Capacity Limits of MIMO Channels

Markku Juntti

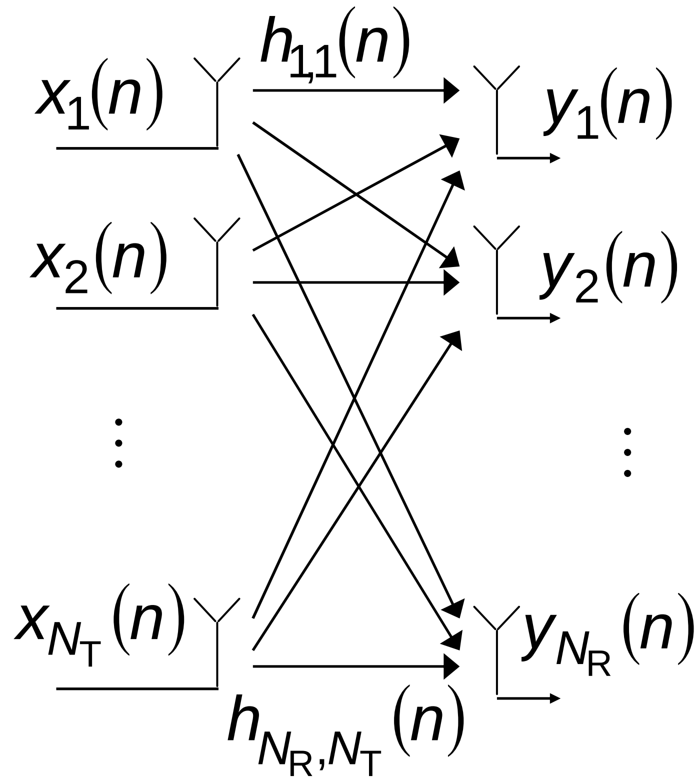
Contents

1. Introduction

- The use of multiple antennas can provide gain due to
  - antenna gain
    - more receive antennas ⇨ more power is collected
  - interference gain
    - interference nulling by beamforming (array gain)
    - interference averaging (to zero) due to independent observations
  - diversity gain against fading
    - receive diversity
    - transmit diversity.
- Information theoretic model of multi-input–multi-output (MIMO) channel is considered.

MIMO Channel Model

- Assume N_T transmit and N_R receive antennae
  - called $N_T \times N_R$ MIMO system.
- Fading radio channels modeled as frequency-flat:
  - fixed
  - time-varying
  - known both/either in the transmitter and/or receiver
    - perfect *channel state information* (CSI)
  - *a priori* unknown.



MIMO channel model.

2. Review of Information Theory

- Information theory (IT) has its origins in analyzing the limits communications.
- Information theory answers two fundamental questions in communication theory:
  - What is the ultimate data compression rate?
    - Answer: entropy.
  - What is the ultimate data transmission rate?
    - Answer: channel capacity.

Basic Concepts

- Assume a discrete valued *random variable* (RV) X with probability mass function $p(x)$.
- The *average information* or *entropy* of RV X:

$$H(X) = -\sum_{x} p(x)\log[p(x)] = -\mathrm{E}[\log p(X)] = \mathrm{E}\left[\log\frac{1}{p(X)}\right].$$

- *Joint entropy* of RV's X and Y:

$$H(X,Y) = -\sum_{x}\sum_{y} p(x,y)\log[p(x,y)] = -\mathrm{E}\{[\log[p(X,Y)]]\}.$$

- *Conditional entropy* of RV Y given $X = x$:

$$H(Y|X) = \sum_{x} p(x)H(Y|X=x) = -\sum_{x}\sum_{y} p(x,y)\log[p(y|x)] = -\mathrm{E}\{\log[p(X,Y)]\}.$$

⇨ Chain rule:

$$H(X,Y) = H(X) + H(Y|X).$$

Mutual Information

- *Mutual information* is the *relative entropy* between the joint distribution and product distribution:

$$I(X;Y) = \sum_x \sum_y p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right] = \mathrm{E}\left\{\log\left[\frac{p(X,Y)}{p(X)p(Y)}\right]\right\}.$$

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$= H(X) + H(Y) - H(X,Y) = I(Y;X).$$

- Measure of the information one random variable (say, X) contains on the other (Y):
  - If X and Y are independent: $I(X;Y) = 0$ (also “only if”).
  - If $Y = X$: $I(X;X) = H(X)$.
- Differential entropy for continuous RV’s.

Gaussian RV's

- For multivariate, real-valued Gaussian RV's X_1, X_2,..., X_n with mean vector μ and covariance matrix K, the differential entropy is

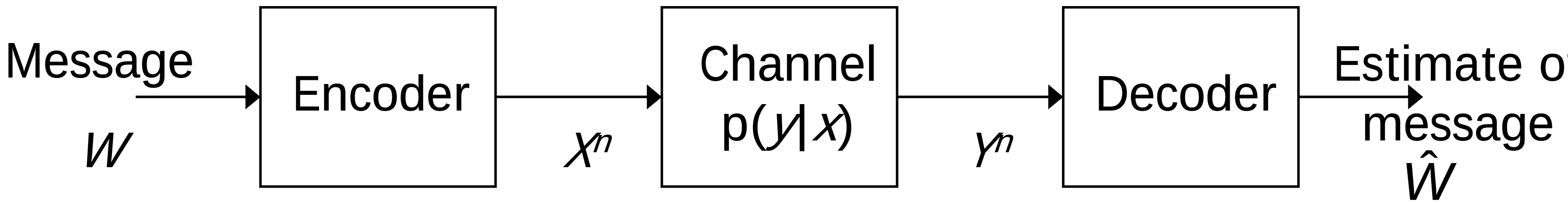
$$h(X_1, X_2, \ldots, X_n) = \frac{1}{2}\log\left[(2\pi e)^n\right]\det(\mathbf{K}).$$

- Gaussian distribution maximizes the entropy over all distributions with the same covariance:

$$h(X_1, X_2, \ldots, X_n) \leq \frac{1}{2}\log\left[(2\pi e)^n\right]\det(\mathbf{K})$$

for any RV's X_1, X_2,..., X_n with equality if and only if they are Gaussian.

Channel Capacity



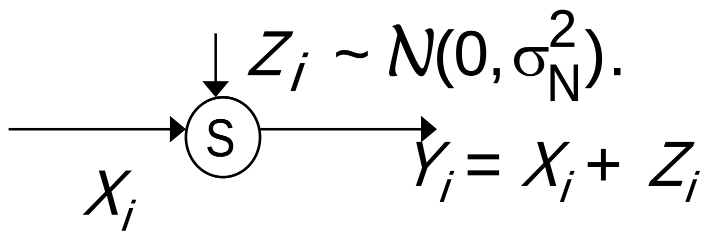
Information theoretic model of a communication system.

- *Channel capacity*: $C = \max_{p(x)} I(X;Y).$

- Code rate R is *achievable*, if there exists a sequence of $(2^{nR}, n)$ codes so that

$$P_{e,\max} \to 0, \text{ as } n \to \infty.$$

Gaussian Channel



The Gaussian channel.

- Channel capacity:

$$C = \max_{\substack{p(x) \\ E(X^2) \le \sigma_S^2}} I(X;Y) = \boxed{\frac{1}{2}\log(1+\gamma)}, \quad \gamma = \frac{\sigma_S^2}{\sigma_N^2}.$$

- Capacity per time unit ((2 W) samples per second):

$$C = \boxed{W \log\left(1 + \frac{P}{N_0 W}\right)}.$$

Parallel Gaussian Channels

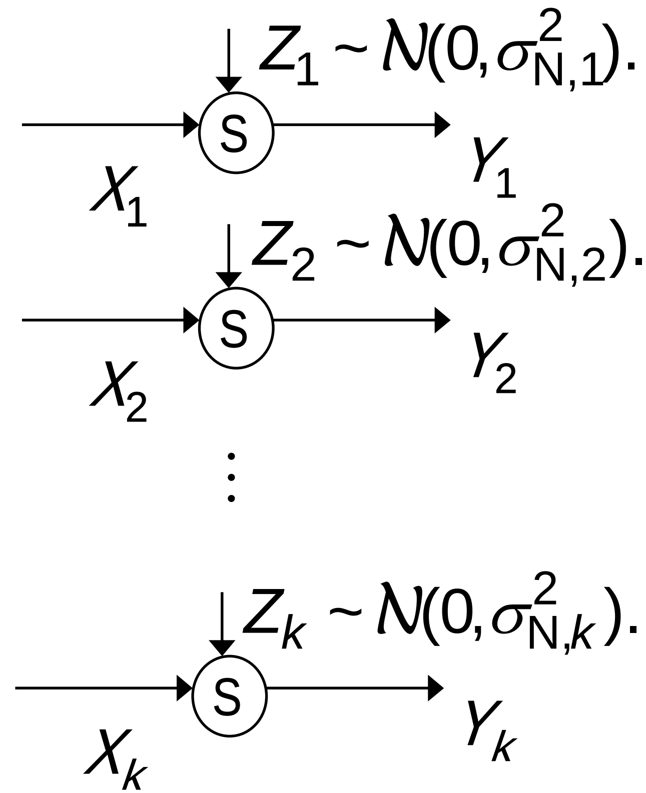
- Capacity:

$$C=\sum_{i=1}^{k}\frac{1}{2}\log\left(1+\frac{\sigma_{S,i}^{2}}{\sigma_{N,i}^{2}}\right)=\sum_{i=1}^{k}\frac{1}{2}\log(1+\gamma_i).$$

- Optimal transmission:

$$\mathbf{X}\sim\mathcal{N}\left(\mathbf{0},\mathrm{diag}\left[\sigma_{S,1}^{2},\sigma_{S,2}^{2},\ldots,\sigma_{S,k}^{2}\right]\right)$$

⇨ *water-filling*.



Parallel Gaussian channels.

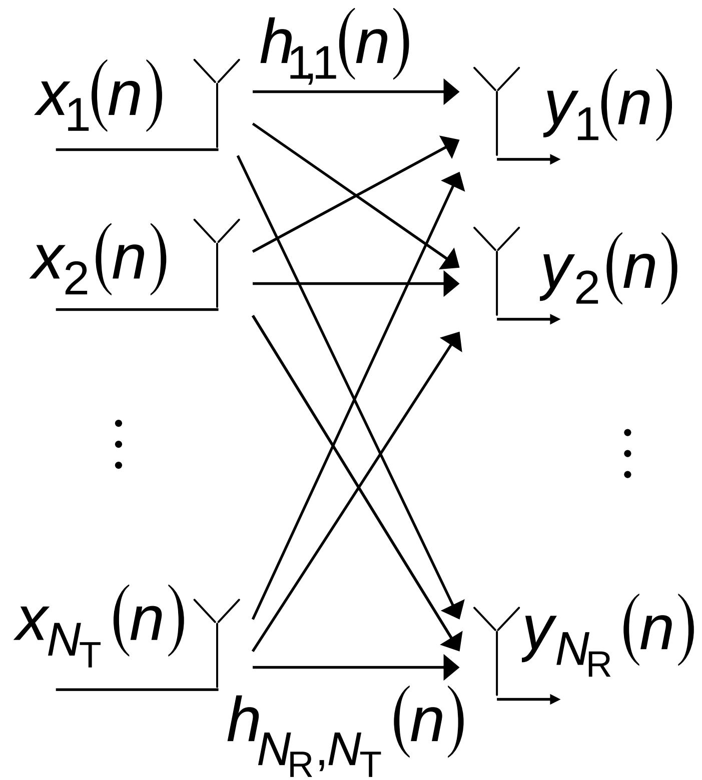
3. Fixed MIMO Channels

- Signal $x_i(n)$ is transmitted at time interval n from antenna i (i=1,2,...,N_T).
- Signal $y_j(n)$ is received at time interval n at antenna j (j=1,2,...,N_R):

$$y_j(n) = \sum_{i=1}^{N_T} h_{ij}(n) x_i(n) + \eta_j(n),$$

where $h_{ij}(n)$ is the complex channel gain with

$$\mathrm{E}\left(\left|h_{ij}(n)\right|^2\right) = 1$$



MIMO channel model.

Matrix Formulation of MIMO Channel Model

- The signal received at all antennas:

$$\mathbf{y}(n) = \mathbf{H}(n)\mathbf{x}(n) + \boldsymbol{\eta}(n),$$

where $\mathbf{x}(n) = \begin{bmatrix} x_1(n) & x_2(n) & \cdots & x_{N_T}(n) \end{bmatrix}^T \in C^{N_T}$,

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) & y_2(n) & \cdots & y_{N_R}(n) \end{bmatrix}^T \in C^{N_R},$$

$$\mathbf{H}(n) = \begin{bmatrix} h_{1,1}(n) & h_{1,2}(n) & \cdots & h_{1,N_T}(n) \\ h_{2,1}(n) & h_{2,2}(n) & \cdots & h_{2,N_T}(n) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1}(n) & h_{N_R,2}(n) & \cdots & h_{N_R,N_T}(n) \end{bmatrix} \in C^{N_R \times N_T}.$$

UNIVERSITY of OULU

Noise Model and Power Constraint

- The noise vector

$$\boldsymbol{\eta}(n) = \left[\eta_1(n) \quad \eta_2(n) \quad \cdots \quad \eta_{N_\mathrm{R}}(n)\right]^\mathrm{T} \in C^{N_\mathrm{R}},$$

  satisfies

$$\boldsymbol{\eta}(n) \sim C\mathcal{N}\left(\mathbf{0}, \sigma_\mathrm{N}^2 \mathbf{I}\right).$$

- The transmitted signal satisfies the average power constraint:

$$\mathrm{E}\left(\mathbf{x}^\mathrm{H}(n)\mathbf{x}(n)\right) = \sum_{i=1}^{N_\mathrm{T}} \mathrm{E}\left(\left|x_i(n)\right|^2\right) = \sum_{i=1}^{N_\mathrm{T}} \sigma_{\mathrm{S},i}^2 \le \sigma_\mathrm{S}^2.$$

Singular Value Decomposition

- The MIMO model is a special case of parallel Gaussian channels.
- The channel transfer matrix has *singular value decomposition* (SVD):

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^{\mathrm{H}},$$

where

$$\mathbf{U} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}, \mathbf{V} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$$

are unitary matrices, and

$$\mathbf{\Lambda}^{1/2} \in \mathbb{R}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$$

is a "diagonal" matrix of the singular values of $\boldsymbol{H}$.

Equivalent Channel Model

- Let $\tilde{\mathbf{x}}(n) = \mathbf{V}^{H}\mathbf{x}(n), \quad \tilde{\mathbf{y}}(n) = \mathbf{U}^{H}\mathbf{y}(n), \quad \tilde{\boldsymbol{\eta}}(n) = \mathbf{U}^{H}\boldsymbol{\eta}(n).$
- Since ***U*** and ***V*** are unitary:

$$\mathrm{E}\left(\tilde{\mathbf{x}}^{H}(n)\tilde{\mathbf{x}}(n)\right) \leq \sigma_{S}^{2},$$

$$\tilde{\boldsymbol{\eta}}(n) \sim CN\left(\mathbf{0}, \sigma_{N}^{2}\mathbf{I}\right).$$

⇨ Equivalent channel model

$$\tilde{\mathbf{y}}(n) = \mathbf{\Lambda}^{1/2}\tilde{\mathbf{x}}(n) + \tilde{\boldsymbol{\eta}}(n).$$

"diagonal" matrix of sixe $N_R \times N_T$

⇨ Independent parallel Gaussian channels.

⇨ Capacity achieved with Gaussian input and by water-filling.

Derivation of Channel Capacity

- The rank of matrix $\mathbf{H}$ is $\text{rank}(\mathbf{H}) \le \min(N_R, N_T)$.

⇨ The number of positive singular values is $\text{rank}(\mathbf{H})$.

⇨ The capacity of MIMO AWGN channel:

$$C = \sum_{i=1}^{\text{rank}(\mathbf{H})} \log\left(1 + \frac{\lambda_i \sigma_{S,i}^2}{\sigma_N^2}\right) = \sum_{i=1}^{\text{rank}(\mathbf{H})} \log(1 + \lambda_i \gamma_i), \quad \gamma_i = \frac{\sigma_{S,i}^2}{\sigma_N^2},$$

where the signal powers are solved via water-filling

$$\sigma_{S,i}^2 = \max\left\{0, \left(\frac{\mu - \sigma_N^2}{\lambda_i}\right)\right\}, \; i = 1,2,\ldots,\text{rank}(\mathbf{H}),$$

and μ is chosen so that the power constraint is satisfied or

$$\sum_{i=1}^{\text{rank}(\mathbf{H})} \sigma_{S,i}^2 \le \sigma_S^2.$$

MIMO Channel Capacity for Full–Rank Channel Matrix

- No CSI at the transmitter (and full–rank **H**):

$$C = \log\left[\det\left(\boldsymbol{I}_{N_\mathrm{R}} + \frac{\gamma}{N_\mathrm{T}}\boldsymbol{H}\boldsymbol{H}^\mathrm{H}\right)\right].$$

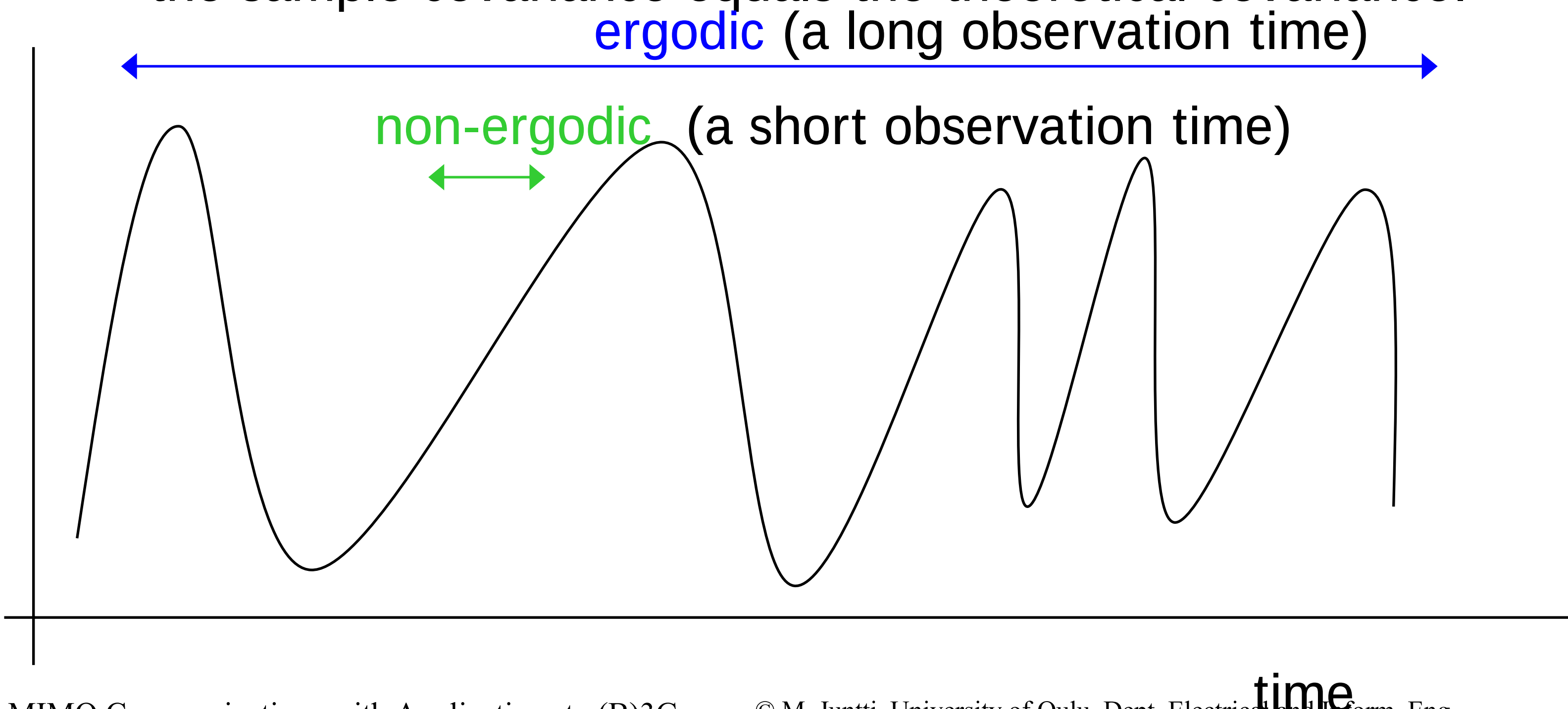
- CSI at the transmitter (and full–rank **H**):

$$C = \max_{\mathbf{Q}} \log\left[\det\left(\mathbf{I}_{N_\mathrm{R}} + \frac{\gamma}{N_\mathrm{T}}\mathbf{H}\mathbf{Q}\mathbf{H}^\mathrm{H}\right)\right],$$

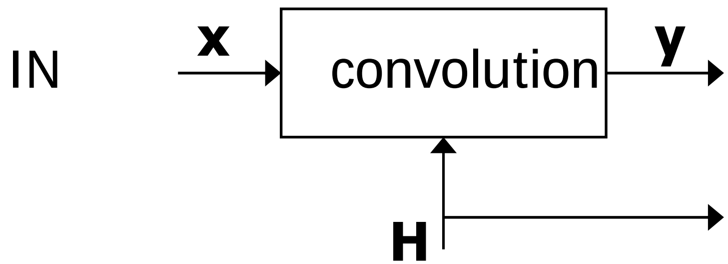
  where **Q** is the covariance matrix of the input vector **x** satisfying the power constraint $\mathrm{tr}(\mathbf{Q}) \leq \sigma_S^2$.
  - No CSI at the transmitter ⇨ **Q** = **I**.

4. Fading MIMO Channels

- The channels are usually assumed to be ergodic: fading is fast enough and gets all realizations so many times that
  - the sample average equals the theoretical mean
  - the sample covariance equals the theoretical covariance.



Fading Channel Model with Perfect Receiver CSI



$$I(\mathbf{x};\mathbf{y},\mathbf{H}) = I(\mathbf{x};\mathbf{H}) + I(\mathbf{x};\mathbf{y}|\mathbf{H}) = I(\mathbf{x};\mathbf{y}|\mathbf{H}).$$

OUT

= 0

RV conditioned on channel realization

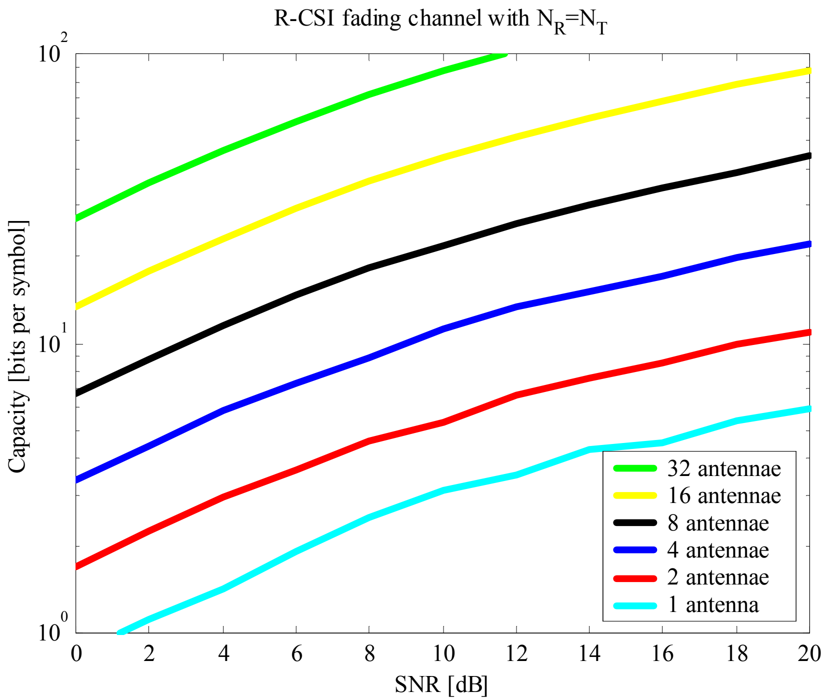
- The effective channel output: the actual channel output **y** and the channel realization **H**.
- Assuming that the channel is memoryless (independent channel state for each transmission), the capacity equals the mean of the mutual information:

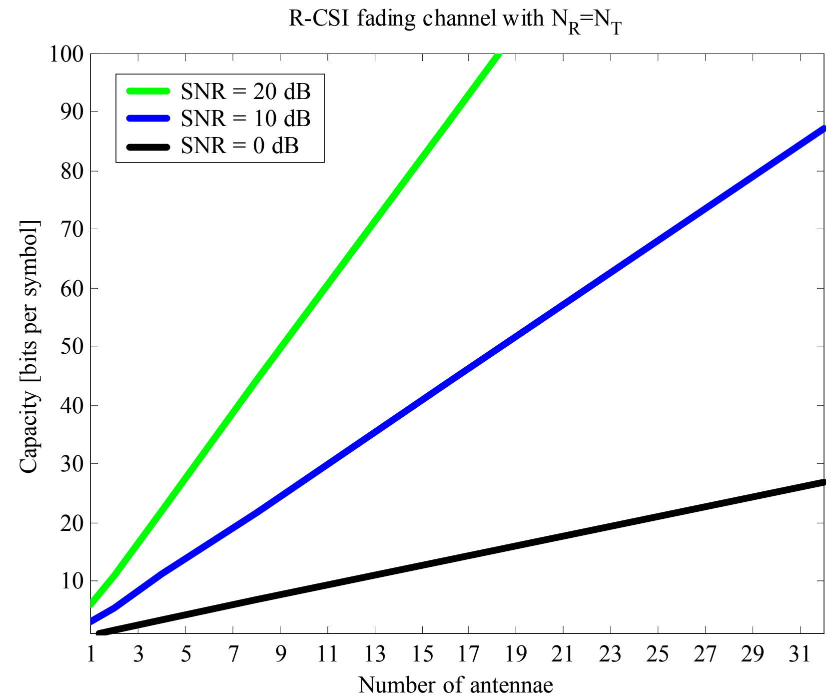
$$C = \mathrm{E}_{\boldsymbol{H}}\left\{\log\left[\det\left(\boldsymbol{I}_{N_\mathrm{R}} + \frac{\gamma}{N_\mathrm{T}}\boldsymbol{H}\boldsymbol{H}^\mathrm{H}\right)\right]\right\}.$$

Capacity Evaluation

- The evaluation of the fading MIMO channel capacity is complicated:
  - Wishart distribution ⇨ Laguerre polynomials [Telatar 1999]
  - bounds [Foschini & Gans 1998]
  - **Monte Carlo computer simulations**
  - random matrix theory ⇨ mutual information tends to Gaussian
    - under development.

Example: $N \times N$ MIMO System



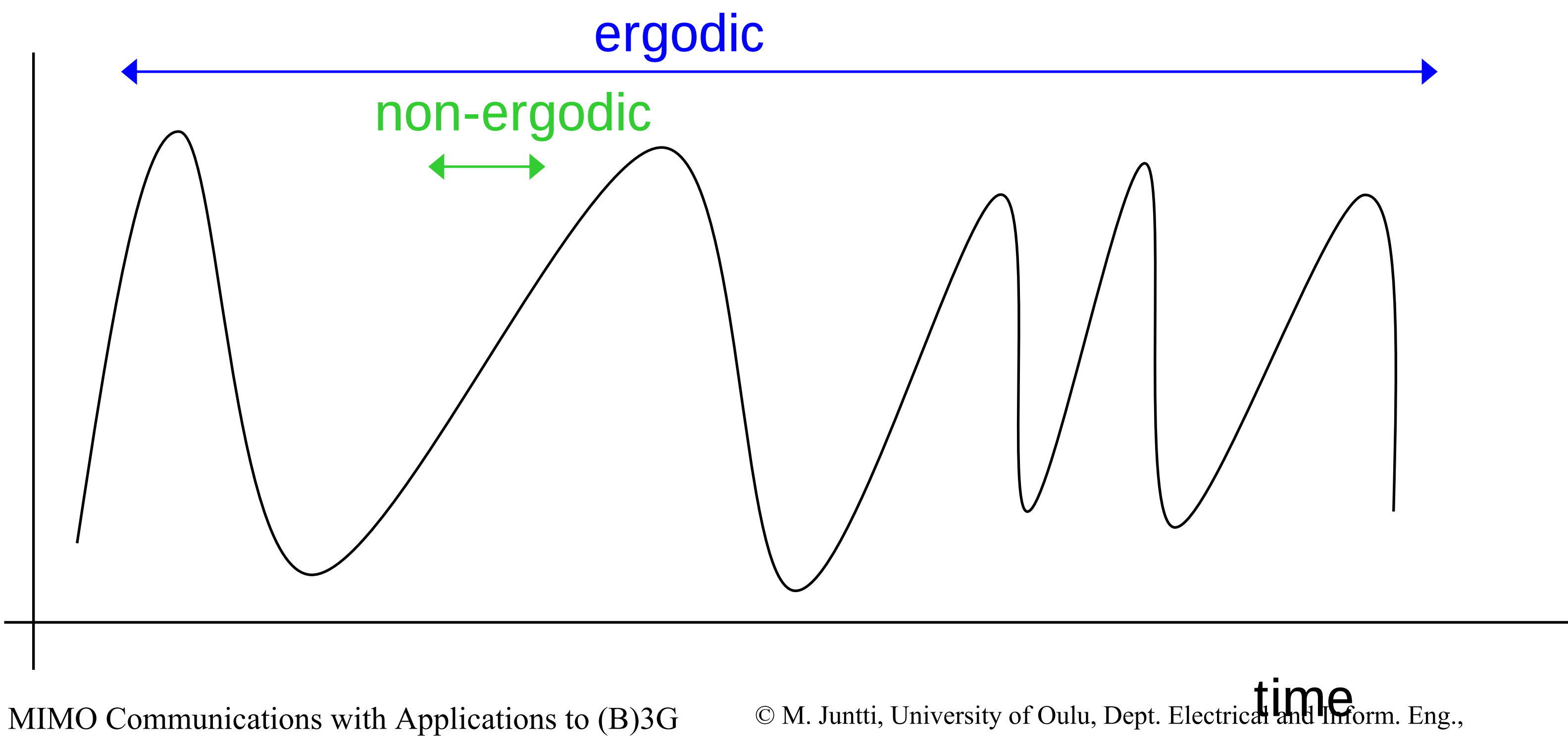


- ⇨ The capacity curves are sifted upwards by introducing more antennae.
- ⇨ The capacity increases linearly vs. the number of antennae.

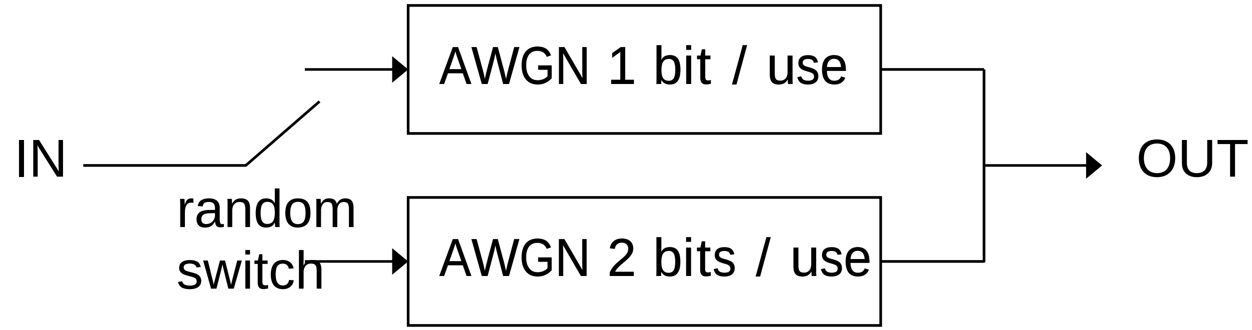
Non-Ergodic Channels

- The channels are not always ergodic: fading can be so slow that it undergoes only some realizations.

⇨ The random process becomes non-ergodic.



Example



- Select one of the channels with equal probability, and keep then fixed.

⇨ Average mutual information is 1.5 bits / channel use.

- However, with probability 0.5 it is not supported.

⇨ The achievable rate $\leq$ 1 bits / channel use.

⇨ **Channel capacity $\neq$ the average maximum mutual information.**

Example: Random and Fixed Channel

- A simple example: generate a channel realization, and keep it fixed during the whole transmission.

⇨ There is a positive probability of an arbitrarily bad channel realization.

⇨ However small a rate, the channel realization may not be able to support it regardless the length of the code word.

⇨ The Shannon capacity of this non-ergodic channel is zero.

⇨ The Shannon capacity is again not equal to the average mutual information.

Outage Probability

- In non-ergodic channels, the capacity is measured by the *probability of outage* for a given rate *R*:

$$P_{\text{out}}(R) = \inf_{\boldsymbol{Q}:\boldsymbol{Q}\geq 0, \text{tr}(\boldsymbol{Q})\leq \sigma_S^2} \Pr\left[I(\boldsymbol{x};\boldsymbol{y}) < R \right]$$

$$= \inf_{\boldsymbol{Q}:\boldsymbol{Q}\geq 0, \text{tr}(\boldsymbol{Q})\leq \sigma_S^2} \Pr\left\{ \log\left[\det\left(\boldsymbol{I}_{N_R} + \frac{\gamma}{N_T} \boldsymbol{HQH}^H \right) \right] < R \right\}.$$

  - Often called capacity versus outage.
- The set-up is encountered in real time applications with transmission delay constraints.
- Similar approach is applicable also for delay constrained communications in ergodic channels.

5. Summary and Conclusions

- AWGN MIMO channels are an extension of parallel Gaussian channels.
  - Another example of parallel channels: channels on different frequencies.
- Introducing both multiple transmit and receive antennae is equivalent to increase in bandwidth.
- The linear capacity increase becomes natural.

$$C = \log\left[\det\left(\boldsymbol{I}_{N_R} + \frac{\gamma}{N_T}\boldsymbol{HQH}^H\right)\right].$$

Fading AWGN MIMO Channel

- Ergodic channels:
  - Channel experiences all its states several times.
  - No delay constraints and/or fast fading.
  - Capacity equals the average mutual information:

$$C = \mathrm{E}_{\boldsymbol{H}}\left\{\log\left[\det\left(\boldsymbol{I}_{N_\mathrm{R}} + \frac{\gamma}{N_\mathrm{T}}\boldsymbol{H}\boldsymbol{H}^\mathrm{H}\right)\right]\right\}.$$

  - Capacity increases linearly with $N_\mathrm{R} = N_\mathrm{T}$.
- Non-ergodic channels:
  - Capacity does not equal the average mutual information.
  - Capacity versus outage probability.

Research Challenges

- Capacity of selective channels
  - time-selective
  - frequency-selective

  with no or imperfect channel state information in the transmitter and the receiver.

⇨ Optimal signal structures (coding and modulation) for real use with issues like
  - amount of training vs. non-coherent detection
  - transceiver complexity constraints
  - limited bandwidth of a non-ideal feedback channel.

References

1. T. M. Cover & J. A. Thomas, Elements of Information Theory. John Wiley & Sons, 1991. ISBN: 0-471-06259-6
2. E. Telatar, Capacity of multi-antenna Gaussian channels. European Transactions on Telecommunications, vol. 10, no. 6, pp. 585-595, Nov.-Dec. 1999.
3. G. J. Foschini & M. J. Gans, On limits of wireless communications in a fading environment when using multiple antennas. Wireless Personal Communications , vol. 6, pp. 311-335, Nov.-Dec. 1999
4. T. L. Marzetta & B. M. Hochwald, Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. IEEE Transactions on Information Theory, vol. 45, no. 1, pp. 139-157, Jan. 1999
5. I. E. Telatar & D. N C. Tse, Capacity and mutual information of wideband multipath fading channels. IEEE Transactions on Information Theory, vol. 46, no. 4, pp. 1384-1400, July 2000.
6. M. Medard, The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. IEEE Transactions on Information Theory, vol. 46, no. 3, pp. 933-945, May 2000.
7. M. Medard & R. G. Gallager, Bandwidth scaling for fading multipath channels. IEEE Transactions on Information Theory, vol. 48, no. 4, pp. 840-852, April 2002.
8. V. G. Subramanian & B. Hajek, Broad-band fading channels: signal burstiness and capacity. IEEE Transactions on Information Theory, vol. 48, no. 4, pp. 809-827, April 2002.