathbf{H}} + \tilde{\mathbf{H}}\right)\left(\mathbf{x}_1 + \mathbf{x}_2\right) + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}$$

Rate splitting: How does it help? (1 of 3)

Decompose $\mathbf{x}_{\mathrm{G}}$ into a sum of two independent Gaussian signals:

$$\underbrace{\mathbf{x}_{\mathrm{G}}}_{\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{Q})} = \underbrace{\mathbf{x}_1}_{\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{Q}_1)} + \underbrace{\mathbf{x}_2}_{\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{Q}_2)} \qquad (\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2)$$

$$\boxed{\mathbf{y} = \left(\hat{\mathbf{H}} + \tilde{\mathbf{H}}\right)\left(\mathbf{x}_1 + \mathbf{x}_2\right) + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}$$

Chain rule

$$\begin{aligned} I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}}) &= I(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}|\hat{\mathbf{H}}) \\ &= I(\mathbf{x}_1;\mathbf{y}|\hat{\mathbf{H}}) + I(\mathbf{x}_2;\mathbf{y}|\hat{\mathbf{H}},\mathbf{x}_1) \end{aligned}$$

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\text{WCN},1}$:

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\text{WCN},2}$:

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\mathrm{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\mathrm{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\mathrm{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\mathrm{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\text{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \frac{1}{\sqrt{\text{snr}}}\mathbf{z}}_{\text{effective noise}}$$

$$R_{\text{WCN},1} = \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}_1 + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}\right] + \frac{1}{\text{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_1\hat{\mathbf{H}}^\dagger\right)\right]$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\text{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\mathrm{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \frac{1}{\sqrt{\mathrm{snr}}}\mathbf{z}}_{\text{effective noise}}$$

$$R_{\mathrm{WCN},1} = \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}_1 + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}\right] + \frac{1}{\mathrm{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_1\hat{\mathbf{H}}^\dagger\right)\right]$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\mathrm{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\text{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \tfrac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}$$

$$R_{\text{WCN},1} = \mathsf{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathsf{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}_1 + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}\right] + \tfrac{1}{\mathsf{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_1\hat{\mathbf{H}}^\dagger\right)\right]$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\text{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\text{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \tfrac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}$$

$$R_{\text{WCN},1} = \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}_1 + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}\right] + \tfrac{1}{\mathsf{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_1\hat{\mathbf{H}}^\dagger\right)\right]$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\text{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

$$\begin{aligned}\mathbf{y}' &= \mathbf{y} - \hat{\mathbf{H}}\mathbf{x}_1 \\ &= \hat{\mathbf{H}}\mathbf{x}_2 + \underbrace{\tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \tfrac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}\end{aligned}$$

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\mathrm{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \tfrac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}$$

$$R_{\mathrm{WCN},1} = \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}_1 + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}\right] + \tfrac{1}{\mathsf{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_1\hat{\mathbf{H}}^\dagger\right)\right]$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\mathrm{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

$$\begin{aligned}\mathbf{y}' &= \mathbf{y} - \hat{\mathbf{H}}\mathbf{x}_1 \\ &= \hat{\mathbf{H}}\mathbf{x}_2 + \underbrace{\tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \tfrac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}\end{aligned}$$

$$R_{\mathrm{WCN},2} = \mathrm{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{x}_1\mathbf{x}_1^\dagger + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}, \mathbf{x}_1\right] + \tfrac{1}{\mathsf{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger\right)\right]$$

Rate splitting: How does it help? (2 of 3)

1) First term $I(\mathbf{x}_1; \mathbf{y}|\hat{\mathbf{H}}) \geq R_{\mathrm{WCN},1}$: ("decode $\mathbf{x}_1$ knowing $\hat{\mathbf{H}}$")

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x}_1 + \underbrace{\hat{\mathbf{H}}\mathbf{x}_2 + \tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}$$

$$R_{\mathrm{WCN},1} = \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}_1 + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}\right] + \frac{1}{\mathsf{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_1\hat{\mathbf{H}}^\dagger\right)\right]$$

2) Second term $I(\mathbf{x}_2; \mathbf{y}|\hat{\mathbf{H}}, \mathbf{x}_1) \geq R_{\mathrm{WCN},2}$: ("decode $\mathbf{x}_2$ knowing $\hat{\mathbf{H}}$ and $\mathbf{x}_1$")

$$\begin{aligned}\mathbf{y}' &= \mathbf{y} - \hat{\mathbf{H}}\mathbf{x}_1 \\ &= \hat{\mathbf{H}}\mathbf{x}_2 + \underbrace{\tilde{\mathbf{H}}(\mathbf{x}_1 + \mathbf{x}_2) + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{z}}_{\text{effective noise}}\end{aligned}$$

$$R_{\mathrm{WCN},2} = \mathrm{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I} + \underbrace{\left(\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{x}_1\mathbf{x}_1^\dagger + \mathbf{Q}_2)\tilde{\mathbf{H}}^\dagger \mid \hat{\mathbf{H}}, \mathbf{x}_1\right] + \frac{1}{\mathsf{snr}}\mathbf{I}\right)}_{\text{effective noise covariance}}^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger\right)\right]$$

Rate splitting: How does it help? (3 of 3)

$$\left.\begin{array}{rrcl} \text{WCN bound:} & R_{\text{WCN}} & \leq & I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \\ \text{WCN bound with RS:} & R_{\text{WCN},1} + R_{\text{WCN},2} & \leq & I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \end{array}\right\} \text{which is best?}$$

Rate splitting: How does it help? (3 of 3)

$$\left.\begin{array}{rrcl} \text{WCN bound:} & R_{\mathrm{WCN}} & \leq & I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}}) \\ \text{WCN bound with RS:} & R_{\mathrm{WCN},1}+R_{\mathrm{WCN},2} & \leq & I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}}) \end{array}\right\}\ \text{which is best?}$$

$$R_{\mathrm{WCN},2} = \mathrm{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I}+\left(\ldots\mathbf{x}_1\mathbf{x}_1^{\dagger}\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^{\dagger}\right)\right]$$

Rate splitting: How does it help? (3 of 3)

$$\left.\begin{array}{rrcl} \text{WCN bound:} & R_{\text{WCN}} & \leq & I(\mathbf{x}_G; \mathbf{y}|\hat{\mathbf{H}}) \\ \text{WCN bound with RS:} & R_{\text{WCN},1} + R_{\text{WCN},2} & \leq & I(\mathbf{x}_G; \mathbf{y}|\hat{\mathbf{H}}) \end{array}\right\} \text{which is best?}$$

$$R_{\text{WCN},2} = \mathrm{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I} + \left(\ldots \mathbf{x}_1 \mathbf{x}_1^\dagger \ldots\right)^{-1} \hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger\right)\right]$$

Rate splitting: How does it help? (3 of 3)

$$\left.\begin{array}{rrcl} \text{WCN bound:} & R_{\text{WCN}} & \leq & I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \\ \text{WCN bound with RS:} & R_{\text{WCN},1}+R_{\text{WCN},2} & \leq & I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \end{array}\right\} \text{which is best?}$$

$$\begin{aligned} R_{\text{WCN},2} &= \mathrm{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I}+\left(\ldots\mathbf{x}_1\mathbf{x}_1^{\dagger}\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^{\dagger}\right)\right] \\ &\overset{\text{Jensen's ineq.}}{\geq} \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I}+\left(\ldots\mathbf{Q}_1\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^{\dagger}\right)\right] \end{aligned}$$

Rate splitting: How does it help? (3 of 3)

$$\left.\begin{array}{rrcl} \text{WCN bound:} & R_{\text{WCN}} & \leq & I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \\ \text{WCN bound with RS:} & R_{\text{WCN},1} + R_{\text{WCN},2} & \leq & I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}) \end{array}\right\} \text{ which is best?}$$

$$\begin{aligned} R_{\text{WCN},2} &= \mathsf{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I} + \left(\ldots\mathbf{x}_1\mathbf{x}_1^\dagger\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger\right)\right] \\ &\overset{\text{Jensen's ineq.}}{\geq} \mathsf{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I} + \left(\ldots\mathbf{Q}_1\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^\dagger\right)\right] \\ &= R_{\text{WCN}} - R_{\text{WCN},1} \end{aligned}$$

Rate splitting: How does it help? (3 of 3)

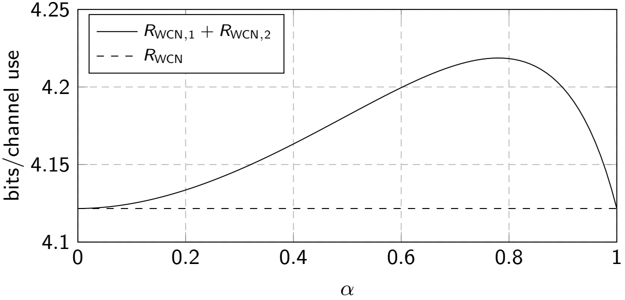
$$\left.\begin{array}{rrcl} \text{WCN bound:} & R_{\mathrm{WCN}} & \leq & I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}}) \\ \text{WCN bound with RS:} & R_{\mathrm{WCN},1}+R_{\mathrm{WCN},2} & \leq & I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}}) \end{array}\right\} \text{which is best?}$$

$$\begin{aligned} R_{\mathrm{WCN},2} &= \mathrm{E}_{\hat{\mathbf{H}},\mathbf{x}_1}\left[\log\det\left(\mathbf{I}+\left(\ldots\mathbf{x}_1\mathbf{x}_1^{\dagger}\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^{\dagger}\right)\right] \\ &\overset{\text{Jensen's ineq.}}{\geq} \mathrm{E}_{\hat{\mathbf{H}}}\left[\log\det\left(\mathbf{I}+\left(\ldots\mathbf{Q}_1\ldots\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}_2\hat{\mathbf{H}}^{\dagger}\right)\right] \\ &= R_{\mathrm{WCN}} - R_{\mathrm{WCN},1} \end{aligned}$$

Rate splitting helps!

$$\boxed{R_{\mathrm{WCN}} \leq R_{\mathrm{WCN},1} + R_{\mathrm{WCN},2} \triangleq R_{\mathrm{RS}}}$$

Rate splitting: numerical example



Rate-splitting bound $R_{\mathrm{WCN},1} + R_{\mathrm{WCN},2}$ compared to R_{WCN} for $(\mathbf{Q}_1, \mathbf{Q}_2) = (\alpha\mathbf{Q}, (1-\alpha)\mathbf{Q})$ and $\mathrm{snr} = 10\,\mathrm{dB}$, $n_{\mathrm{R}} = 6$, $n_{\mathrm{T}} = 4$

Rate splitting: general formula

Can we generalize this rate-splitting approach?

Rate splitting: general formula

Can we generalize this rate-splitting approach?

- $(\mathbf{Q}_1, \mathbf{Q}_2) = (\alpha\mathbf{Q}, (1-\alpha)\mathbf{Q})$ is not the only possible decomposition! For example:

$$\mathbf{Q}_1 = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\dagger$$

$$\mathbf{Q}_2 = \sum_{i=1}^{n_\mathsf{T}} \lambda_i \mathbf{u}_i \mathbf{u}_i^\dagger$$

Rate splitting: general formula

Can we generalize this rate-splitting approach?

- $(\mathbf{Q}_1, \mathbf{Q}_2) = (\alpha\mathbf{Q}, (1-\alpha)\mathbf{Q})$ is not the only possible decomposition!
  For example:

$$\mathbf{Q}_1 = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\dagger$$
$$\mathbf{Q}_2 = \sum_{i=1}^{n_\mathsf{T}} \lambda_i \mathbf{u}_i \mathbf{u}_i^\dagger$$

- How about more than 2 splits?

$$\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2 + \ldots + \mathbf{Q}_L$$

Rate splitting: general formula

Decompose $\mathbf{x}_\mathrm{G}$ into a sum of L signals (layers):

$$\underbrace{\mathbf{x}_\mathrm{G}}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q})} = \sum_{\ell=1}^{L} \underbrace{\mathbf{x}_\ell}_{\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{Q}_\ell)} \qquad \mathbf{Q} = \sum_{\ell=1}^{L} \mathbf{Q}_\ell$$

Worst-case noise bound with L-layer rate splitting

$$R_\mathrm{WCN} \overset{\text{Jensen}}{\leq} \underbrace{\boxed{\sum_{\ell=1}^{L} \mathrm{E}\log\det\left(\mathbf{I} + \hat{\mathbf{H}}^\dagger \boldsymbol{\Gamma}_\ell^{-1} \hat{\mathbf{H}} \mathbf{Q}_\ell\right)}}_{R_\mathrm{RS}(\mathbf{Q}_1,\ldots,\mathbf{Q}_L)} \overset{\substack{\text{chain rule} \\ +\text{ WCN}}}{\leq} I\left(\mathbf{x}_\mathrm{G};\mathbf{y}\middle|\hat{\mathbf{H}}\right)$$

with

$$\boldsymbol{\Gamma}_\ell = \underbrace{\mathrm{E}\left[\tilde{\mathbf{H}}\left(\sum_{\ell'=1}^{\ell-1}\mathbf{x}_{\ell'}\right)\left(\sum_{\ell'=1}^{\ell-1}\mathbf{x}_{\ell'}\right)^\dagger \tilde{\mathbf{H}}^\dagger \,\middle|\, \hat{\mathbf{H}}, \mathbf{x}_1,\ldots,\mathbf{x}_{\ell-1}\right]}_{\substack{\text{residual interference} \\ \text{from decoded layers}}} + \underbrace{\mathrm{E}\left[\tilde{\mathbf{H}}\mathbf{Q}_\ell\tilde{\mathbf{H}}^\dagger \,\middle|\, \hat{\mathbf{H}}\right]}_{\substack{\text{CSIR imperfection} \\ \text{for layer } \ell}}$$

$$+ \underbrace{\hat{\mathbf{H}}\left(\mathbf{Q} - \sum_{\ell'=1}^{\ell}\mathbf{Q}_{\ell'}\right)\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}\left(\mathbf{Q} - \sum_{\ell'=1}^{\ell}\mathbf{Q}_{\ell'}\right)\tilde{\mathbf{H}}^\dagger \,\middle|\, \hat{\mathbf{H}}\right]}_{\text{interference from layers yet to be decoded}} + \underbrace{\frac{1}{\mathrm{snr}}\mathbf{I}}_{\substack{\text{additive} \\ \text{noise}}}$$

Towards the best rate-splitting allocation (1 of 3)

- Riemann integration:

$$\sum_{i=1}^{n} f(x_i)\Delta x_i \xrightarrow{n\to\infty} \int_{x_0}^{x_\infty} f(x)\,\mathrm{d}x$$

- Riemann-Stieltjes integration:

$$\sum_{i=1}^{n} f(x_i)g(\Delta x_i) \xrightarrow{n\to\infty} \int_{x_0}^{x_\infty} f(x)\,\mathrm{d}g(x)$$

Towards the best rate-splitting allocation (2 of 3)

Layering functions

A layering function $\mathbf{K}\colon [0;1] \to \mathbb{C}_{+}^{n_\mathsf{T} \times n_\mathsf{T}}$ has the following properties:

- continuity
- $0 \le \lambda_1 < \lambda_2 \le 1 \Rightarrow \mathbf{K}(\lambda_1) \overset{\neq}{\preceq} \mathbf{K}(\lambda_2)$
- $\mathbf{K}(0) = \mathbf{0}$
- $\mathbf{K}(1) = \mathbf{Q}$

Towards the best rate-splitting allocation (2 of 3)

Layering functions

A layering function $\mathbf{K}\colon [0;1] \rightarrow \mathbb{C}_{+}^{n_\mathsf{T} \times n_\mathsf{T}}$ has the following properties:

- continuity
- $0 \leq \lambda_1 < \lambda_2 \leq 1 \Rightarrow \mathbf{K}(\lambda_1) \overset{\neq}{\preceq} \mathbf{K}(\lambda_2)$
- $\mathbf{K}(0) = \mathbf{0}$
- $\mathbf{K}(1) = \mathbf{Q}$

Towards the best rate-splitting allocation (2 of 3)

Layering functions

A layering function $\mathbf{K}\colon [0;1] \to \mathbb{C}_+^{n_\mathsf{T}\times n_\mathsf{T}}$ has the following properties:

- continuity
- $0 \le \lambda_1 < \lambda_2 \le 1 \Rightarrow \mathbf{K}(\lambda_1) \overset{\neq}{\preceq} \mathbf{K}(\lambda_2)$
- $\mathbf{K}(0) = \mathbf{0}$
- $\mathbf{K}(1) = \mathbf{Q}$

Towards the best rate-splitting allocation (2 of 3)

Layering functions

A layering function $\mathbf{K}\colon [0;1] \to \mathbb{C}_{+}^{n_\mathsf{T} \times n_\mathsf{T}}$ has the following properties:

- continuity
- $0 \leq \lambda_1 < \lambda_2 \leq 1 \Rightarrow \mathbf{K}(\lambda_1) \overset{\neq}{\preceq} \mathbf{K}(\lambda_2)$
- $\mathbf{K}(0) = \mathbf{0}$
- $\mathbf{K}(1) = \mathbf{Q}$

Towards the best rate-splitting allocation (2 of 3)

Layering functions

A layering function $\mathbf{K}\colon [0;1] \to \mathbb{C}_+^{n_\mathrm{T} \times n_\mathrm{T}}$ has the following properties:

- continuity
- $0 \le \lambda_1 < \lambda_2 \le 1 \Rightarrow \mathbf{K}(\lambda_1) \overset{\neq}{\preceq} \mathbf{K}(\lambda_2)$
- $\mathbf{K}(0) = \mathbf{0}$
- $\mathbf{K}(1) = \mathbf{Q}$

Given a pair $(\mathbf{K}, L)$, a tuple $(\mathbf{Q}_1, \ldots, \mathbf{Q}_L)$ is uniquely determined via

$$\mathbf{Q}_\ell = \mathbf{K}\left(\tfrac{\ell}{L}\right) - \mathbf{K}\left(\tfrac{\ell-1}{L}\right), \qquad \ell = 1, \ldots, L$$

such that

$$\boxed{R_\mathrm{RS}(\mathbf{Q}_1, \ldots, \mathbf{Q}_L) \triangleq R_\mathrm{RS}(\mathbf{K}, L)}$$

Towards the best rate-splitting allocation (3 of 3)

The best rate-splitting bound is

$$R_{\mathsf{RS}}^{\star\star} = \sup_{L} \sup_{\substack{\mathbf{Q}_1,\ldots,\mathbf{Q}_L \\ \sum_{\ell=1}^{L} \mathbf{Q}_\ell = \mathbf{Q}}} R_{\mathsf{RS}}(\mathbf{Q}_1, \ldots, \mathbf{Q}_L)$$

Towards the best rate-splitting allocation (3 of 3)

The best rate-splitting bound is

$$R_{\mathsf{RS}}^{\star\star} = \sup_{L} \sup_{\substack{\mathbf{Q}_1,\ldots,\mathbf{Q}_L \\ \sum_{\ell=1}^{L} \mathbf{Q}_\ell = \mathbf{Q}}} R_{\mathsf{RS}}(\mathbf{Q}_1, \ldots, \mathbf{Q}_L) = \sup_{L} \sup_{\mathbf{K}} R_{\mathsf{RS}}(\mathbf{K}, L)$$

Towards the best rate-splitting allocation (3 of 3)

The best rate-splitting bound is

$$R_{\mathrm{RS}}^{\star\star} = \sup_{L} \sup_{\substack{\mathbf{Q}_1,\ldots,\mathbf{Q}_L \\ \sum_{\ell=1}^{L} \mathbf{Q}_\ell = \mathbf{Q}}} R_{\mathrm{RS}}(\mathbf{Q}_1, \ldots, \mathbf{Q}_L) = \sup_{\mathbf{K}} \sup_{L} R_{\mathrm{RS}}(\mathbf{K}, L)$$

Towards the best rate-splitting allocation (3 of 3)

The best rate-splitting bound is

$$R_{\mathrm{RS}}^{\star\star} = \sup_{L} \sup_{\substack{\mathbf{Q}_1,\ldots,\mathbf{Q}_L \\ \sum_{\ell=1}^{L} \mathbf{Q}_\ell = \mathbf{Q}}} R_{\mathrm{RS}}(\mathbf{Q}_1, \ldots, \mathbf{Q}_L) = \sup_{\mathbf{K}} \underbrace{\sup_{L} R_{\mathrm{RS}}(\mathbf{K}, L)}_{R_{\mathrm{RS}}^{\star}(\mathbf{K})}$$

Towards the best rate-splitting allocation (3 of 3)

The best rate-splitting bound is

$$R_{\mathrm{RS}}^{\star\star} = \sup_{L} \sup_{\substack{\mathbf{Q}_1,\ldots,\mathbf{Q}_L \\ \sum_{\ell=1}^{L}\mathbf{Q}_\ell=\mathbf{Q}}} R_{\mathrm{RS}}(\mathbf{Q}_1,\ldots,\mathbf{Q}_L) = \sup_{\mathbf{K}} \underbrace{\sup_{L} R_{\mathrm{RS}}(\mathbf{K},L)}_{R_{\mathrm{RS}}^{\star}(\mathbf{K})}$$

Theorem ("Infinite" layering)

$R_{\mathrm{RS}}^{\star}(\mathbf{K})$ is achieved for $L \to \infty$ and given by the Riemann-Stieltjes integral

$$R_{\mathrm{RS}}^{\star}(\mathbf{K}) = \lim_{L\to\infty} R_{\mathrm{RS}}(\mathbf{K},L) = \int_0^1 \mathrm{E}\left[\mathrm{tr}\left\{\hat{\mathbf{H}}^\dagger \boldsymbol{\Gamma}(\lambda)^{-1}\hat{\mathbf{H}}\,\mathrm{d}\mathbf{K}(\lambda)\right\}\right]$$

with $\boldsymbol{\xi} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{I})$ and $\boldsymbol{\Gamma}(\lambda) =$

$$\underbrace{\mathrm{E}\left[\tilde{\mathbf{H}}\mathbf{K}(\lambda)^{\frac{1}{2}}\boldsymbol{\xi}\boldsymbol{\xi}^\dagger\mathbf{K}(\lambda)^{\frac{1}{2}}\tilde{\mathbf{H}}^\dagger \mid \boldsymbol{\xi}\right]}_{\text{residual interference from decoded layers}} + \underbrace{\hat{\mathbf{H}}(\mathbf{Q}-\mathbf{K}(\lambda))\hat{\mathbf{H}}^\dagger + \mathrm{E}\left[\tilde{\mathbf{H}}(\mathbf{Q}-\mathbf{K}(\lambda))\tilde{\mathbf{H}}^\dagger\right]}_{\text{interference from layers yet to be decoded}} + \underbrace{\frac{1}{\mathrm{snr}}\mathbf{I}_{n_R}}_{\text{additive noise}}$$

Towards the best rate-splitting allocation (3 of 3)

The best rate-splitting bound is

$$R_{\mathrm{RS}}^{\star\star} = \sup_{L} \sup_{\substack{\mathbf{Q}_1,\dots,\mathbf{Q}_L \\ \sum_{\ell=1}^{L}\mathbf{Q}_\ell = \mathbf{Q}}} R_{\mathrm{RS}}(\mathbf{Q}_1,\dots,\mathbf{Q}_L) = \sup_{\mathbf{K}} \underbrace{\sup_{L} R_{\mathrm{RS}}(\mathbf{K},L)}_{R_{\mathrm{RS}}^{\star}(\mathbf{K})}$$

Theorem ("Infinite" layering)

$R_{\mathrm{RS}}^{\star}(\mathbf{K})$ is achieved for $L \to \infty$ and given by the Riemann-Stieltjes integral

$$R_{\mathrm{RS}}^{\star}(\mathbf{K}) = \lim_{L\to\infty} R_{\mathrm{RS}}(\mathbf{K},L) = \int_0^1 \mathrm{E}\Big[\mathrm{tr}\Big\{\hat{\mathbf{H}}^{\dagger}\boldsymbol{\Gamma}(\lambda)^{-1}\hat{\mathbf{H}}\,\mathrm{d}\mathbf{K}(\lambda)\Big\}\Big]$$

with $\boldsymbol{\xi} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{I})$ and $\boldsymbol{\Gamma}(\lambda) =$

$$\underbrace{\mathrm{E}\Big[\tilde{\mathbf{H}}\mathbf{K}(\lambda)^{\frac{1}{2}}\boldsymbol{\xi}\boldsymbol{\xi}^{\dagger}\mathbf{K}(\lambda)^{\frac{1}{2}}\tilde{\mathbf{H}}^{\dagger} \;\Big|\; \boldsymbol{\xi}\Big]}_{\text{residual interference from decoded layers}} + \underbrace{\hat{\mathbf{H}}(\mathbf{Q}-\mathbf{K}(\lambda))\hat{\mathbf{H}}^{\dagger} + \mathrm{E}\Big[\tilde{\mathbf{H}}(\mathbf{Q}-\mathbf{K}(\lambda))\tilde{\mathbf{H}}^{\dagger}\Big]}_{\text{interference from layers yet to be decoded}} + \underbrace{\tfrac{1}{\mathrm{snr}}\mathbf{I}_{n_R}}_{\text{additive noise}}$$

Remarks:

- Compact notation for $\int_0^1 \mathrm{tr}\big[\mathbf{A}(\lambda)\,\mathrm{d}\mathbf{B}(\lambda)\big] = \sum_{i,j}\int_0^1 A_{i,j}(\lambda)\,\mathrm{d}B_{j,i}(\lambda)$
- For continuously differentiable $\mathbf{K}$, we have $\mathrm{d}\mathbf{K}(\lambda) = \mathbf{K}'(\lambda)\,\mathrm{d}\lambda$ and we obtain a Riemann integral
- R_{WCN} can be recovered via Jensen's inequality

Proof sketch of the infinite-layering formula

First step ("adding layers always helps"):

$$R_{\mathsf{RS}}(\dots, \mathbf{Q}_\ell + \mathbf{Q}_{\ell+1}, \dots) \leq R_{\mathsf{RS}}(\dots, \mathbf{Q}_\ell, \mathbf{Q}_{\ell+1}, \dots) \Rightarrow \boxed{R^{\star}_{\mathsf{RS}}(\mathbf{K}) = \lim_{L\to\infty} R_{\mathsf{RS}}(\mathbf{K}, L)}$$

Proof sketch of the infinite-layering formula

First step ("adding layers always helps"):

$$R_{\mathrm{RS}}(\dots, \mathbf{Q}_\ell + \mathbf{Q}_{\ell+1}, \dots) \leq R_{\mathrm{RS}}(\dots, \mathbf{Q}_\ell, \mathbf{Q}_{\ell+1}, \dots) \Rightarrow \boxed{R^{\star}_{\mathrm{RS}}(\mathbf{K}) = \lim_{L \to \infty} R_{\mathrm{RS}}(\mathbf{K}, L)}$$

Second step ("from sum to integral"):

Denote $\Delta \mathbf{K}_\ell = \mathbf{K}\big(\frac{\ell}{L}\big) - \mathbf{K}\big(\frac{\ell-1}{L}\big) = \mathbf{Q}_\ell$.

Proof sketch of the infinite-layering formula

First step ("adding layers always helps"):

$$R_{\text{RS}}(\dots,\mathbf{Q}_\ell+\mathbf{Q}_{\ell+1},\dots) \leq R_{\text{RS}}(\dots,\mathbf{Q}_\ell,\mathbf{Q}_{\ell+1},\dots) \Rightarrow \boxed{R^\star_{\text{RS}}(\mathbf{K}) = \lim_{L\to\infty} R_{\text{RS}}(\mathbf{K},L)}$$

Second step ("from sum to integral"):

Denote $\Delta\mathbf{K}_\ell = \mathbf{K}\big(\tfrac{\ell}{L}\big) - \mathbf{K}\big(\tfrac{\ell-1}{L}\big) = \mathbf{Q}_\ell$.

$$\begin{aligned} R_{\text{RS}}(\mathbf{K},L) &= \sum_{\ell=1}^{L} \mathrm{E}\Big[\log\det\Big(\mathbf{I} + \hat{\mathbf{H}}^\dagger \boldsymbol{\Gamma}_\ell^{-1}\hat{\mathbf{H}}\Delta\mathbf{K}_\ell\Big)\Big] \\ &= \sum_{\ell=1}^{L}\Big[\mathrm{E}\big[\mathrm{tr}\{\hat{\mathbf{H}}^\dagger\boldsymbol{\Gamma}_\ell^{-1}\hat{\mathbf{H}}\Delta\mathbf{K}_\ell\}\big] + \mathcal{O}(\Delta\mathbf{K}_\ell^2)\Big] \\ &\xrightarrow{L\to\infty} \boxed{\int_0^1 \mathrm{E}\Big[\mathrm{tr}\Big\{\hat{\mathbf{H}}^\dagger\boldsymbol{\Gamma}(\lambda)^{-1}\hat{\mathbf{H}}\,\mathrm{d}\mathbf{K}(\lambda)\Big\}\Big]} + \underbrace{\int_0^1 \mathrm{d}\lambda^2}_{=0} \end{aligned}$$

Three additional properties of layering functions

Continuity

Provided that $\mathsf{E}[\|\hat{\mathbf{H}}\|_{\mathsf{F}}^4] < \infty$, the function $R_{\mathsf{RS}}^{\star}(\mathbf{K})$ is uniformly continuous in $\mathbf{K}$:

$$\forall \epsilon > 0 \colon \exists \delta > 0 \colon \|\mathbf{K}_1 - \mathbf{K}_2\|_\infty \leq \delta \Rightarrow \left| R_{\mathsf{RS}}^{\star}(\mathbf{K}_1) - R_{\mathsf{RS}}^{\star}(\mathbf{K}_2) \right| \leq \epsilon$$

Three additional properties of layering functions

Continuity

Provided that $\mathsf{E}[\|\hat{\mathbf{H}}\|_{\mathsf{F}}^4] < \infty$, the function $R_{\mathsf{RS}}^{\star}(\mathbf{K})$ is uniformly continuous in $\mathbf{K}$:

$$\forall \epsilon > 0\colon \exists \delta > 0\colon \|\mathbf{K}_1 - \mathbf{K}_2\|_\infty \le \delta \Rightarrow \left| R_{\mathsf{RS}}^{\star}(\mathbf{K}_1) - R_{\mathsf{RS}}^{\star}(\mathbf{K}_2) \right| \le \epsilon$$

Continuous differentiable layering functions

Provided that $\mathsf{E}[\|\hat{\mathbf{H}}\|_{\mathsf{F}}^4] < \infty$, the best rate-splitting bound $R_{\mathsf{RS}}^{\star\star}$ is the supremum of $R_{\mathsf{RS}}^{\star}(\mathbf{K})$ over continuously differentiable layering functions, i.e.,

$$R_{\mathsf{RS}}^{\star\star} = \sup_{\mathbf{K}} R_{\mathsf{RS}}^{\star}(\mathbf{K}) = \sup_{\text{differentiable } \mathbf{K}} R_{\mathsf{RS}}^{\star}(\mathbf{K})$$

Three additional properties of layering functions

Continuity

Provided that $\mathrm{E}[\|\hat{\mathbf{H}}\|_\mathrm{F}^4] < \infty$, the function $R_\mathrm{RS}^\star(\mathbf{K})$ is uniformly continuous in $\mathbf{K}$:

$$\forall \epsilon > 0 \colon \exists \delta > 0 \colon \|\mathbf{K}_1 - \mathbf{K}_2\|_\infty \leq \delta \Rightarrow \left| R_\mathrm{RS}^\star(\mathbf{K}_1) - R_\mathrm{RS}^\star(\mathbf{K}_2) \right| \leq \epsilon$$

Continuous differentiable layering functions

Provided that $\mathrm{E}[\|\hat{\mathbf{H}}\|_\mathrm{F}^4] < \infty$, the best rate-splitting bound $R_\mathrm{RS}^{\star\star}$ is the supremum of $R_\mathrm{RS}^\star(\mathbf{K})$ over continuously differentiable layering functions, i.e.,

$$R_\mathrm{RS}^{\star\star} = \sup_{\mathbf{K}} R_\mathrm{RS}^\star(\mathbf{K}) = \sup_{\text{differentiable } \mathbf{K}} R_\mathrm{RS}^\star(\mathbf{K})$$

Optimal layering

Provided that $\mathrm{E}[\|\hat{\mathbf{H}}\|_\mathrm{F}^4] < \infty$, there exists an optimal layering function

$$\mathbf{K}^\star = \operatorname*{argmax}_{\mathbf{K}} R_\mathrm{RS}^\star(\mathbf{K})$$

such that $R_\mathrm{RS}^{\star\star} = R_\mathrm{RS}^\star(\mathbf{K}^\star)$.

Example 1: SIMO with white channel error

Assumptions:

- $n_\mathrm{T} = 1$, $(\hat{\mathbf{H}}, \tilde{\mathbf{H}}) \to (\hat{\mathbf{h}}, \tilde{\mathbf{h}})$, $\mathbf{K}(\lambda) \to k(\lambda)$
- $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are mutually independent
- $\mathrm{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] = \tilde{V}\mathbf{I}_{n_\mathrm{R}}$ (spatially white estimation error)

Example 1: SIMO with white channel error

Assumptions:

- $n_\mathsf{T} = 1$, $(\hat{\mathbf{H}}, \tilde{\mathbf{H}}) \to (\hat{\mathbf{h}}, \tilde{\mathbf{h}})$, $\mathbf{K}(\lambda) \to k(\lambda)$
- $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are mutually independent
- $\mathsf{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] = \tilde{V}\mathbf{I}_{n_\mathsf{R}}$ (spatially white estimation error)

After some algebra:

$$R^\star_{\mathsf{RS}}(k) = \int_0^1 \mathsf{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - k(\lambda) + \beta(k(\lambda))\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}k(\lambda)$$

where $\beta(k) = \tilde{V}\big((\Xi_1 - 1)k + 1\big) + \mathsf{snr}^{-1}$ with $\Xi_1 \sim \mathrm{gamma}(1, 1)$

Example 1: SIMO with white channel error

Assumptions:

- $n_\mathsf{T} = 1$, $(\hat{\mathbf{H}}, \tilde{\mathbf{H}}) \to (\hat{\mathbf{h}}, \tilde{\mathbf{h}})$, $\mathbf{K}(\lambda) \to k(\lambda)$
- $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are mutually independent
- $\mathsf{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] = \tilde{V}\mathbf{I}_{n_\mathsf{R}}$ (spatially white estimation error)

After some algebra:

$$R^\star_{\mathsf{RS}}(k) = \int_0^1 \mathsf{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - k(\lambda) + \beta(k(\lambda))\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}k(\lambda)$$

where $\beta(k) = \tilde{V}\big((\Xi_1 - 1)k + 1\big) + \mathsf{snr}^{-1}$ with $\Xi_1 \sim \mathrm{gamma}(1,1)$

Example 1: SIMO with white channel error

Assumptions:

- $n_\mathsf{T} = 1$, $(\hat{\mathbf{H}}, \tilde{\mathbf{H}}) \to (\hat{\mathbf{h}}, \tilde{\mathbf{h}})$, $\mathbf{K}(\lambda) \to k(\lambda)$
- $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are mutually independent
- $\mathsf{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] = \tilde{V}\mathbf{I}_{n_\mathsf{R}}$ (spatially white estimation error)

After some algebra:

$$R^\star_{\mathsf{RS}}(k) = \int_0^1 \mathsf{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - k(\lambda) + \beta(k(\lambda))\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}k(\lambda)$$
$$= \int_0^1 \mathsf{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - \lambda + \beta(\lambda)\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}\lambda$$

where $\beta(k) = \tilde{V}\big((\Xi_1 - 1)k + 1\big) + \mathsf{snr}^{-1}$ with $\Xi_1 \sim \mathrm{gamma}(1, 1)$

Example 1: SIMO with white channel error

Assumptions:

- $n_\mathrm{T} = 1$, $(\hat{\mathbf{H}}, \tilde{\mathbf{H}}) \to (\hat{\mathbf{h}}, \tilde{\mathbf{h}})$, $\mathbf{K}(\lambda) \to k(\lambda)$
- $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are mutually independent
- $\mathrm{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] = \tilde{V}\mathbf{I}_{n_\mathrm{R}}$ (spatially white estimation error)

After some algebra:

$$R^\star_{\mathrm{RS}}(k) = \int_0^1 \mathrm{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - k(\lambda) + \beta(k(\lambda))\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}k(\lambda)$$
$$= \int_0^1 \mathrm{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - \lambda + \beta(\lambda)\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}\lambda$$

where $\beta(k) = \tilde{V}\big((\Xi_1 - 1)k + 1\big) + \mathrm{snr}^{-1}$ with $\Xi_1 \sim \mathrm{gamma}(1,1)$

$$\boxed{R^\star_{\mathrm{RS}} = \mathrm{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{\tilde{V}(\Xi_1 - 1) - \|\hat{\mathbf{h}}\|_2^2} \log\left(1 + \frac{\tilde{V}(\Xi_1 - 1) - \|\hat{\mathbf{h}}\|_2^2}{\|\hat{\mathbf{h}}\|_2^2 + \tilde{V} + \mathrm{snr}^{-1}}\right)\right]}$$

Example 1: SIMO with white channel error

Assumptions:

- $n_\mathrm{T} = 1$, $(\hat{\mathbf{H}}, \tilde{\mathbf{H}}) \to (\hat{\mathbf{h}}, \tilde{\mathbf{h}})$, $\mathbf{K}(\lambda) \to k(\lambda)$
- $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are mutually independent
- $\mathrm{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] = \tilde{V}\mathbf{I}_{n_\mathrm{R}}$ (spatially white estimation error)

After some algebra:

$$R^\star_\mathrm{RS}(k) = \int_0^1 \mathrm{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - k(\lambda) + \beta(k(\lambda))\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}k(\lambda)$$
$$= \int_0^1 \mathrm{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{1 - \lambda + \beta(\lambda)\|\hat{\mathbf{h}}\|_2^2}\right] \mathrm{d}\lambda$$

where $\beta(k) = \tilde{V}\big((\Xi_1 - 1)k + 1\big) + \mathrm{snr}^{-1}$ with $\Xi_1 \sim \mathrm{gamma}(1,1)$

$$\boxed{R^\star_\mathrm{RS} = \mathrm{E}\left[\frac{\|\hat{\mathbf{h}}\|_2^2}{\tilde{V}(\Xi_1 - 1) - \|\hat{\mathbf{h}}\|_2^2} \log\left(1 + \frac{\tilde{V}(\Xi_1 - 1) - \|\hat{\mathbf{h}}\|_2^2}{\|\hat{\mathbf{h}}\|_2^2 + \tilde{V} + \mathrm{snr}^{-1}}\right)\right]}$$

Layering function has disappeared!

Example 2: MIMO with i.i.d. fading and two exemplary layerings (1 of 2)

Assumptions:

- $\text{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_\text{R} n_\text{T}})$ and $\text{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_\text{R} n_\text{T}})$ are independent
- $\mathbf{Q} = \frac{1}{n_\text{T}}\mathbf{I}_{n_\text{T}}$

Example 2: MIMO with i.i.d. fading and two exemplary layerings (1 of 2)

Assumptions:

- $\mathrm{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ and $\mathrm{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ are independent
- $\mathbf{Q} = \frac{1}{n_\mathrm{T}}\mathbf{I}_{n_\mathrm{T}}$

Levelled layering $\mathbf{K}_{\mathrm{lev}}$

$$\mathbf{K}_{\mathrm{lev}}(\lambda) = \lambda\mathbf{Q} = \lambda\frac{1}{n_\mathrm{T}}\mathbf{I}_{n_\mathrm{T}}$$

Example 2: MIMO with i.i.d. fading and two exemplary layerings (1 of 2)

Assumptions:

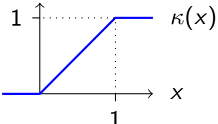
- $\mathrm{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_{\mathrm{R}}n_{\mathrm{T}}})$ and $\mathrm{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_{\mathrm{R}}n_{\mathrm{T}}})$ are independent
- $\mathbf{Q} = \frac{1}{n_{\mathrm{T}}}\mathbf{I}_{n_{\mathrm{T}}}$

Levelled layering $\mathbf{K}_{\mathrm{lev}}$

$$\mathbf{K}_{\mathrm{lev}}(\lambda) = \lambda\mathbf{Q} = \lambda\frac{1}{n_{\mathrm{T}}}\mathbf{I}_{n_{\mathrm{T}}}$$

Staggered layering $\mathbf{K}_{\mathrm{stag}}$

$$\mathbf{K}_{\mathrm{stag}}(\lambda) = \frac{1}{n_{\mathrm{T}}}\begin{bmatrix} \kappa(\lambda n_{\mathrm{T}}) & & & 0 \\ & \kappa(\lambda n_{\mathrm{T}} - 1) & & \\ & & \ddots & \\ 0 & & & \kappa(\lambda n_{\mathrm{T}} - n_{\mathrm{T}} + 1) \end{bmatrix}$$

where 

Example 2: MIMO with i.i.d. fading and two exemplary layerings (2 of 2)

- **Levelled layering:**

$$R^{\star}_{\mathsf{RS}}(\mathbf{K}_{\mathsf{lev}}) = \frac{1}{n_{\mathsf{T}}} \int_0^1 \mathsf{E}\left[\mathrm{tr}\left\{\hat{\mathbf{H}}^{\dagger} \boldsymbol{\Gamma}_{\mathsf{lev}}(\lambda)^{-1} \hat{\mathbf{H}}\right\}\right] \mathrm{d}\lambda$$

where

$$\boldsymbol{\Gamma}_{\mathsf{lev}}(\lambda) = \left(\lambda \tilde{V} \frac{\Xi_{n_{\mathsf{T}}}}{n_{\mathsf{T}}} + \tilde{V}(1-\lambda) + \frac{1}{\mathsf{snr}}\right) \mathbf{I}_{n_{\mathsf{R}}} + (1-\lambda)\frac{1}{n_{\mathsf{T}}} \hat{\mathbf{H}}\hat{\mathbf{H}}^{\dagger}$$

and $\Xi_{n_{\mathsf{T}}} \sim \mathrm{gamma}(n_{\mathsf{T}}, 1)$

- **Staggered layering:**

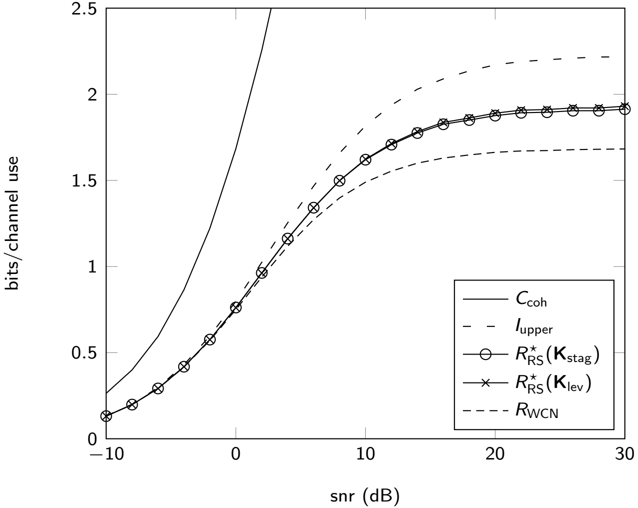
$$R^{\star}_{\mathsf{RS}}(\mathbf{K}_{\mathsf{stag}}) = \sum_{i=1}^{n_{\mathsf{T}}} \int_0^1 \mathsf{E}\left[\frac{\hat{\mathbf{h}}_i^{\dagger} \mathbf{A}_i(\lambda)^{-1} \hat{\mathbf{h}}_i}{1 + (1-\lambda)\hat{\mathbf{h}}_i^{\dagger} \mathbf{A}_i(\lambda)^{-1} \hat{\mathbf{h}}_i}\right] \mathrm{d}\lambda$$

where

$$\mathbf{A}_i(\lambda) = \left(\tilde{V}\left(\Xi_{i-1}^{(1)} + \lambda \Xi_1^{(2)} + 1 - \lambda + n_{\mathsf{T}} - i\right) + \frac{n_{\mathsf{T}}}{\mathsf{snr}}\right) \mathbf{I}_{n_{\mathsf{R}}} + \hat{\mathbf{H}}_{(i+1):n_{\mathsf{T}}} \hat{\mathbf{H}}^{\dagger}_{(i+1):n_{\mathsf{T}}}$$

and the $\Xi_i^{(j)} \sim \mathrm{gamma}(i, 1)$ are independent

Numerical example



Capacity and MI bounds for a 2×2 channel with $\hat{\mathbf{H}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \frac{1}{2}\mathbf{I}_{n_\text{R} n_\text{T}})$ and $\tilde{\mathbf{H}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \frac{1}{2}\mathbf{I}_{n_\text{R} n_\text{T}})$

Large antenna arrays (1 of 2)

Assumptions:

- $\text{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_{\mathsf{R}}n_{\mathsf{T}}})$ and $\text{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_{\mathsf{R}}n_{\mathsf{T}}})$ independent
- $\mathbf{Q} = \frac{1}{n_{\mathsf{T}}}\mathbf{I}_{n_{\mathsf{T}}}$

Large antenna arrays (1 of 2)

Assumptions:

- $\mathrm{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ and $\mathrm{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ independent
- $\mathbf{Q} = \frac{1}{n_\mathrm{T}}\mathbf{I}_{n_\mathrm{T}}$

Large transmit array ($n_\mathrm{T} \to \infty$)

Consider a sequence $(\mathbf{K}_{n_\mathrm{T}})_{n_\mathrm{T} \in \mathbb{N}}$ of $n_\mathrm{T} \times n_\mathrm{T}$ layering functions. Then

$$\lim_{n_\mathrm{T} \to \infty} \left\{ R^\star_\mathrm{RS}(\mathbf{K}_{n_\mathrm{T}}) - R_\mathrm{WCN} \right\} = 0$$

Large antenna arrays (1 of 2)

Assumptions:

- $\text{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_R n_T})$ and $\text{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_R n_T})$ independent
- $\mathbf{Q} = \frac{1}{n_T}\mathbf{I}_{n_T}$

Large transmit array ($n_T \to \infty$)

Consider a sequence $(\mathbf{K}_{n_T})_{n_T \in \mathbb{N}}$ of $n_T \times n_T$ layering functions. Then

$$\lim_{n_T \to \infty} \left\{ R^{\star}_{RS}(\mathbf{K}_{n_T}) - R_{WCN} \right\} = 0$$

Large receive array ($n_R \to \infty$)

$$\lim_{n_R \to \infty} \left\{ R^{\star}_{RS}(\mathbf{K}) - R_{WCN} \right\} = n_T \, \mathrm{E}\left[\log \left(\frac{\tilde{V} + \text{snr}^{-1}}{\tilde{V}\frac{\Xi_{n_T}}{n_T} + \text{snr}^{-1}} \right) \right]$$

Large antenna arrays (1 of 2)

Assumptions:

- $\text{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}\mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ and $\text{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}\mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ independent
- $\mathbf{Q} = \frac{1}{n_\mathrm{T}}\mathbf{I}_{n_\mathrm{T}}$

Large transmit array ($n_\mathrm{T} \to \infty$)

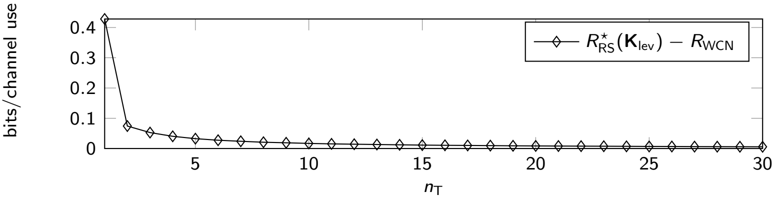
Consider a sequence $(\mathbf{K}_{n_\mathrm{T}})_{n_\mathrm{T} \in \mathbb{N}}$ of $n_\mathrm{T} \times n_\mathrm{T}$ layering functions. Then

$$\lim_{n_\mathrm{T} \to \infty} \left\{ R^{\star}_{\mathrm{RS}}(\mathbf{K}_{n_\mathrm{T}}) - R_{\mathrm{WCN}} \right\} = 0$$

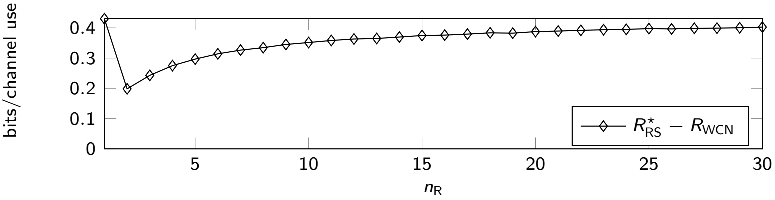
Large receive array ($n_\mathrm{R} \to \infty$)

$$\lim_{n_\mathrm{R} \to \infty} \left\{ R^{\star}_{\mathrm{RS}}(\mathbf{K}) - R_{\mathrm{WCN}} \right\} = n_\mathrm{T}\, \mathrm{E}\left[\log \left(\frac{\tilde{V} + \mathrm{snr}^{-1}}{\tilde{V}\frac{\Xi_{n_\mathrm{T}}}{n_\mathrm{T}} + \mathrm{snr}^{-1}} \right) \right] \xrightarrow{n_\mathrm{T} \to \infty} \frac{1}{2}$$

Large antenna arrays (2 of 2)



Bound difference $R^\star_{RS}(\mathbf{K}_{lev}) - R_{WCN}$ for a MISO channel ($n_R = 1$)



Bound difference $R^\star_{RS} - R_{WCN}$ for a SIMO channel ($n_T = 1$)

Asymptotically perfect CSI

Assumptions:

- $\text{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}_{\text{snr}}\mathbf{I}_{n_R n_T})$ and $\text{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}_{\text{snr}}\mathbf{I}_{n_R n_T})$ are mutually independent
- We have an arbitrary family of layering functions $\mathbf{K}_{\text{snr}}$

Asymptotically perfect CSI

Assumptions:

- $\mathrm{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}_{\mathsf{snr}}\mathbf{I}_{n_{\mathrm{R}} n_{\mathrm{T}}})$ and $\mathrm{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}_{\mathsf{snr}}\mathbf{I}_{n_{\mathrm{R}} n_{\mathrm{T}}})$ are mutually independent
- We have an arbitrary family of layering functions $\mathbf{K}_{\mathsf{snr}}$

Asymptotically perfect CSI

Assumptions:

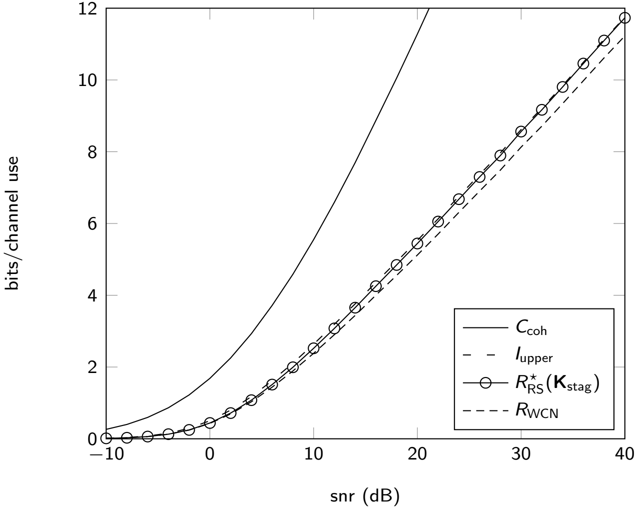
- $\mathrm{vec}(\hat{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{V}_{\mathrm{snr}} \mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ and $\mathrm{vec}(\tilde{\mathbf{H}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tilde{V}_{\mathrm{snr}} \mathbf{I}_{n_\mathrm{R} n_\mathrm{T}})$ are mutually independent
- We have an arbitrary family of layering functions $\mathbf{K}_{\mathrm{snr}}$

Asymptotic tightness

If $\lim_{\mathrm{snr}\to\infty} \tilde{V}_{\mathrm{snr}} = 0$, then $R^{\star}_{\mathrm{RS}}(\mathbf{K}_{\mathrm{snr}})$ is asymptotically tight:

$$\lim_{\mathrm{snr}\to\infty} \left\{ I(\mathbf{x}_\mathrm{G}; \mathbf{y} | \hat{\mathbf{H}}) - R^{\star}_{\mathrm{RS}}(\mathbf{K}_{\mathrm{snr}}) \right\} = 0$$

Asymptotically perfect CSI: numerical example



Capacity and MI bounds for a 2×2 channel with $\tilde{V}_{snr} = (1 + snr)^{-1/2}$

Introduction

- Pilot-aided communication achieves good performance-complexity tradeoff

Introduction

- Pilot-aided communication achieves good performance-complexity tradeoff
- Forward pilot symbols transmitted periodically

Introduction

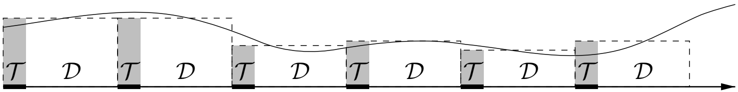
- Pilot-aided communication achieves good performance-complexity tradeoff
- Forward pilot symbols transmitted periodically
- Imperfect CSIR for the decoding task

Introduction

- Pilot-aided communication achieves good performance-complexity tradeoff
- Forward pilot symbols transmitted periodically
- Imperfect CSIR for the decoding task
- Variables to be optimized: **linear precoder** and **training sequence**

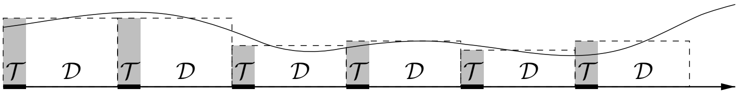
Rayleigh block fading

Block-fading with training preambles:



Rayleigh block fading

Block-fading with training preambles:



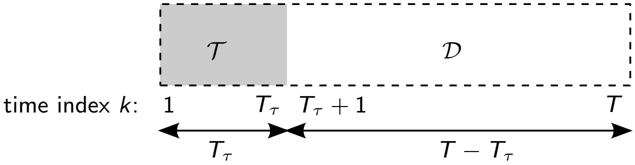
Transmit-side spatial correlation:

$$\mathbf{H} = \mathbf{W}\mathbf{R}^{\frac{1}{2}}$$

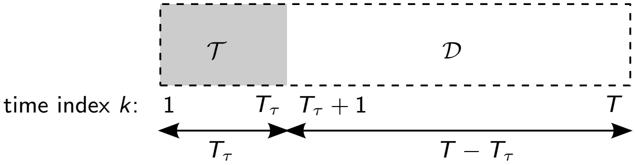
where

- $[\mathbf{W}]_{i,j}$ are random i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$
- $\mathbf{R} = \mathrm{E}[\mathbf{H}^{\dagger}\mathbf{H}]$ is the channel correlation

System model



System model



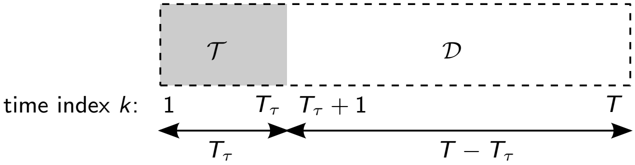
Training phase $\mathcal{T}$:

$$\mathbf{y}_\tau^{(k)} = \mathbf{H}\mathbf{t}^{(k)} + \mathbf{z}^{(k)}$$

Data transmission phase $\mathcal{D}$:

$$\mathbf{y}^{(k)} = \mathbf{H}\mathbf{F}\mathbf{x}^{(k)} + \mathbf{z}^{(k)}$$

System model



Training phase $\mathcal{T}$:

$$\mathbf{y}_\tau^{(k)} = \mathbf{H}\mathbf{t}^{(k)} + \mathbf{z}^{(k)}$$

Data transmission phase $\mathcal{D}$:

$$\mathbf{y}^{(k)} = \mathbf{H}\mathbf{F}\mathbf{x}^{(k)} + \mathbf{z}^{(k)}$$

Two matrix variables to be optimized:

- *pilot Gram* $\mathbf{P} = \left[\mathbf{t}^{(1)} \dots \mathbf{t}^{(T_\tau)}\right] \begin{bmatrix} \mathbf{t}^{(1)\dagger} \\ \vdots \\ \mathbf{t}^{(T_\tau)\dagger} \end{bmatrix}$
- $\mathrm{tr}(\mathbf{P})$ is *training energy*
- $\mathrm{rank}(\mathbf{P}) = T_\tau$ is *training duration*

- *transmit covariance* $\mathbf{Q} = \mathbf{F}\mathbf{F}^\dagger$
- $\mathrm{tr}(\mathbf{Q})$ is *transmit power*
- $\mathrm{rank}(\mathbf{Q})$ is *number of streams*

Worst-case noise lower bound

The WCN bound reads as

$$\begin{aligned} I\left(\mathbf{x}_G; \mathbf{y} \middle| \hat{\mathbf{H}}\right) &\geq \mathrm{E}_{\mathbf{W}}\left[\log\det\left(\mathbf{I} + \mathbf{W} \underbrace{\frac{\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{R} - (\mathbf{R}^{-1} + \mathbf{P})^{-1}\right)\mathbf{Q}^{\frac{1}{2}}}{1 + \mathrm{tr}\left(\mathbf{Q}(\mathbf{R}^{-1} + \mathbf{P})^{-1}\right)}}_{\mathbf{S}} \mathbf{W}^{\dagger}\right)\right] \\ &= \mathrm{E}_{\mathbf{W}}\left[\log\det(\mathbf{I} + \mathbf{W}\mathbf{S}\mathbf{W}^{\dagger})\right] \\ &= R_{\mathrm{WCN}}(\mathbf{S}) \\ &= R_{\mathrm{WCN}}(\mathbf{s}) \end{aligned}$$

and is a function of $\mathbf{s} = \boldsymbol{\lambda}(\mathbf{S})$.

Utility functions I

Definition

$$
\begin{aligned}
f: \quad \mathbf{S} &\mapsto f(\mathbf{S}) \\
\mathbb{C}_{+}^{n_T \times n_T} &\rightarrow \mathbb{R}
\end{aligned}
$$

Utility functions I

Definition

$$
\begin{array}{rccc}
f: & \mathbf{S} & \mapsto & f(\mathbf{S}) \\
 & \mathbb{C}_{+}^{n_T \times n_T} & \rightarrow & \mathbb{R}
\end{array}
$$

A utility function f

- *depends only on the unordered eigenvalues* $\mathbf{s} = \boldsymbol{\lambda}(\mathbf{S})$

Utility functions I

Definition

$$\begin{array}{rccc} f: & \mathbf{S} & \mapsto & f(\mathbf{S}) \\ & \mathbb{C}_+^{n_T \times n_T} & \to & \mathbb{R} \end{array}$$

A utility function f

- *depends only on the unordered eigenvalues* $\mathbf{s} = \boldsymbol{\lambda}(\mathbf{S})$
- *is monotonic in the vector* $\mathbf{s}$, *that is,* $\mathbf{s} \leq \mathbf{s}' \Rightarrow f(\mathbf{s}) \leq f(\mathbf{s}')$

Utility functions I

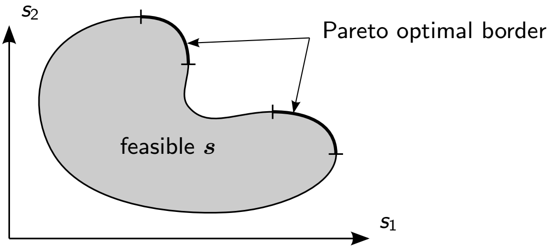
Definition

$$\begin{array}{rccc} f: & \mathbf{S} & \mapsto & f(\mathbf{S}) \\ & \mathbb{C}_+^{n_T \times n_T} & \to & \mathbb{R} \end{array}$$

A utility function f

- *depends only on the unordered eigenvalues* $\mathbf{s} = \boldsymbol{\lambda}(\mathbf{S})$
- *is monotonic in the vector* $\mathbf{s}$, *that is,* $\mathbf{s} \leq \mathbf{s}' \Rightarrow f(\mathbf{s}) \leq f(\mathbf{s}')$

Utility optimization comprises a multiobjective (Pareto) optimization:



Utility functions II

Examples of utility functions:

- $f(\mathbf{S}) = \mathrm{E}\big[\log_2 \det\big(\mathbf{I} + \mathbf{WSW}^\dagger\big)\big]$ $\quad f(\mathbf{S}) = \Pr\big\{\det\big(\mathbf{I} + \mathbf{WSW}^\dagger\big) > \eta\big\}$
- $f(\mathbf{S}) = -\operatorname{tr}\mathrm{E}\big[(\mathbf{I} + \mathbf{WSW}^\dagger)^{-1}\big]$ $\quad f(\mathbf{S}) = \Pr\big\{-\operatorname{tr}\big[(\mathbf{I} + \mathbf{WSW}^\dagger)^{-1}\big] > \eta\big\}$
- $f(\mathbf{S}) = \lambda_{\mathrm{min}}(\mathbf{S})$ $\quad f(\mathbf{S}) = \lambda_{\mathrm{max}}(\mathbf{S})$
- $f(\mathbf{S}) = \det(\mathbf{S})$ $\quad f(\mathbf{S}) = \operatorname{tr}(\mathbf{S})$
- $f(\mathbf{S}) = \operatorname{tr}(\mathbf{S}^{-1})^{-1}$ $\quad f(\mathbf{S}) = \det(\mathbf{I} + \nu\mathbf{S})$

etc.

Joint pilot/precoder optimization

$$(\mathbf{P}^\star, \mathbf{Q}^\star) = \underset{\substack{(\mathbf{P},\mathbf{Q}) \\ \mathrm{rank}(\mathbf{P})=T_\tau}}{\mathrm{argmax}} \left\{ \frac{T - T_\tau}{T} f\big(\mathbf{S}(\mathbf{P},\mathbf{Q})\big) \right\}$$

subject to:

- Positive semidefiniteness: $\mathbf{P} \succeq \mathbf{0}$ and $\mathbf{Q} \succeq \mathbf{0}$
- Energy conservation: $\mathrm{tr}(\mathbf{P}) + (T - T_\tau)\,\mathrm{tr}(\mathbf{Q}) \leq T\mu$

Joint pilot/precoder optimization

$$(\mathbf{P}^\star, \mathbf{Q}^\star) = \underset{\substack{(\mathbf{P},\mathbf{Q}) \\ \text{rank}(\mathbf{P})=T_\tau}}{\text{argmax}} \left\{ \frac{T - T_\tau}{T} f\big(\mathbf{S}(\mathbf{P},\mathbf{Q})\big) \right\}$$

subject to:

- Positive semidefiniteness: $\mathbf{P} \succeq \mathbf{0}$ and $\mathbf{Q} \succeq \mathbf{0}$
- Energy conservation: $\text{tr}(\mathbf{P}) + (T - T_\tau)\,\text{tr}(\mathbf{Q}) \leq T\mu$

Reformulation:

$$(\mathbf{P}^\star, \mathbf{Q}^\star) = \underset{T_\tau=1,\ldots,n_\mathrm{T}}{\text{argmax}} \left\{ \frac{T - T_\tau}{T} \max_{\substack{(\mathbf{P},\mathbf{Q}) \\ \text{rank}(\mathbf{P})\leq T_\tau}} f(\mathbf{S}(\mathbf{P},\mathbf{Q})) \right\}$$

Partial optimization problems

- Prescribed precoder: $\mathbf{P}^\star(\mathbf{Q}) = \underset{\mathbf{P}}{\operatorname{argmax}}\ f(\mathbf{S}(\mathbf{P},\mathbf{Q})) = \underset{\mathbf{S}(\bullet,\mathbf{Q})}{\operatorname{argmax}}\ f(\mathbf{S})$
- Prescribed pilots: $\mathbf{Q}^\star(\mathbf{P}) = \underset{\mathbf{Q}}{\operatorname{argmax}}\ f(\mathbf{S}(\mathbf{P},\mathbf{Q})) = \underset{\mathbf{S}(\mathbf{P},\bullet)}{\operatorname{argmax}}\ f(\mathbf{S})$

Partial optimization problems

- Prescribed precoder: $\mathbf{P}^\star(\mathbf{Q}) = \underset{\mathbf{P}}{\text{argmax}}\ f(\mathbf{S}(\mathbf{P},\mathbf{Q})) = \underset{\mathbf{S}(\bullet,\mathbf{Q})}{\text{argmax}}\ f(\mathbf{S})$
- Prescribed pilots: $\mathbf{Q}^\star(\mathbf{P}) = \underset{\mathbf{Q}}{\text{argmax}}\ f(\mathbf{S}(\mathbf{P},\mathbf{Q})) = \underset{\mathbf{S}(\mathbf{P},\bullet)}{\text{argmax}}\ f(\mathbf{S})$

Theorem (rank inequalities)

$$\text{rank}(\mathbf{P}^\star(\mathbf{Q})) \leq \text{rank}(\mathbf{Q}) \qquad \text{rank}(\mathbf{Q}^\star(\mathbf{P})) \leq \text{rank}(\mathbf{P})$$

$$\#\textit{pilots} \leq \#\textit{streams} \qquad \#\textit{streams} \leq \#\textit{pilots}$$

Partial optimization problems

- Prescribed precoder: $\mathbf{P}^\star(\mathbf{Q}) = \underset{\mathbf{P}}{\operatorname{argmax}}\ f(\mathbf{S}(\mathbf{P},\mathbf{Q})) = \underset{\mathbf{S}(\bullet,\mathbf{Q})}{\operatorname{argmax}}\ f(\mathbf{S})$
- Prescribed pilots: $\mathbf{Q}^\star(\mathbf{P}) = \underset{\mathbf{Q}}{\operatorname{argmax}}\ f(\mathbf{S}(\mathbf{P},\mathbf{Q})) = \underset{\mathbf{S}(\mathbf{P},\bullet)}{\operatorname{argmax}}\ f(\mathbf{S})$

Theorem (rank inequalities)

$$\operatorname{rank}(\mathbf{P}^\star(\mathbf{Q})) \le \operatorname{rank}(\mathbf{Q}) \qquad \operatorname{rank}(\mathbf{Q}^\star(\mathbf{P})) \le \operatorname{rank}(\mathbf{P})$$

#pilots ≤ #streams *#streams ≤ #pilots*

Theorem (convex reformulations)

In the **S***-domain, for (quasi-)concave utilities f the partial problems are* **(quasi-)convex***!*

Joint training and transmit directions

Spectral decompositions:

$$\mathbf{R} = \mathbf{U}_{\mathbf{R}} \operatorname{diag}(\mathbf{r}) \mathbf{U}_{\mathbf{R}}^{\dagger} \qquad \mathbf{P} = \mathbf{U}_{\mathbf{P}} \operatorname{diag}(\mathbf{p}) \mathbf{U}_{\mathbf{P}}^{\dagger} \qquad \mathbf{Q} = \mathbf{U}_{\mathbf{Q}} \operatorname{diag}(\mathbf{q}) \mathbf{U}_{\mathbf{Q}}^{\dagger}$$

Joint training and transmit directions

Spectral decompositions:

$$\mathbf{R} = \mathbf{U}_{\mathbf{R}} \operatorname{diag}(\mathbf{r}) \mathbf{U}_{\mathbf{R}}^{\dagger} \qquad \mathbf{P} = \mathbf{U}_{\mathbf{P}} \operatorname{diag}(\mathbf{p}) \mathbf{U}_{\mathbf{P}}^{\dagger} \qquad \mathbf{Q} = \mathbf{U}_{\mathbf{Q}} \operatorname{diag}(\mathbf{q}) \mathbf{U}_{\mathbf{Q}}^{\dagger}$$

Theorem (eigenbasis alignment)

For any utility function, it is optimal to set

$$\mathbf{U}_{\mathbf{P}} = \mathbf{U}_{\mathbf{Q}} = \mathbf{U}_{\mathbf{R}}$$

Pareto optimization I

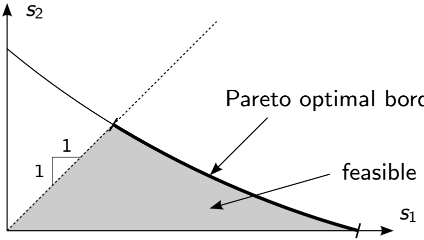
Iterating between partial optimizations does not guarantee optimality!

With alignment $\mathbf{U_P} = \mathbf{U_Q} = \mathbf{U_R}$, we get

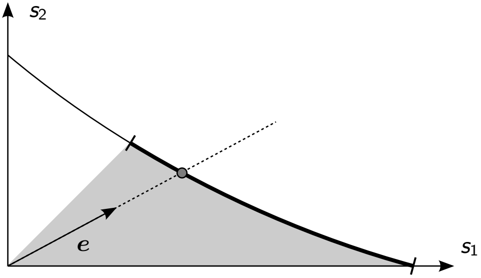
$$s_i = \frac{r_i^2 p_i q_i}{(1 + r_i p_i)\left(1 + \sum_{j=1}^{n_\mathrm{T}} \frac{r_j q_j}{1+r_j p_j}\right)}$$

Multiobjective (Pareto) optimization:

$$\max_{\substack{p_i, q_i \geq 0 \\ \sum_i p_i + (T - T_\tau) \sum_i q_i \leq T\mu}} (s_1, \ldots, s_{n_\mathrm{T}})$$



Pareto optimization II



Theorem (Pareto optimization)

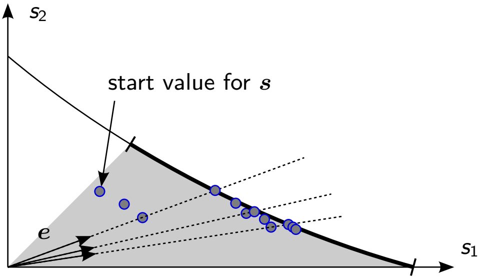
The optimization along a fixed direction $\mathbf{e}$

$$\max_{\mathbf{s}/\|\mathbf{s}\|=\mathbf{e}} \|\mathbf{s}\|$$

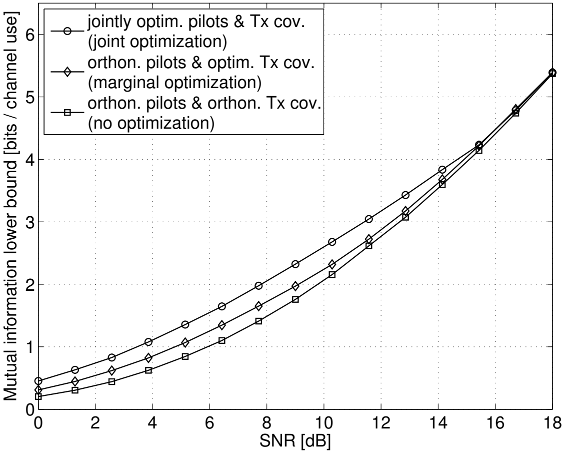
is a **quasi-convex** *problem*

Joint optimization algorithm

1. Pilot optimization: $\mathbf{s} \leftarrow \text{argmax}_{\mathbf{s}(\bullet,\mathbf{Q})} f(\mathbf{s})$
2. Precoder optimization: $\mathbf{s} \leftarrow \text{argmax}_{\mathbf{s}(\mathbf{P},\bullet)} f(\mathbf{s})$
3. Pareto optimization: $\mathbf{s} \leftarrow \text{argmax}_{\mathbf{s}/\|\mathbf{s}\|=\mathbf{e}} \|\mathbf{s}\|$

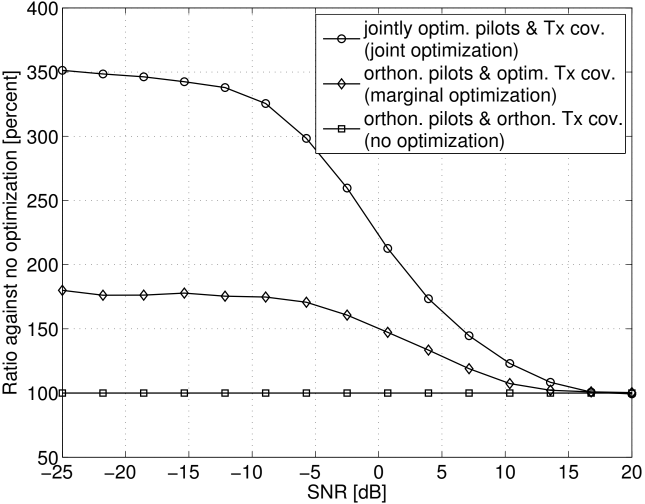


Simulations I



2x2 MIMO system with $r_1 = 0.7$, $r_2 = 0.3$

Simulations II



Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Joint Pilot-Precoder Design

- tackling a multidimensional joint optimization
- does the alignment property hold more generally?
- joint pilot-precoder optimization with rate-splitting bounds?

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Joint Pilot-Precoder Design

- tackling a multidimensional joint optimization
- does the alignment property hold more generally?
- joint pilot-precoder optimization with rate-splitting bounds?

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Joint Pilot-Precoder Design

- tackling a multidimensional joint optimization
- does the alignment property hold more generally?
- joint pilot-precoder optimization with rate-splitting bounds?

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Joint Pilot-Precoder Design

- tackling a multidimensional joint optimization
- does the alignment property hold more generally?
- joint pilot-precoder optimization with rate-splitting bounds?

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Joint Pilot-Precoder Design

- tackling a multidimensional joint optimization
- does the alignment property hold more generally?
- joint pilot-precoder optimization with rate-splitting bounds?

Related literature:

- A. Pastore, T. Koch, J.R. Fonollosa, *"A rate-splitting approach to fading channels with imperfect channel-state information,"* *IEEE Trans. Inf. Theory*, Vol. 60, Issue 7, p. 4266–4285
- A. Pastore, J. Hoydis, J.R. Fonollosa, *"Sharpened capacity lower bounds of fading MIMO channels with imperfect CSI,"* Proc. IEEE Int. Symp. on Inf. Theory 2013, Istanbul
- A. Pastore, M. Joham, J.R. Fonollosa, *"A Framework for Joint Design of Pilot Sequence and Linear Precoder,"* submitted to *IEEE Trans. Inf. Theory* Available: arXiv:1210.1470

Conclusions

Improved Capacity Bounds for Imperfect CSI:

- novel bound on MIMO capacity with imperfect CSIR
- sharper yet a bit more complex
- asymptotically tight
- dependency of layering functions not well understood

Joint Pilot-Precoder Design

- tackling a multidimensional joint optimization
- does the alignment property hold more generally?
- joint pilot-precoder optimization with rate-splitting bounds?

Related literature:

- A. Pastore, T. Koch, J.R. Fonollosa, *"A rate-splitting approach to fading channels with imperfect channel-state information,"* *IEEE Trans. Inf. Theory*, Vol. 60, Issue 7, p. 4266–4285
- A. Pastore, J. Hoydis, J.R. Fonollosa, *"Sharpened capacity lower bounds of fading MIMO channels with imperfect CSI,"* Proc. IEEE Int. Symp. on Inf. Theory 2013, Istanbul
- A. Pastore, M. Joham, J.R. Fonollosa, *"A Framework for Joint Design of Pilot Sequence and Linear Precoder,"* submitted to *IEEE Trans. Inf. Theory* Available: arXiv:1210.1470

Thank you!

Backup slides

Large random-matrix approximation

Large random-matrix approximation [Hachem, Khorunzhiy, Loubaton, Najim, Pastur '08]

As $n_\mathrm{R}, n_\mathrm{T} \to \infty$ with $0 < \liminf(n_\mathrm{R}/n_\mathrm{T}) \le \limsup(n_\mathrm{R}/n_\mathrm{T}) < \infty$,

$$\boxed{R^\star_{\mathrm{RS}}(\mathbf{K}) = \bar{R}^\star_{\mathrm{RS}}(\mathbf{K}) + \mathcal{O}(1/n_\mathrm{T})}$$

where

$$\boxed{\bar{R}^\star_{\mathrm{RS}}(\mathbf{K}) = \int_0^\infty \int_{\sigma^2(x)}^\infty \frac{n_\mathrm{R} - n_\mathrm{T} g(x,u) d(g(x,u))}{u} f(x)\,\mathrm{d}u\,\mathrm{d}x}$$

with

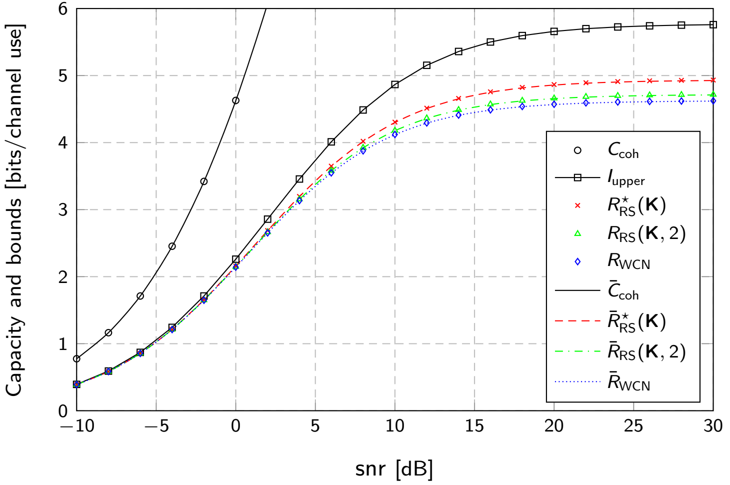
$$\sigma^2(x) = \frac{1}{\hat{V}}\left(\frac{\tilde{V}}{n_\mathrm{T}} x + \rho^{-1}\right)$$

$$f(x) = \frac{x^{n_\mathrm{T}-1} e^{-x}}{(n_\mathrm{T}-1)!}$$

$$g(x,u) = \frac{\tilde{V}}{\hat{V}}\left(1 - \frac{x}{n_\mathrm{T}}\right) + u$$

$$d(x) = \frac{\frac{n_\mathrm{R}}{n_\mathrm{T}} - 1}{2x} - \frac{1}{2} + \frac{\sqrt{\left(1 - \frac{n_\mathrm{R}}{n_\mathrm{T}} + x\right)^2 + 4\frac{n_\mathrm{R}}{n_\mathrm{T}} x}}{2x}$$

Large random-matrix approximation: numerical example



Capacity and MI bounds and their asymptotic approximations vs. SNR. Parameter values: $n_{\text{R}} = 6$, $n_{\text{T}} = 4$, $\hat{V} = \tilde{V} = 0.5$.

Asymptotically perfect CSI (SISO)

Assumptions:

- The joint law of $(\hat{H}, \tilde{H}) = (\hat{H}_{\mathsf{snr}}, \tilde{H}_{\mathsf{snr}})$ depends on the SNR
- $\limsup_{\mathsf{snr}\to\infty} \left\{ \frac{\tilde{V}_{\mathsf{snr}}(\hat{h}_{\mathsf{snr}})}{\frac{1}{\pi \mathrm{e}} \mathrm{e}^{h(\tilde{H}_{\mathsf{snr}} | \hat{H}_{\mathsf{snr}} = \hat{h}_{\mathsf{snr}})}} \right\} \leq M$ for (almost) every $\hat{H}_{\mathsf{snr}} = \hat{h}_{\mathsf{snr}}$

Asymptotically perfect CSI (SISO)

Assumptions:

- The joint law of $(\hat{H}, \tilde{H}) = (\hat{H}_{\mathsf{snr}}, \tilde{H}_{\mathsf{snr}})$ depends on the SNR
- $\limsup_{\mathsf{snr}\to\infty} \left\{ \frac{\tilde{V}_{\mathsf{snr}}(\hat{h}_{\mathsf{snr}})}{\frac{1}{\pi e} e^{h(\tilde{H}_{\mathsf{snr}} \mid \hat{H}_{\mathsf{snr}} = \hat{h}_{\mathsf{snr}})}} \right\} \leq M$ for (almost) every $\hat{H}_{\mathsf{snr}} = \hat{h}_{\mathsf{snr}}$

Asymptotically perfect CSI (SISO)

Assumptions:

- The joint law of $(\hat{H}, \tilde{H}) = (\hat{H}_{\mathsf{snr}}, \tilde{H}_{\mathsf{snr}})$ depends on the SNR
- $\limsup_{\mathsf{snr}\to\infty} \left\{ \frac{\tilde{V}_{\mathsf{snr}}(\hat{h}_{\mathsf{snr}})}{\frac{1}{\pi e} e^{h(\tilde{H}_{\mathsf{snr}} \mid \hat{H}_{\mathsf{snr}} = \hat{h}_{\mathsf{snr}})}} \right\} \le M$ for (almost) every $\hat{H}_{\mathsf{snr}} = \hat{h}_{\mathsf{snr}}$

Asymptotically perfect CSI (SISO)

Assumptions:

- The joint law of $(\hat{H}, \tilde{H}) = (\hat{H}_{\text{snr}}, \tilde{H}_{\text{snr}})$ depends on the SNR
- $\limsup_{\text{snr}\to\infty} \left\{ \dfrac{\tilde{V}_{\text{snr}}(\hat{h}_{\text{snr}})}{\frac{1}{\pi e} e^{h(\tilde{H}_{\text{snr}} | \hat{H}_{\text{snr}} = \hat{h}_{\text{snr}})}} \right\} \leq M$ for (almost) every $\hat{H}_{\text{snr}} = \hat{h}_{\text{snr}}$

Asymptotic offset

If $\lim_{\text{snr}\to\infty} \mathrm{E}\left[\tilde{V}_{\text{snr}}(\hat{H}_{\text{snr}})\right] = 0$, then $R^\star_{\text{SD}}$ is asymptotically tight to within $\log(M)$:

$$\lim_{\text{snr}\to\infty} \left\{ I(X_{\text{G}}; Y | \hat{H}) - R^\star_{\text{SD}} \right\} \leq \log(M)$$

Asymptotically perfect CSI (SISO)

Assumptions:

- The joint law of $(\hat{H}, \tilde{H}) = (\hat{H}_{\text{snr}}, \tilde{H}_{\text{snr}})$ depends on the SNR
- $\limsup_{\text{snr}\to\infty} \left\{ \frac{\tilde{V}_{\text{snr}}(\hat{h}_{\text{snr}})}{\frac{1}{\pi e} e^{h(\tilde{H}_{\text{snr}} | \hat{H}_{\text{snr}} = \hat{h}_{\text{snr}})}} \right\} \leq M$ for (almost) every $\hat{H}_{\text{snr}} = \hat{h}_{\text{snr}}$

Asymptotic offset

If $\lim_{\text{snr}\to\infty} \mathsf{E}\left[\tilde{V}_{\text{snr}}(\hat{H}_{\text{snr}})\right] = 0$, then $R^{\star}_{\text{SD}}$ is asymptotically tight to within $\log(M)$:

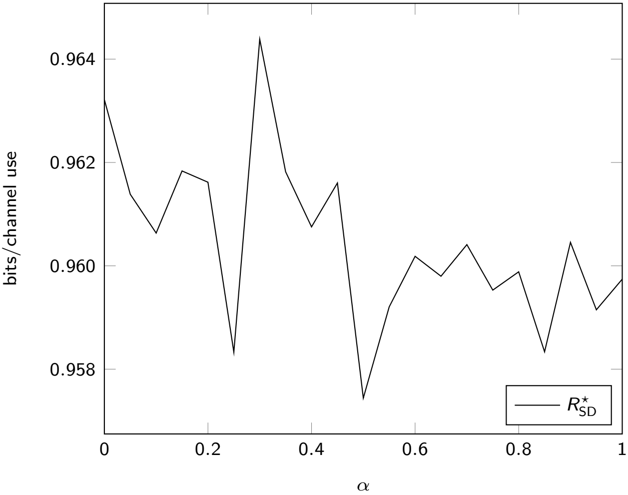
$$\lim_{\text{snr}\to\infty} \left\{ I(X_{\text{G}}; Y|\hat{H}) - R^{\star}_{\text{SD}} \right\} \leq \log(M)$$

Corollary (asymptotic tightness)

If $\tilde{H}_{\text{snr}}|\hat{H}_{\text{snr}}$ is Gaussian,

$$\lim_{\text{snr}\to\infty} \left\{ I(X_{\text{G}}; Y|\hat{H}) - R^{\star}_{\text{SD}} \right\} = 0$$

Dependency on the layering



Infinite-layer bound $R^\star_{\mathsf{SD}}(\alpha \mathbf{K}_{\mathsf{lev}} + (1-\alpha)\mathbf{K}_{\mathsf{stag}})$ as a function of α for a $(r, t) = (1, 2)$ MISO channel, $\mathsf{snr} = 13\,\mathrm{dB}$