An Introduction to MIMO communications

Pascal Vallet

Bordeaux INP/Laboratoire IMS

Evolution of cellular communications

- **1980-2000.** 1G cellular networks (AMPS, NMT, Radiocom2000)
- **1990-?** 2G cellular networks (GSM, DCS, GPRS, EDGE)
- **2000-?** 3G cellular networks (CDMA2000, UMTS, HSDPA, HSPA+)
- **2010-?** 4G celullar networks (LTE, LTE Advanced)
- **2020-?** 5G cellular networks
- $\sim$**2030** 6G cellular networks

Review of "classical" channel access techniques

- **TDMA.** Time Division Multiple Access (2G, 3G)
  - $\oplus$ flexible data rates
  - $\ominus$ synchronization/equalization, roaming, multipath interference
- **FMDA.** Frequency Division Multiple Access (1G, 2G)
  - $\oplus$ simplicity
  - $\ominus$ fading, inefficient use of bandwidth, static data rates
- **CMDA.** Code Division Multiple Access (3G)
  - $\oplus$ random access
  - $\ominus$ receiver complexity, complex power allocation
- **OFDMA.** Orthogonal Frequency Division Multiple Access (4G)
  - $\oplus$ simplicity, reduced multi-user interference and fading
  - $\ominus$ non-robust to frequency offsets

SDMA

- **SDMA.** Space Division Multiple Access
- The use of multiple antennas at the BS allows to form *virtual beams* focused on a specific user.
- Mutiple users may share the same time slots and frequency bands.
- Increased capacity if the UTs have multiple antennas (MIMO communications).
- Signal processing techniques for SDMA rely on array processing methods.

SDMA and mobile communications (1)

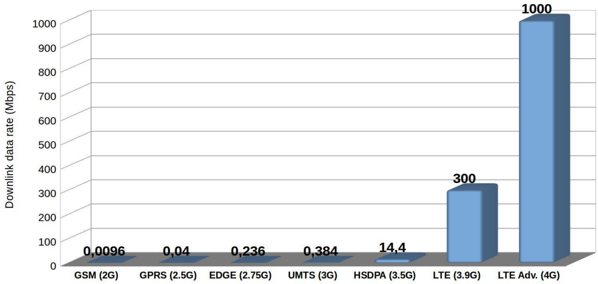


Figure: Evolution of downlink data rates (Mbps), from 2G to 4G

- TDMA, FDMA, CDMA, OFDMA.

- SDMA: No exploitation until LTE (MIMO 4x4) and LTE Adv. (MIMO 8x8).
- SDMA will be one of the main features in 5G standards.

SDMA and mobile communications (2)

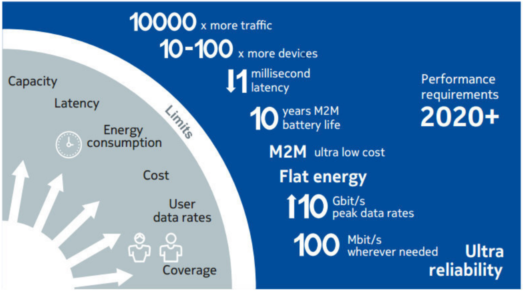
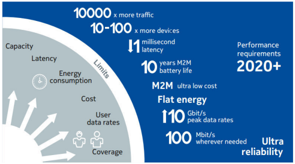


Figure: Requirements for future 2020 mobile standards (*source: Nokia*)

SDMA and mobile communications (3)



Key features

- Extreme densification of cells
- mmWave (30 GHz to 300 GHz)
- Massive MIMO (up to 120 antennas at base stations)

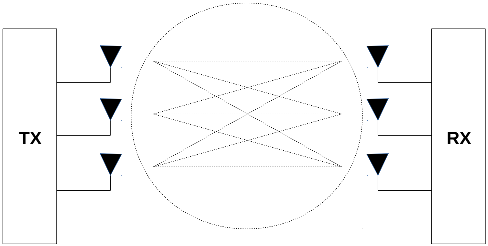
Challenges

- Green communications
- Co-user and co-channel interference
- Propagation of mmWaves

SDMA and mobile communications (4)

Figure: Exemple of a massive MIMO phased array for 5G communications (Aurora CMM.100.A, 5-6GHz, 64 antennas)

MIMO techniques



MIMO channel

- **Diversity.** MIMO allows to exploit spatial diversity, as a complement of the standard time and frequency diversities, to combat channel fading.
- **Capacity.** Under certain conditions, the use of M antennas at TX and N antennas at RX increases the capacity by a factor $\min(M, N)$, compared to the classical case where both TX and RX have a single antenna.
- **Space-Time Codes.** MIMO offers a new layer of coding, by carefully choosing how to organize symbols across each TX antenna and time.

Outline

Outline

SISO model (1)

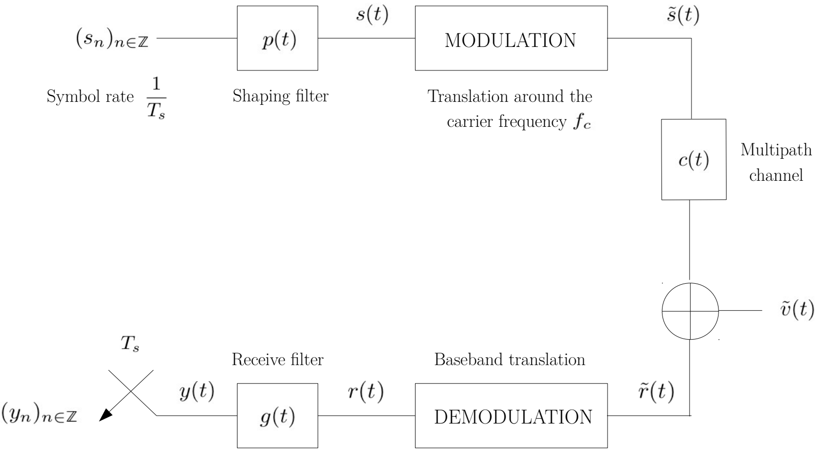


Figure: SISO transmit/receive chain

SISO model (2)

- **Transmit signal.** $\tilde{s}(t) = \mathrm{Re}\left(s(t)e^{i2\pi f_c t}\right)$ with

$$s(t) = \sum_{k\in\mathbb{Z}} s_k p(t - kT_s).$$

- **Receive signal.** $\tilde{r}(t) = \mathrm{Re}\left(r(t)e^{i2\pi f_c t}\right)$ with

$$r(t) = \sum_{p=1}^{P} \gamma_p s(t - T_p) + v(t)$$

where

- P is the multipath number
- T_p is the p-th path delay
- γ_p is the p-th path fading coefficient
- $v(t)$ is the baseband additive noise

SISO model (3)

- **Filtered signal.**

$$
\begin{aligned}
y(t) &= (r \star g)(t) \\
&= \sum_{k \in \mathbb{Z}} s_k h(t - kT_s) + w(t)
\end{aligned}
$$

where

- $h(t) = \sum_{p=1}^{P} \gamma_p (p \star g)(t - T_p)$ is the "equivalent" channel
- $w(t) = (g \star v)(t)$.

MISO model (1)

- N colocated transmit antennas radiating isotropically
- Narrowband signals (cf. array processing chapter)
- P far-field scatterers

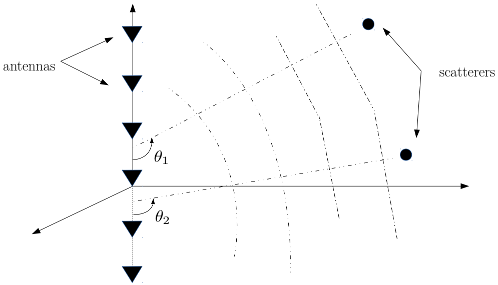


Figure: Example with a ULA at TX

MISO model (2)

- The signal received at the p-th scatterer is modelled as

$$\tilde{x}_p(t) = \alpha_p \sum_{n=1}^{N} \tilde{s}_n \left(t - T_{n,p}\right)$$

where

  - α_p is a fading coefficient between the antenna array and the p-th scatterer
  - $T_{n,p}$ is the propagation delay between the n-th antenna and the p-th scatterer
  - $\tilde{s}_n(t) = \mathrm{Re}\left(s_n(t)\mathrm{e}^{\mathrm{i}2\pi f_c t}\right)$ is the signal transmitted by the n-th antenna.

MISO model (3)

- The baseband model of $\tilde{x}_p(t)$ can be approximated as

$$x_p(t) = \alpha_p \mathrm{e}^{-\mathrm{i}2\pi f_c T_{1,p}} \mathbf{a}(\theta_p, \phi_p)^T \mathbf{s}(t - T_{1,p})$$

  with

  - $\mathbf{a}(\theta, \phi)$ is the steering vector associated with the transmit N-antenna array
  - θ_p, ϕ_p are the elevation and azimuth of the p-th scatterer
  - $\mathbf{s}(t) = (s_1(t), \ldots, s_N(t))^T$.

MISO model (4)

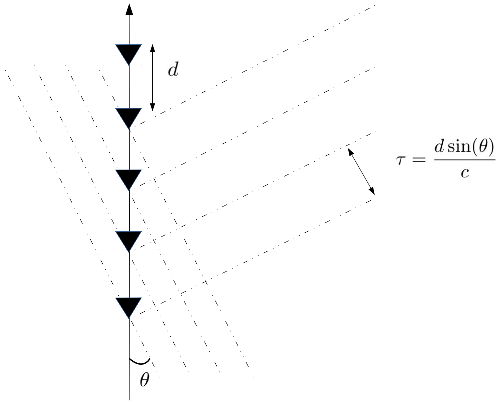
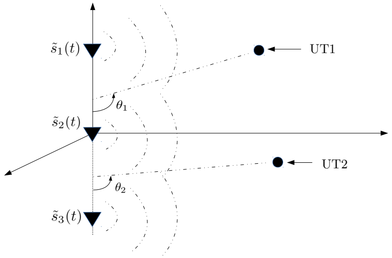


Figure: Example with ULA at TX, $\mathbf{a}(\theta) = \left(1, \mathrm{e}^{-\mathrm{i}2\pi\frac{d}{\lambda_c}\sin(\theta)}, \ldots, \mathrm{e}^{-\mathrm{i}2\pi(N-1)\frac{d}{\lambda_c}\sin(\theta)}\right)^T$

Digression on a 1st type of beamforming (1)

- **Scenario.** Instead of scatterers, consider two single-antenna UTs with directions (θ_1, ϕ_1) and (θ_2, ϕ_2), in LOS with the array.
- **Transmit signal.** On the n-th antenna, we transmit (in baseband)

$$s_n(t) = \overline{[\mathbf{a}(\theta_1, \phi_1)]_n} \quad \underbrace{s^{(1)}(t)}_{\text{signal for UT1}} \quad + \overline{[\mathbf{a}(\theta_2, \phi_2)]_n} \quad \underbrace{s^{(2)}(t)}_{\text{signal for UT2}}$$



Digression on a 1st type of beamforming (2)

- **Signal received at UT1.**

$$x_1(t) = \alpha_1\ \mathrm{e}^{-\mathrm{i}2\pi f_c T_1} \underbrace{\|\mathbf{a}(\theta_1, \phi_1)\|_2^2\ \ s^{(1)}(t - T_1)}_{\text{useful signal}}$$

$$+ \alpha_1\ \mathrm{e}^{-\mathrm{i}2\pi f_c T_1}\ \underbrace{\mathbf{a}(\theta_1, \phi_1)^T \overline{\mathbf{a}(\theta_2, \phi_2)} s^{(2)}(t - T_1)}_{\text{interference}}$$

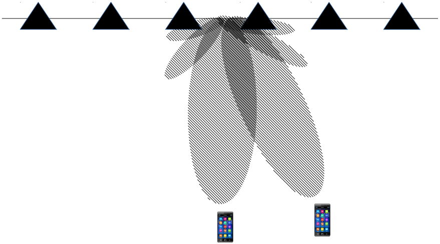
$+$ noise.

Digression on a 1st type of beamforming (3)

- **Example for ULA.** Denoting $\nu(\theta) = \frac{d}{\lambda_c}\sin(\theta)$, we have

$$\|\mathbf{a}(\theta_1)\|_2^2 = N$$

$$\left|\mathbf{a}(\theta_2)^T\overline{\mathbf{a}(\theta_1)}\right| = \left|\frac{\sin\Big(\pi(N-1)(\nu(\theta_1)-\nu(\theta_2)\Big)}{\sin\left(\pi(\nu(\theta_1)-\nu(\theta_2))\right)}\right|$$



SIMO model (1)

- M colocated isotropic received antennas
- One far-field source backscattering signal $\tilde{x}_p(t)$ to the RX.

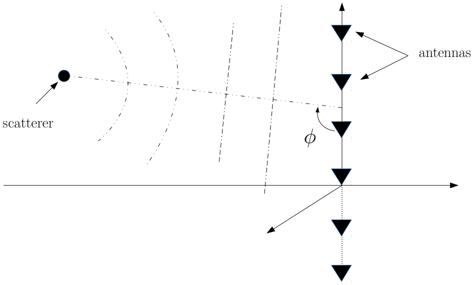


Figure: Example with ULA at RX

SIMO model (2)

- The complex envelope of the signal received by the RX array from the p-th scatterer is modelled as (cf. array processing chapter)

$$\mathbf{r}_p(t) = \beta_p \mathrm{e}^{-\mathrm{i}2\pi f_c \bar{T}_{1,p}} \, \mathbf{b}\left(\bar{\theta}_p, \bar{\phi}_p\right) x_p\left(t - \bar{T}_{1,p}\right)$$

where

  - β_p is a fading coefficient between p-th scatterer/RX
  - $\bar{T}_{m,p}$ is the propagation delay between p-th scatterer/m-th antenna
  - $\bar{\theta}_p, \bar{\phi}_p$ are the elevation and azimuth of the scatterer
  - $\mathbf{b}(\bar{\theta}, \bar{\phi})$ is the steering vector associated with the received M antenna array.

SIMO model (3)

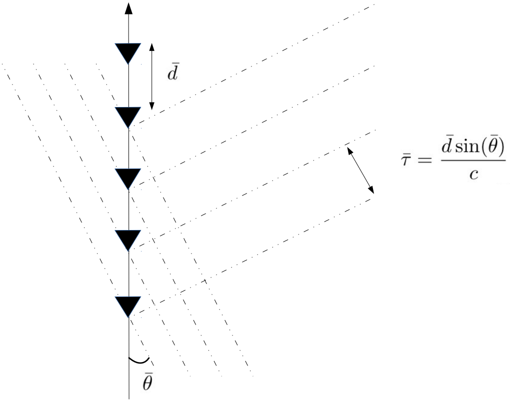


Figure: Example with ULA at RX, $\mathbf{b}(\bar{\theta}) = \left(1, \mathrm{e}^{-\mathrm{i}2\pi \frac{\bar{d}}{\lambda_c} \sin(\bar{\theta})}, \ldots, \mathrm{e}^{-\mathrm{i}2\pi(M-1)\frac{\bar{d}}{\lambda_c} \sin(\bar{\theta})}\right)^T$

MIMO model (1)

- Taking into account the P scatterers and additive noise,

$$\begin{aligned}\mathbf{r}(t) &= \sum_{p=1}^{P} \mathbf{r}_p(t) + \mathbf{v}(t)\\ &= \sum_{p=1}^{P} \gamma_p \mathbf{b}\left(\bar{\theta}_p, \bar{\phi}_p\right) \mathbf{a}(\theta_p, \phi_p)^T \mathbf{s}(t-\delta_p) + \mathbf{v}(t)\end{aligned}$$

  where

  - $\gamma_p = \alpha_p \beta_p \mathrm{e}^{-\mathrm{i}2\pi f_c \delta_p}$
  - $\delta_p = T_{1,p} + \bar{T}_{1,p}$.

MIMO model (2)

- The transmit signal used is given by

$$\mathbf{s}(t) = \sum_{k \in \mathbb{Z}} \begin{pmatrix} s_{k,1} \\ \vdots \\ s_{k,N} \end{pmatrix} p(t - kT_s)$$

  where $p(t)$ is a shaping filter common to the N TX antennas.

- We use $g(t)$ as a filter common to the M RX antennas.

MIMO model (3)

MIMO model - continuous time

The received signal after filtering is modelled as

$$\mathbf{y}(t) = (\mathbf{r} \star g)(t)$$

$$= \sum_{k \in \mathbb{Z}} \mathbf{H}(t - kT_s)\mathbf{s}_k + \mathbf{w}(t)$$

with

- $\mathbf{H}(t) = \sum_{p=1}^{P} \gamma_p \mathbf{b}\left(\bar{\theta}_p, \bar{\phi}_p\right) \mathbf{a}(\theta_p, \phi_p)^T (p \star g)(t - \delta_p)$ the MIMO channel impulse response
- $\mathbf{w}(t) = \left((v_1 \star g)(t), \ldots, (v_M \star g)(t)\right)^T$ the filtered noise

MIMO model (4)

MIMO model - discrete time

After sampling at frequency $\frac{1}{T_s}$, we get

$$\mathbf{y}_n = \mathbf{y}(nT_s)$$

$$= \sum_{k \in \mathbb{Z}} \mathbf{H}_k \mathbf{s}_{n-k} + \mathbf{w}_n$$

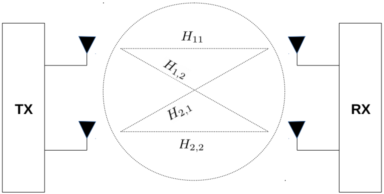
where $\mathbf{H}_k = \mathbf{H}(kT_s)$ and $\mathbf{w}_n = \mathbf{w}(nT_s)$.

MIMO model (5)

- **Remark.** In practice, we consider finite-length and causal channels:

$$\mathbf{y}_n = \sum_{k=0}^{K} \mathbf{H}_k s_{n-k} + \mathbf{w}_n$$

  - If $K > 1$, the channel is said to be *frequency selective*.
  - If $K = 1$, the channel is said to be *flat fading*: $\mathbf{y}_n = \mathbf{H}\mathbf{s}_n + \mathbf{w}_n$.



MIMO channel matrix

Kronecker model (1)

- **Assumption 1.** The scatterers are colocated and induce approximately the same delays

$$\delta_1 = \ldots = \delta_P = \delta$$

- **Assumption 2.** $\gamma_1, \ldots, \gamma_p$ are modelled as i.i.d. complex circular rv with variance $\frac{1}{P}$ (normalized total transmit energy is spread uniformly among the scatterers).

- **Assumption 3.**
  - The DoD and DoA (θ_1, ϕ_1), $(\bar{\theta}_1, \bar{\phi}_1)$, ..., (θ_P, ϕ_P), $(\bar{\theta}_P, \bar{\phi}_P)$ are independent, and mutually independent of $\gamma_1, \ldots, \gamma_P$.
  - $(\theta_1, \phi_1), \ldots, (\theta_P, \phi_P)$ are i.i.d., distributed as a generic rv (θ, ϕ)
  - $(\bar{\theta}_1, \bar{\phi}_1), \ldots, (\bar{\theta}_P, \bar{\phi}_P)$ are i.i.d., distributed as a generic rv $(\bar{\theta}, \bar{\phi})$

Kronecker model (2)

- **TLC.** As $P \to \infty$, the channel matrix

$$\mathbf{H} = (p \star g)(0) \sum_{p=1}^{P} \gamma_p \mathbf{b}\left(\bar{\theta}_p, \bar{\phi}_p\right) \mathbf{a}(\theta_p, \phi_p)^T$$

  converge in distribution to a Gaussian random matrix, with correlated entries.

- **Correlation analysis.** Assuming $(p \star g)(0) = 1$, straightfoward computations show that

$$\mathbb{E}\left[h_{i,j}\overline{h}_{k,l}\right] = \mathbb{E}\left[[\mathbf{b}\left(\bar{\theta}, \bar{\phi}\right)\mathbf{a}\left(\theta, \phi\right)^T]_{i,j}\overline{[\mathbf{b}\left(\bar{\theta}, \bar{\phi}\right)\mathbf{a}\left(\theta, \phi\right)^T]_{l,k}}\right]$$

$$= r_{i,l}\ \tilde{r}_{j,k}$$

  where

$$r_{i,l} = \mathbb{E}\left[[\mathbf{b}\left(\bar{\theta}, \bar{\phi}\right)]_i\ \overline{[\mathbf{b}\left(\bar{\theta}, \bar{\phi}\right)]_l}\right] \text{ and } \tilde{r}_{j,k} = \mathbb{E}\left[[\mathbf{a}\left(\theta, \phi\right)]_j\overline{[\mathbf{a}\left(\theta, \phi\right)]_k}\right].$$

Kronecker model (3)

- **RX covariance matrix.**

$$\mathbf{R} = (r_{i,l})_{i,l=1,\ldots,M} = \mathbb{E}\left[\mathbf{b}\left(\bar{\theta},\bar{\phi}\right)\mathbf{b}\left(\bar{\theta},\bar{\phi}\right)^*\right]$$

- **TX covariance matrix.**

$$\tilde{\mathbf{R}} = (\tilde{r}_{j,k})_{j,k=1,\ldots,N} = \mathbb{E}\left[\mathbf{a}\left(\theta,\phi\right)\mathbf{a}\left(\theta,\phi\right)^*\right]$$

Kronecker model

Under the previous assumptions, and if the number of scatterers P is large, one can model $\mathbf{H}$ as

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

where $\mathbf{X}$ is a $M \times N$ random matrix with i.i.d $\mathcal{N}_{\mathbb{C}}(0,1)$ entries.

Kronecker model for a ULA

- If the DoA are uniformly spread in the angular sector $\bar{\Theta}$, i.e.

$$\bar{\theta} \sim \mathcal{U}\left(\bar{\Theta}\right)$$

then

$$r_{k,l} = \frac{1}{|\bar{\Theta}|} \int_{\bar{\Theta}} \mathrm{e}^{-\mathrm{i}2\pi \frac{\bar{d}}{\lambda_c}(k-l)\sin(\bar{\theta})} \mathrm{d}\bar{\theta} = \frac{1}{|\bar{\Theta}|} \int_{\sin(\bar{\Theta})} \frac{\mathrm{e}^{-\mathrm{i}2\pi \frac{\bar{d}}{\lambda_c}(k-l)u}}{\sqrt{1-u^2}} \mathrm{d}u.$$

so that

$$\mathbf{R} \xrightarrow[\frac{\bar{d}}{\lambda_c} \to \infty]{} \mathbf{I}.$$

- If $\bar{d} \gg \lambda_c$, the RX antennas may be considered as "statistically uncorrelated" (id. for TX antennas).

Rice model

Rice model

Under the assumptions of the Kronecker model, and in the presence of a direct path, one can model $\mathbf{H}$ as

$$\mathbf{H} = \sqrt{\frac{\kappa}{\kappa+1}}\mathbf{C} + \sqrt{\frac{1}{\kappa+1}}\mathbf{R}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{R}}^{\frac{1}{2}}$$

where

- $\mathbf{C} = \mathbf{b}\left(\bar{\theta}_0, \bar{\phi}_0\right)\mathbf{a}\left(\theta_0, \phi_0\right)^T$ with $\left(\bar{\theta}_0, \bar{\phi}_0\right)$ and $\left(\theta_0, \phi_0\right)$ deterministic
- κ is the ratio between the energy carried by the direct path over the energy reflected by the scatterers.

Outline

Reminder on channel codes (1)

- **Context.** We consider a flat fading $M \times N$ MIMO channel. On a symbol time, we receive

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

where

- $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the channel matrix,
- $\mathbf{v} \sim \mathcal{N}_{\mathbb{C}^M}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ is the additive Gaussian noise, assumed spatially white.
- $\mathbf{x} \in \mathbb{C}^N$ is the transmit signal, satisfying the power constraint

$$\mathbb{E}\,\|\mathbf{x}\|_2^2 \leq P \quad \Leftrightarrow \quad \mathrm{tr}\mathbf{Q} \leq P$$

where $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^*]$.

Reminder on channel codes (2)

- **Code.** A code is a set of codewords $\mathcal{C} = \{\mathbf{X}_1, \ldots, \mathbf{X}_{|\mathcal{C}|}\}$, where for $k = 1, \ldots, |\mathcal{C}|$, $\mathbf{X}_k$ is a $N \times L$ matrix of symbols, with L the codelength.

  In the context of MIMO communications, such codes are refered to as Space-Time Block Codes (STBC).

- **Rate.** $R = \frac{\log|\mathcal{C}|}{L}$.

- **Decoding function.** $\phi : \mathbb{C}^{M\times L} \mapsto \mathcal{C}$.

- **Error probability.** If $\mathbf{X}$ is a (randomly) transmitted codeword and if we receive the $M \times L$ matrix $\mathbf{Y}$, then

$$P_e = \mathbb{P}\left(\phi(\mathbf{Y}) \neq \mathbf{X}\right).$$

Reminder on channel codes (3)

- **Achievability.** A rate $R > 0$ is said to be *achievable* if there exists a sequence of STBC $(\mathcal{C}_L)_{L \geq 1}$ with length L, rate R_L and error probability $P_{e,L}$ such that

$$R_L \xrightarrow[L \to \infty]{} R \quad \text{and} \quad P_{e,L} \xrightarrow[L \to \infty]{} 0.$$

- **Channel capacity.** The capacity of the MIMO channel is defined as

$$C = \sup \left\{ R \geq 0 : R \text{ is achievable} \right\}.$$

Reminder on the SVD (1)

- **SVD.** Any $M \times N$ complex matrix $\mathbf{H}$ can be factorized as

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

where

  - $\mathbf{\Sigma} = (\sigma_{i,j})$ is a $M \times N$ diagonal matrix of singular values, with

$$\sigma_{1,1} \geq \ldots \geq \sigma_{r,r} > 0$$

    and $\sigma_{i,i} = 0$ for all $i > r$, with $r = \mathrm{rank}(\mathbf{H})$.

  - $\mathbf{U}$ is a $M \times M$ unitary matrix of left singular vectors.

  - $\mathbf{V}$ is $N \times N$ matrix of right singular vectors.

Reminder on the SVD (2)

- **EVD (1).** An eigenvalue decomposition of matrix [1] $\mathbf{H}\mathbf{H}^* \succeq \mathbf{0}$ is given by

$$\mathbf{H}\mathbf{H}^* = \mathbf{U} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_M \end{pmatrix} \mathbf{U}^*,$$

where $\lambda_1 = \sigma_{1,1}^2 \geq \ldots \lambda_r = \sigma_{r,r}^2 > \lambda_{r+1} = \ldots = \lambda_M = 0$.

- **EVD (2).** An eigenvalue decomposition of matrix $\mathbf{H}^*\mathbf{H} \succeq \mathbf{0}$ is given by

$$\mathbf{H}^*\mathbf{H} = \mathbf{V} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix} \mathbf{V}^*,$$

where $\lambda_1 = \sigma_{1,1}^2 \geq \ldots \lambda_r = \sigma_{r,r}^2 > \lambda_{r+1} = \ldots = \lambda_N = 0$.

[1]**N.B.** $\mathbf{A} \succeq \mathbf{0} \Leftrightarrow \mathbf{A}$ is positive semidefinite.

Capacity for a deterministic channel $\mathbf{H}$ (1)

- **Information capacity.** The *information capacity* of the channel is defined by

$$C = \sup_{f_{\mathbf{x}}:\ \mathbb{E}\|\mathbf{x}\|_2^2 \leq P} \mathbb{I}(\mathbf{x};\mathbf{y})$$

where $\mathbb{I}(\mathbf{x};\mathbf{y})$ is the mutual information between the input/output of the channel.

- **Mutual information.**

$$\begin{aligned} \mathbb{I}\left(\mathbf{x};\mathbf{y}\right) &= \mathbb{H}(\mathbf{y}) - \mathbb{H}(\mathbf{y}|\mathbf{x}) \\ &= \mathbb{H}(\mathbf{y}) - M\log(\pi e \sigma^2). \end{aligned}$$

where $\mathbb{H}(.)$ is the entropy and $\mathbb{H}(.|.)$ the conditional entropy.

Capacity for a deterministic channel $\mathbf{H}$ (2)

- **Output entropy.** Since $\mathbb{E}[\mathbf{y}\mathbf{y}^*] = \mathbf{H}\mathbf{Q}\mathbf{H}^* + \sigma^2\mathbf{I}$ with $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^*]$, we have

$$\mathbb{H}(\mathbf{y}) \leq \log\Big((\pi e)^M \det\left(\mathbf{H}\mathbf{Q}\mathbf{H}^* + \sigma^2\mathbf{I}\right)\Big)$$

with equality iff $\mathbf{y} \sim \mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{H}\mathbf{Q}\mathbf{H}^* + \sigma^2\mathbf{I})$, which is verified if $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}^N}(\mathbf{0}, \mathbf{Q})$.

Information capacity (1st formula)

$$C = \sup_{\substack{\mathbf{Q}\succeq\mathbf{0} \\ \mathrm{tr}\mathbf{Q}\leq P}} \operatorname{logdet}\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^*}{\sigma^2}\right)$$

Capacity for a deterministic channel $\mathbf{H}$ (3)

- Consider the EVD of matrix $\mathbf{H}^*\mathbf{H} \succeq \mathbf{0}$

$$\mathbf{H}^*\mathbf{H} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^*,$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, \underbrace{0, \ldots, 0}_{N-r})$ and $\mathbf{V}$ is $N \times N$ unitary.

- Since $\mathbf{Q} \mapsto \mathbf{V}^*\mathbf{Q}\mathbf{V}$ is an isometric isomorphism conserving the trace,

$$C = \sup_{\substack{\mathbf{Q}\succeq\mathbf{0} \\ \mathrm{tr}\mathbf{Q}\leq P}} \mathrm{logdet}\left(\mathbf{I} + \frac{\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^*\mathbf{Q}\mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}}{\sigma^2}\right) = \sup_{\substack{\mathbf{Q}\succeq\mathbf{0} \\ \mathrm{tr}\mathbf{Q}\leq P}} \mathrm{logdet}\left(\mathbf{I} + \frac{\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}}{\sigma^2}\right)$$

Capacity for a deterministic channel $\mathbf{H}$ (4)

- Using Hadamard's inequality,

$$\operatorname{logdet}\left(\mathbf{I} + \frac{\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{Q}\mathbf{\Lambda}^{\frac{1}{2}}}{\sigma^2}\right) \leq \sum_{i=1}^{r} \log\left(1 + \frac{q_i \lambda_i}{\sigma^2}\right)$$

with equality iff $\mathbf{Q} = \mathrm{diag}\,(q_1, \ldots, q_N)$.

Information capacity (2nd formula)

$$C = \sup_{\substack{q_1,\ldots,q_r \geq 0 \\ q_1+\ldots+q_r \leq P}} \sum_{i=1}^{r} \log\left(1 + \frac{q_i \lambda_i}{\sigma^2}\right)$$

Capacity for a deterministic channel $\mathbf{H}$ (5)

- **Water-Filling.** Using KKT Theorem, there exists unique maximizers $q_1^*, \ldots, q_r^*$ such that

$$q_i^* = \left(\frac{1}{\gamma^*} - \frac{1}{\frac{\lambda_i}{\sigma^2}}\right)^+ \quad \forall i = 1, \ldots, r$$

where γ^* is the unique solution to the equation $f\left(\frac{1}{\gamma}\right) = P$ with

$$f\left(\frac{1}{\gamma}\right) = \sum_{i=1}^{r} \left(\frac{1}{\gamma} - \frac{1}{\frac{\lambda_i}{\sigma^2}}\right)^+ .$$

Capacity for a deterministic channel $\mathbf{H}$ (6)

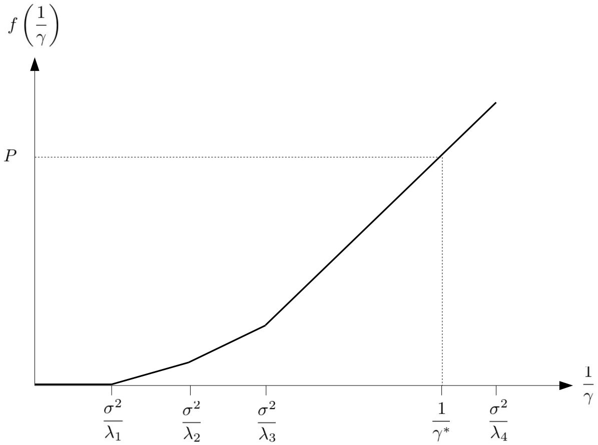


Figure: Illustration of the water-filling problem

Capacity for a deterministic channel $\mathbf{H}$ (7)

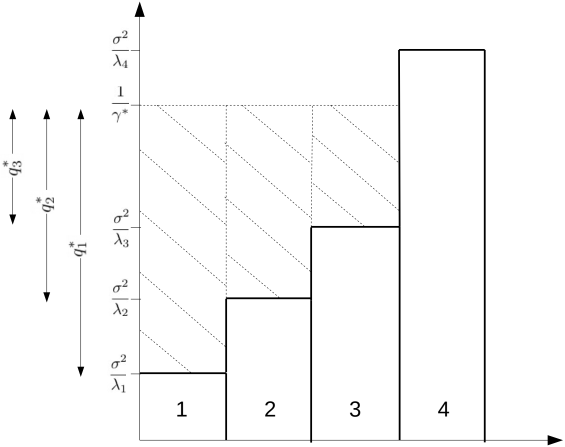


Figure: Illustration of the water-filling problem (no power allocated to sub-channel 4)

Precoding/postcoding and optimal power allocation (1)

- From the previous study, the optimal capacity-achieving covariance matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \mathbf{V} \begin{pmatrix} q_1^* & & & & & \\ & \ddots & & & & \\ & & q_r^* & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \mathbf{V}^*$$

  where the allocated powers $q_1^*, \ldots, q_r^*$ are adjusted by water-filling, and $\mathbf{V}$ is the $N \times N$ unitary eigenvectors matrix of $\mathbf{H}^*\mathbf{H}$.

- **Remark.** The knowledge of the channel matrix $\mathbf{H}$ is required, which may be unrealistic depending on the context.

Precoding/postcoding and optimal power allocation (2)

- **Transmit signal/Precoding.** We use

$$\mathbf{x} = \mathbf{V} \begin{pmatrix} \sqrt{q_1^*} & & & & & \\ & \ddots & & & & \\ & & \sqrt{q_r^*} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \mathbf{s}$$

where $\mathbf{s} \in \mathbb{C}^N$ is a vector of transmit symbols, such that

$$\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \mathrm{diag}\left(\underbrace{1, \ldots, 1}_{r \text{ times}}, 0, \ldots, 0\right)$$

Precoding/postcoding and optimal power allocation (3)

- **Receive signal.** Using the SVD of $\mathbf{H}$, the receive signal writes

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

$$= \mathbf{U} \begin{pmatrix} \sqrt{q_1^* \lambda_1} s_1 \\ \vdots \\ \sqrt{q_r^* \lambda_r} s_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathbf{v}$$

Precoding/postcoding and optimal power allocation (4)

- **Postcoding.** The receiver computes

$$\mathbf{z} = \mathbf{U}^* \mathbf{y}$$

$$= \begin{pmatrix} \sqrt{q_1^* \lambda_1} s_1 \\ \vdots \\ \sqrt{q_r^* \lambda_r} s_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathbf{w}$$

where $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}^M}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$.

Precoding/postcoding and optimal power allocation (5)

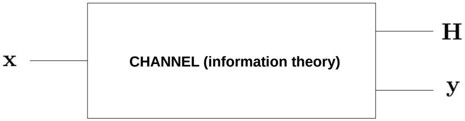
Summary of precoding/postcoding

- After precoding/postcoding, we retrieve $r = \mathrm{rank}(\mathbf{H})$ parallel AWGN subchannels, with optimized SNR (the channel capacity and the information capacity coincides). Thus, a maximum of r symbols per channel use can be transmitted.
- Precoding/postcoding may be interpreted as a beamforming along "virtual" directions (i.e. along the singular vectors of $\mathbf{H}$).
- Precoding/postcoding and optimal power allocation can only be performed if $\mathbf{H}$ is known at both the transmitter and the receiver.

Capacity for a fast fading Gaussian $\mathbf{H}$ (1)

- **Model.** The channel matrix $\mathbf{H}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries (and independent of $\mathbf{x}$), changes at each symbol time (fast fading) and is known to the receiver.

- **IT model.**



- **Information capacity (1).**

$$C = \sup_{f_{\mathbf{x}}:\ \mathbb{E}\|\mathbf{x}\|_2^2 \leq P} \mathbb{I}\left(\mathbf{x};(\mathbf{y},\mathbf{H})\right)$$

Capacity for a fast fading Gaussian $\mathbf{H}$ (2)

- **Information capacity (2).** Conditioning on $\mathbf{H}$, we get

$$C = \sup_{f_{\mathbf{x}}:\ \mathbb{E}\|\mathbf{x}\|_2^2 \leq P} \mathbb{I}\left(\mathbf{x};\mathbf{H}\right) + \mathbb{I}\left(\mathbf{x};\mathbf{y}|\mathbf{H}\right)$$

where

$$\begin{aligned} \mathbb{I}\left(\mathbf{x};\mathbf{y}|\mathbf{H}\right) &:= \mathbb{H}(\mathbf{x}|\mathbf{H}) - \mathbb{H}(\mathbf{x}|\mathbf{y},\mathbf{H}) \\ &:= \mathbb{E}\left[J(\mathbf{H})\right] \end{aligned}$$

with

$$\mathbf{Z} \mapsto J(\mathbf{Z}) = \int\int f(\mathbf{x},\mathbf{y}|\mathbf{H}=\mathbf{Z}) \log\left(\frac{f(\mathbf{x},\mathbf{y}|\mathbf{H}=\mathbf{Z})}{f(\mathbf{x}|\mathbf{H}=\mathbf{Z})f(\mathbf{y}|\mathbf{H}=\mathbf{Z})}\right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

Capacity for a fast fading Gaussian $\mathbf{H}$ (3)

- **Information capacity (3).** Since $\mathbf{x}$ and $\mathbf{H}$ are independent,

$$C = \sup_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \mathrm{tr}\mathbf{Q} \leq P}} \sup_{f_{\mathbf{x}}:\ \mathbb{E}[\mathbf{x}\mathbf{x}^*] = \mathbf{Q}} \mathbb{E}[J(\mathbf{H})].$$

- **Information capacity (4).** From the previous section,

$$\sup_{f_{\mathbf{x}}:\ \mathbb{E}[\mathbf{x}\mathbf{x}^*] = \mathbf{Q}} J(\mathbf{H}) = \mathrm{logdet}\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^*}{\sigma^2}\right).$$

with equality if $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}^N}(\mathbf{0}, \mathbf{Q})$.

Capacity for a fast fading Gaussian $\mathbf{H}$ (4)

Information capacity (1st formula)

$$C = \sup_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \mathrm{tr}\mathbf{Q} \leq P}} \mathbb{E}\left[\log\det\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^*}{\sigma^2}\right)\right].$$

Capacity for a fast fading Gaussian $\mathbf{H}$ (5)

- **Information capacity (6).** Using the unitary invariance of the standard complex Gaussian distribution, namely

$$\mathbf{HU} \stackrel{\mathcal{D}}{=} \mathbf{H}$$

for any $N \times N$ unitary matrix $\mathbf{U}$, we deduce from the EVD of $\mathbf{Q}$ that

$$C = \sup_{\substack{\mathbf{D} \succeq \mathbf{0} \text{ diagonal} \\ \mathrm{tr}\mathbf{D} \leq P}} \underbrace{\mathbb{E}\left[\operatorname{logdet}\left(\mathbf{I} + \frac{\mathbf{HDH}^*}{\sigma^2}\right)\right]}_{=\Psi(\mathbf{D})}.$$

- **Information capacity (6).** Moreover, if $\mathcal{P}_N$ is the set of $N \times N$ permutation matrices,

$$\Psi(\mathbf{D}) = \Psi(\mathbf{PDP}^*) \quad \forall \mathbf{P} \in \mathcal{P}_N.$$

Capacity for a fast fading Gaussian $\mathbf{H}$ (6)

- **Information capacity (7).** Using the concavity of $\mathbf{D} \mapsto \Psi(\mathbf{D})$, we obtain

$$\begin{aligned}\Psi(\mathbf{D}) &= \frac{1}{|\mathcal{P}_N|} \sum_{\mathbf{P} \in \mathcal{P}_N} \Psi(\mathbf{P}\mathbf{D}\mathbf{P}^*) \\ &\leq \Psi\Bigg(\underbrace{\frac{1}{|\mathcal{P}_N|} \sum_{\mathbf{P} \in \mathcal{P}_N} \mathbf{P}\mathbf{D}\mathbf{P}^*}_{= \text{multiple of the identity}} \Bigg).\end{aligned}$$

  where $|\mathcal{P}_N| = N!$ is the cardinality of $\mathcal{P}_N$.

- **Information capacity (8).** It follows that $\mathbf{D} \mapsto \Psi(\mathbf{D})$ achieves its maximum for $\mathbf{D} = \alpha \mathbf{I}$, with $\alpha \geq 0$.

Capacity for a fast fading Gaussian $\mathbf{H}$ (7)

Information capacity (2nd formula)

The information capacity is given by

$$C = \mathbb{E}\left[\operatorname{logdet}\left(\mathbf{I} + \frac{P}{N\sigma^2}\mathbf{H}\mathbf{H}^*\right)\right]$$

and is achieved for $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}^N}(\mathbf{0}, \frac{P}{N}\mathbf{I})$.

- **Remark (1).** This information capacity is usually called *ergodic capacity* and coincides with the channel capacity if $\mathbf{H}$ is known at the receiver.
- **Remark (2).** Uniform power allocation $\frac{P}{N}$ per antenna, and no specific direction favored.
- **Remark (3).** No simple expression (i.e. by solving expectation) is known.

Large SNR analysis (1)

- The ergodic capacity may be written as

$$
\begin{aligned}
C(\rho) &= \mathbb{E}\left[\operatorname{logdet}\left(\mathbf{I} + \rho \mathbf{H}\mathbf{H}^*\right)\right] \\
&= \mathbb{E}[\mathsf{rank}(\mathbf{H})] \log(\rho) + \mathbb{E}\left[\sum_{i=1}^{\mathsf{rank}(\mathbf{H})} \log\left(\lambda_i(\mathbf{H}\mathbf{H}^*) + \frac{1}{\rho}\right)\right]
\end{aligned}
$$

  where $\lambda_1(\mathbf{H}\mathbf{H}^*) \geq \ldots \geq \lambda_M(\mathbf{H}\mathbf{H}^*) \geq 0$ are the eigenvalues of $\mathbf{H}\mathbf{H}^*$ and ρ is the SNR per antenna.

- It holds that

$$
\mathbb{P}\left(\mathsf{rank}(\mathbf{H}) = \min(M, N)\right) = 1
$$

$$
\mathbb{E}\left[\sum_{i=1}^{\mathsf{rank}(\mathbf{H})} \log\left(\lambda_i(\mathbf{H}\mathbf{H}^*) + \frac{1}{\rho}\right)\right] \xrightarrow[\rho \to \infty]{} \sum_{i=1}^{\min(M,N)} \mathbb{E}\left[\log\left(\lambda_i(\mathbf{H}\mathbf{H}^*)\right)\right].
$$

Large SNR analysis (2)

Behaviour for large SNR

As $\rho \to \infty$,

$$C(\rho) = \min(M, N) \log(\rho) + \mathcal{O}(1).$$

- **Remark (1).** At high SNR, the Gaussian MIMO ergodic capacity behaves as the capacity of $\min(M, N)$ scalar Gaussian channels.
- **Remark (2).** At high SNR, the ergodic capacity may be further increased by increasing *both* M and N.

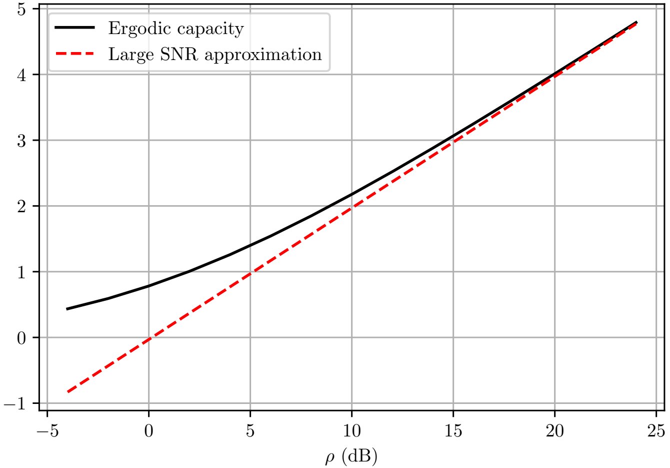


Figure: Ergodic capacity and large SNR approximation ($M = N = 2$)

Large N analysis (1)

- As $N \to \infty$, the LLN implies

$$\frac{\mathbf{H}\mathbf{H}^*}{N} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{h}_n \mathbf{h}_n^* \xrightarrow[N\to\infty]{a.s.} \mathbf{I}.$$

Behaviour for large N

$$C\left(\frac{P}{N\sigma^2}\right) \xrightarrow[N\to\infty]{} M \log\left(1 + \frac{P}{\sigma^2}\right).$$

- **Remark.** Increasing the number of transmit antennas N to increase the capacity is pointless if the number of receive antennas M does not increase as well.

Large N analysis (2)

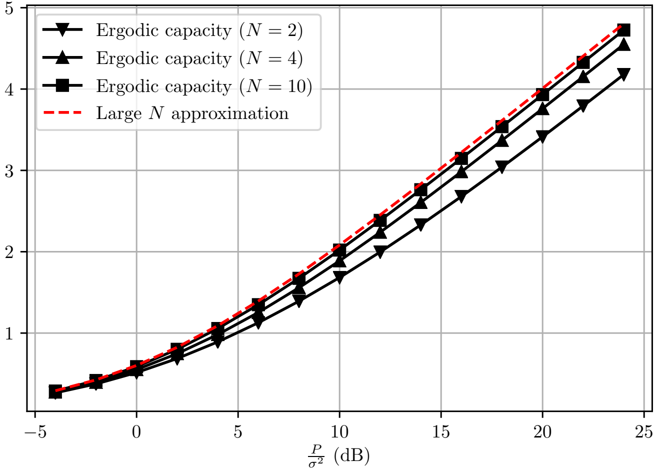


Figure: Ergodic capacity and large N approximation ($M = 2$) vs. total SNR $\frac{P}{\sigma^2}$

Large M analysis (1)

- Using the fact that $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$,

$$\mathbb{E}\left[\operatorname{logdet}\left(\mathbf{I} + \mathbf{H}\mathbf{H}^* \frac{P}{N\sigma^2}\right)\right] = N\log(M) + \mathbb{E}\left[\operatorname{logdet}\left(\mathbf{I} + \frac{\mathbf{H}^*\mathbf{H}}{M}\frac{P}{N\sigma^2}\right)\right]$$

Behaviour for large M

As $M \to \infty$,

$$C\left(\frac{P}{N\sigma^2}\right) = N\log(M) + N\log\left(\frac{P}{N\sigma^2}\right) + o(1).$$

- **Remark.** Increasing the number of receive antennas M "slowly" increases the capacity (i.e. with $\log$ speed).

Large M analysis (2)

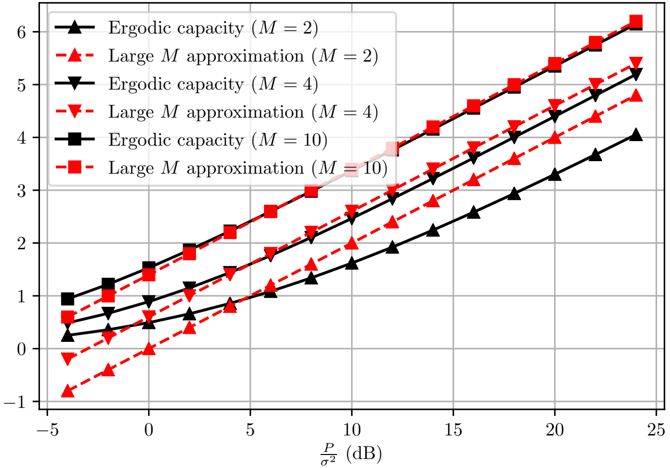


Figure: Ergodic capacity and large M approximation ($N = 2$) vs. total SNR $\frac{P}{\sigma^2}$

Capacity for a block fading Gaussian $\mathbf{H}$ (1)

- **Model.** The channel matrix $\mathbf{H}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries (and independent of $\mathbf{x}$), is constant during the transmission of a codeword (block fading) and is known to the receiver.

- For a fixed realization of $\mathbf{H}$, the maximum achievable rate is given by

$$C(\mathbf{H}) = \sup_{\substack{\mathbf{Q}\succeq\mathbf{0} \\ \mathrm{tr}\mathbf{Q}\leq P}} \operatorname{logdet}\left(\mathbf{I} + \frac{\mathbf{HQH}^*}{\sigma^2}\right).$$

- If $R > 0$ is a fixed target rate, then

$$\mathbb{P}\left(C(\mathbf{H}) < R\right) > 0.$$

$\Rightarrow$ The channel capacity is zero.

Capacity for a block fading Gaussian $\mathbf{H}$ (2)

- In the context of block fading channels, we use the concepts of *outage probability* and *outage capacity* to extend the notion of channel capacity.

- **Outage probability.** For a fixed target rate R, we define

$$P_{\mathsf{out}}(R) = \inf_{\substack{\mathbf{Q}\succeq \mathbf{0} \\ \mathrm{tr}\mathbf{Q}\leq P}} \mathbb{P}\left(\log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^*}{\sigma^2}\right) < R\right).$$

- **Outage capacity.** For $\epsilon > 0$, we define

$$C_{\mathsf{out}}(\epsilon) = \sup\left\{R > 0 : P_{\mathsf{out}}(R) \leq \epsilon\right\}.$$

Literature

- **SU-MIMO.** Foschini & Ganz'98 [6], Telatar'99 [12], Zheng & Tse'02 [16], Jorswieck & Boche [8]
- **MU-MIMO.** Viswanath & Tse'03 [14], Caire & Shamai'03 [2], Viswanath et al.'01 [15]