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)
  - ⊕ flexible data rates
  - ⊖ synchronization/equalization, roaming, multipath interference
- **FMDA.** Frequency Division Multiple Access (1G, 2G)
  - ⊕ simplicity
  - ⊖ fading, inefficient use of bandwidth, static data rates
- **CMDA.** Code Division Multiple Access (3G)
  - ⊕ random access
  - ⊖ receiver complexity, complex power allocation
- **OFDMA.** Orthogonal Frequency Division Multiple Access (4G)
  - ⊕ simplicity, reduced multi-user interference and fading
  - ⊖ 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)

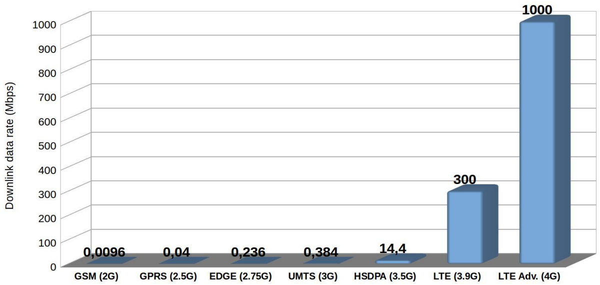


Figure: Evolution of downlink data rates (Mbps), from 2G to 4G

- 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)

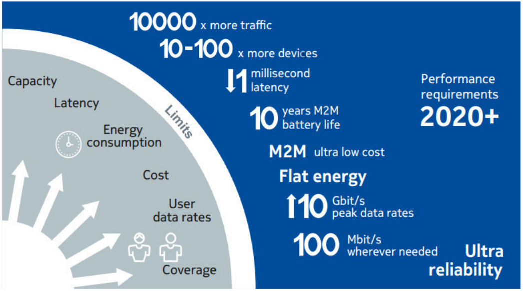
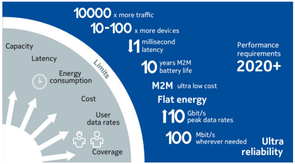


Figure: Requirements for future 2020 mobile standards (*source: Nokia*)

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)

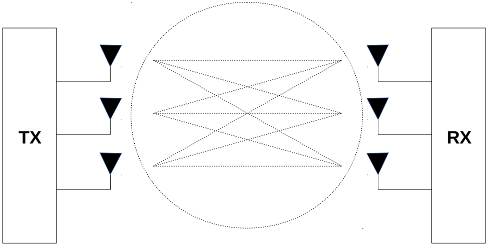
Challenges

- 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)

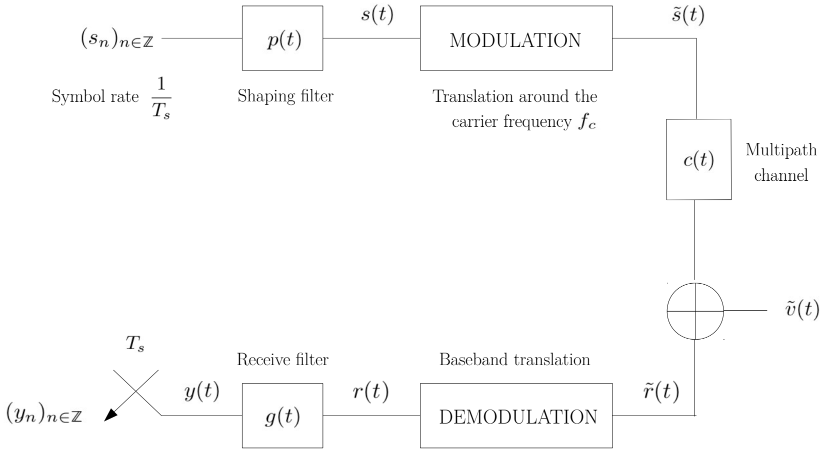


Figure: SISO transmit/receive chain

SISO model (2)

- **Transmit signal.** $\tilde{s}(t) = \mathrm{Re}\left(s(t)\mathrm{e}^{\mathrm{i}2\pi f_c t}\right)$ with

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

- **Receive signal.** $\tilde{r}(t) = \mathrm{Re}\left(r(t)\mathrm{e}^{\mathrm{i}2\pi f_c t}\right)$ 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.**

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

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

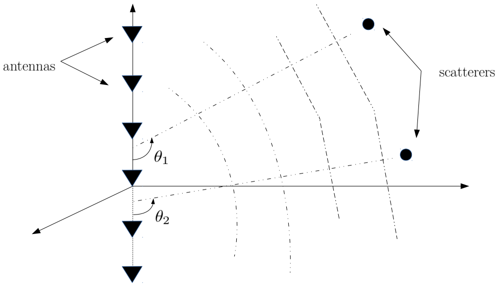


Figure: Example with a ULA at TX

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)

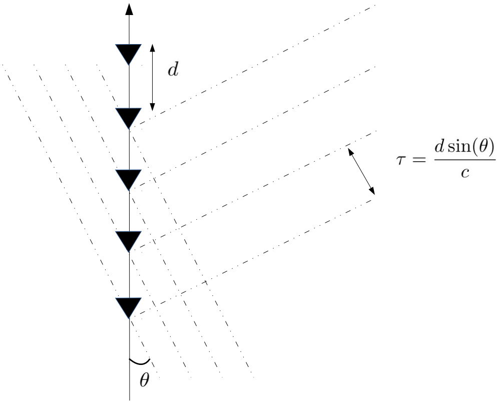


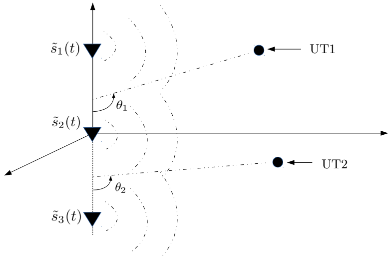
Figure: Example with ULA at TX, $\mathbf{a}(\theta) = \left(1, \mathrm{e}^{-\mathrm{i}2\pi\frac{d}{\lambda_c}\sin(\theta)}, \ldots, \mathrm{e}^{-\mathrm{i}2\pi(N-1)\frac{d}{\lambda_c}\sin(\theta)}\right)^T$

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)

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



Digression on a 1st type of beamforming (2)

- **Signal received at UT1.**

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

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

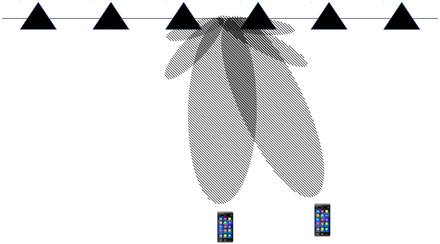
$+$ noise.

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.

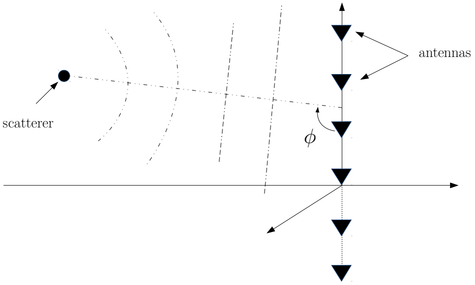


Figure: Example with ULA at RX

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)

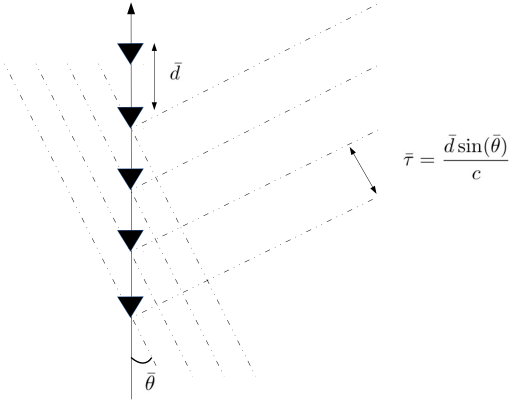


Figure: Example with ULA at RX, $\mathbf{b}(\bar{\theta}) = \left(1, \mathrm{e}^{-\mathrm{i}2\pi\frac{\bar{d}}{\lambda_c}\sin(\bar{\theta})}, \ldots, \mathrm{e}^{-\mathrm{i}2\pi(M-1)\frac{\bar{d}}{\lambda_c}\sin(\bar{\theta})}\right)^T$

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

$$\begin{aligned}\mathbf{y}_n &= \mathbf{y}(nT_s) \\ &= \sum_{k\in\mathbb{Z}} \mathbf{H}_k \mathbf{s}_{n-k} + \mathbf{w}_n\end{aligned}$$

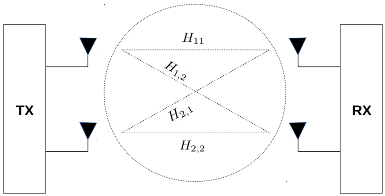
where $\mathbf{H}_k = \mathbf{H}(kT_s)$ and $\mathbf{w}_n = \mathbf{w}(nT_s)$.

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.**

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

- **TX covariance matrix.**

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

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] $\mathbf{HH}^* \succeq \mathbf{0}$ is given by

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

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 $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_r, \underbrace{0, \ldots, 0}_{N-r})$ 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,

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

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 $\mathbf{Q} = \text{diag}\,(q_1, \ldots, q_N)$.

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)

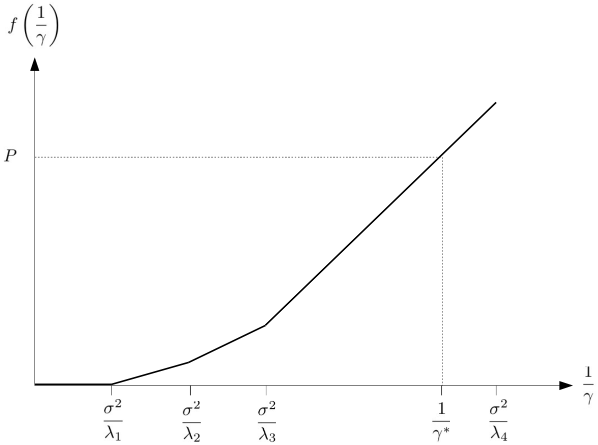


Figure: Illustration of the water-filling problem

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

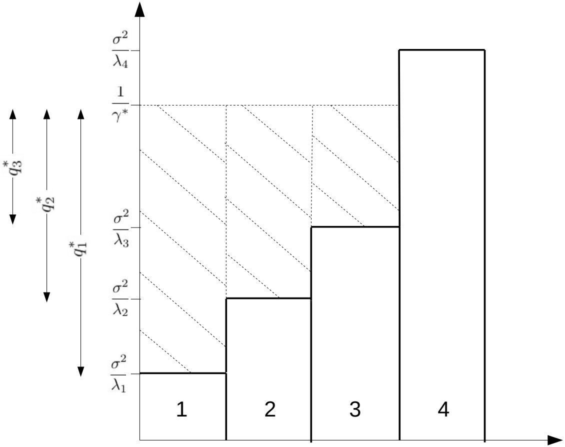


Figure: Illustration of the water-filling problem (no power allocated to sub-channel 4)

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)

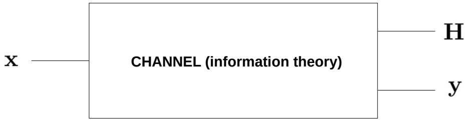
Summary of precoding/postcoding

- 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,

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

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

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

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

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

  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.

Large SNR analysis (3)

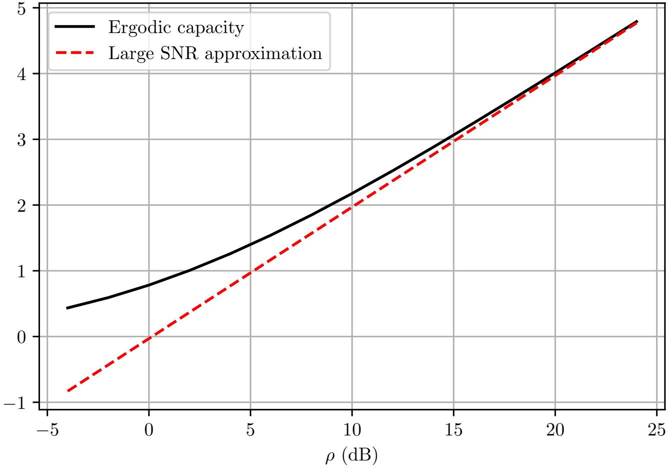


Figure: Ergodic capacity and large SNR approximation ($M = N = 2$)

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)

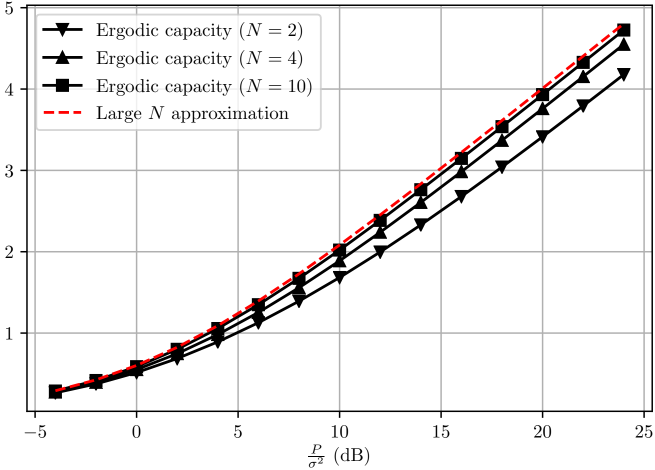


Figure: Ergodic capacity and large N approximation ($M = 2$) vs. total SNR $\frac{P}{\sigma^2}$

Large M analysis (1)

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

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

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)

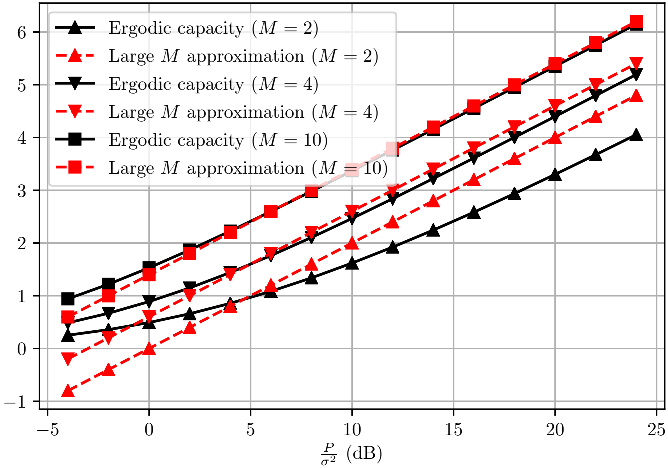


Figure: Ergodic capacity and large M approximation ($N = 2$) vs. total SNR $\frac{P}{\sigma^2}$

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

$$P_{\text{out}}(R) = \inf_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \text{tr}\mathbf{Q} \leq P}} \mathbb{P}\left(\operatorname{logdet}\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^*}{\sigma^2}\right) < R\right).$$

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

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

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]

Outline

Model (1)

- **Scenario.** We consider a $M \times N$ flat fading MIMO channel, constant over L symbols (block fading)

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}$$

where

- $\mathbf{H}$ is a $M \times N$ with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries, known at the receiver.
- $\mathbf{V}$ is a $M \times L$ with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ entries, with σ^2 known at the receiver.
- $\mathbf{X} \in \mathcal{C}$ is a matrix of transmit symbols, where $\mathcal{C} \subset \mathcal{A}^{N \times L}$ is a STBC and $\mathcal{A}$ is the symbol alphabet

Model (2)

- **Decoding.** We recall that a decoder is a mapping $\phi : \mathbb{C}^{M \times L} \mapsto \mathcal{C}$, and we denote

$$\hat{\mathbf{X}} = \phi(\mathbf{Y}).$$

- **Error probability.** If $\mathbf{X}$ is randomly chosen in $\mathcal{C}$, the error probability associated with decoder ϕ is given by

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

Maximum Likelihood (1)

- **Principle.** We estimate $\mathbf{X}$ as

$$\hat{\mathbf{X}} \in \underset{\mathbf{C}\in\mathcal{C}}{\operatorname{argmax}}\, f\left(\mathbf{Y};\mathbf{H},\mathbf{C},\sigma^2\right)$$

where $f\left(\mathbf{Y};\mathbf{H},\mathbf{C},\sigma^2\right)$ is the likelihood function given by

$$f\left(\mathbf{Y};\mathbf{H},\mathbf{C},\sigma^2\right) = \left(\frac{1}{\sigma^2\pi}\right)^{ML} \exp\left(-\frac{1}{\sigma^2}\left\|\mathbf{Y}-\mathbf{HC}\right\|_F^2\right).$$

with $\|.\|_F$ the Frobenius (or Hilbert-Schmidt) norm [2]

[2]For any matrix $M \times L$ matrix $\mathbf{A}$,

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{M}\sum_{j=1}^{L}|a_{i,j}|^2 = \mathrm{tr}(\mathbf{A}^*\mathbf{A}) = \sum_{i=1}^{\min(M,L)} \sigma_i(\mathbf{A})^2.$$

where $\sigma_1(\mathbf{A}) \geq \ldots \geq \sigma_{\min(M,L)}(\mathbf{A}) \geq 0$ are the singular values of $\mathbf{A}$.

Maximum Likelihood (2)

ML decoder

The ML decoder is given by

$$\hat{\mathbf{X}} \in \operatorname*{argmin}_{\mathbf{C} \in \mathcal{C}} \|\mathbf{Y} - \mathbf{H}\mathbf{C}\|_F^2$$

- **Remark (1).** The ML reduces to a least squares optimization problem.
- **Remark (2).** The ML decoder minimizes the error probability among all possible decoders, when the codeword $\mathbf{X}$ is uniformly distributed in $\mathcal{C}$ (cf. 2nd year Channel Coding course).
- **Remark (3).** High computational complexity $\mathcal{O}(|\mathcal{C}|)$ due to the exhaustive search, which may be of the order $\mathcal{O}\left(|\mathcal{A}|^{NL}\right)$ in the worst case scenario where $\mathcal{C} = \mathcal{A}^{N \times L}$.

ZF decoder (1)

- **Principle.** Design a $M \times N$ filtering matrix $\mathbf{F}$ to minimize the ISI due to the spatial mixing induced by the MIMO channel:

$$\begin{aligned}\mathbf{Z} &= \mathbf{F}^*\mathbf{Y} \\ &= \underbrace{\mathbf{F}^*\mathbf{H}}_{\approx \mathbf{I}}\mathbf{X} + \mathbf{F}^*\mathbf{V}\end{aligned}$$

- **Design (1).** The Zero-Forcing (ZF) filter is chosen as the (unique) minimum norm matrix among the set of solutions of the least squares problem

$$\operatorname*{argmin}_{\mathbf{F}\in\mathbb{C}^{M\times N}} \|\mathbf{F}^*\mathbf{H} - \mathbf{I}\|_F^2$$

ZF decoder (2)

- **Design (2).** The solution is given by

$$\mathbf{F}_{ZF} = (\mathbf{H}^*)^+ = (\mathbf{H}^+)^*$$

where $\mathbf{H}^+$ is the pseudo-inverse [3] of $\mathbf{H}$.

- **Remark.** If $\mathrm{rank}(\mathbf{H}) = N$, then $\mathbf{F}_{ZF} = \mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}$ in which case the filtered observed signal writes

$$\mathbf{Z} = \mathbf{X} + \mathbf{F}_{ZF}^* \mathbf{V}.$$

The spatial interference is totally cancelled.

[3] If $r = \mathrm{rank}(\mathbf{H})$ and if a SVD of $\mathbf{H}$ is given by $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ where $\mathbf{U}$ and $\mathbf{V}$ are resp. $M \times r$ and $N \times r$ isometric matrices and $\boldsymbol{\Sigma}$ is a $r \times r$ positive diagonal matrix, then

$$\mathbf{H}^+ = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^*.$$

ZF decoder (3)

ZF decoder

The ZF decoder is given by $\hat{\mathbf{X}} = (\hat{x}_{n,\ell})_{\substack{n=1,\ldots,N \\ \ell=1,\ldots,L}}$ with

$$\hat{x}_{n,\ell} \in \underset{s \in \mathcal{A}}{\operatorname{argmin}} \left| z_{n,\ell} - s \right|^2 ,$$

with $\mathbf{Z} = (z_{n,\ell})_{\substack{n=1,\ldots,N \\ \ell=1,\ldots,L}} = \mathbf{H}^{+}\mathbf{Y}$.

ZF decoder (4)

- **Remark (1).** The decoded matrix $\hat{\mathbf{X}}$ does not belong to $\mathcal{C}$ in general, and additionnal operations may be needed to "project" $\hat{\mathbf{X}}$ onto $\mathcal{C}$.

- **Remark (2).** The ZF decoder is suboptimal (in terms of error probability) but has reduced complexity compared to the ML decoder: in the worst case scenario where $\mathcal{C} = \mathcal{A}^{N\times L}$, we have $\mathcal{O}\left(NL|\mathcal{A}|\right)$.

ZF decoder (5)

- **Remark (3).** The filtered noise $\mathbf{F}_{ZF}^*\mathbf{V}$ is no longer spatially white, with energy

$$\begin{aligned}\frac{1}{L}\mathbb{E}[\|\mathbf{F}_{ZF}^*\mathbf{V}\|_F^2] &= \sigma^2 \|\mathbf{F}_{ZF}\|_F^2 \\ &= \sigma^2 \mathrm{tr}\left(\mathbf{H}^+(\mathbf{H}^+)^*\right) \\ &= \sigma^2 \sum_{i=1}^{\mathrm{rank}(\mathbf{H})} \frac{1}{\sigma_i(\mathbf{H})^2}\end{aligned}$$

where $\sigma_1(\mathbf{H}) \geq \ldots \geq \sigma_{\mathrm{rank}(\mathbf{H})}(\mathbf{H})$ are the singular values of $\mathbf{H}$.

The noise energy is amplified if $\mathbf{H}$ has singular values close to 0.

ZF decoder (6)

- **Remark (4).** After filtering, we have

$$\mathbf{Z} = \mathbf{X} + \underbrace{\left(\mathbf{H}^{+}\mathbf{H} - \mathbf{I}\right)\mathbf{X}}_{ISI} + \mathbf{H}^{+}\mathbf{V},$$

and the ISI energy is given by

$$\frac{1}{L}\mathbb{E}\left\|\left(\mathbf{H}^{+}\mathbf{H} - \mathbf{I}\right)\mathbf{X}\right\|_F^2 = \mathrm{tr}\left(\mathbf{\Pi}\mathbb{E}[\mathbf{X}\mathbf{X}^*]\right)$$

$$= N - r \qquad (\text{if } \mathbb{E}[\mathbf{X}\mathbf{X}^*] = \mathbf{I}),$$

where $\mathbf{\Pi} = \mathbf{I} - \mathbf{H}^{+}\mathbf{H}$ is the orthogonal projection matrix onto $\mathrm{Ker}(\mathbf{H})$

MMSE decoder (1)

- **Principle.** On one symbol time, we observe

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

where $\mathbf{x} \in \mathbb{C}^N$ is the transmit symbol vector and $\mathbf{v} \sim \mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \sigma^2 \mathbf{I})$, and the goal is to design a $M \times N$ filtering matrix $\mathbf{F}$ minimizing the MSE between $\mathbf{F}^* \mathbf{y}$ and $\mathbf{x}$:

$$\mathbf{F}_{MMSE} \in \underset{\mathbf{F} \in \mathbb{C}^{M \times N}}{\operatorname{argmin}} \mathbb{E} \left\| \mathbf{F}^* \mathbf{y} - \mathbf{x} \right\|_2^2 .$$

- **Design.** The MMSE filter is unique and given by

$$\mathbf{F}_{MMSE} = \left(\mathbf{H}\mathbf{Q}\mathbf{H}^* + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{H}\mathbf{Q},$$

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

MMSE decoder (2)

MMSE decoder

The MMSE decoder is given by $\hat{\mathbf{X}} = (\hat{x}_{n,\ell})_{\substack{n=1,\ldots,N \\ \ell=1,\ldots,L}}$ with

$$\hat{x}_{n,\ell} \in \underset{s \in \mathcal{A}}{\operatorname{argmin}} \left| z_{n,\ell} - s \right|^2 ,$$

with $\mathbf{Z} = (z_{n,\ell})_{\substack{n=1,\ldots,N \\ \ell=1,\ldots,L}} = \mathbf{Q}^* \mathbf{H}^* \left(\mathbf{H}\mathbf{Q}\mathbf{H}^* + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{Y}$.

MMSE decoder (3)

- **Remark (1).** When $\mathbf{Q} = \mathbf{I}$, we have

$$\begin{aligned}\mathbf{F}_{MMSE} &= \left(\mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}\right)^{-1}\mathbf{H} \\ &= \mathbf{H}\left(\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I}\right)^{-1} \\ &\xrightarrow[\sigma^2\to 0]{} \mathbf{F}_{ZF}.\end{aligned}$$

MMSE decoder (4)

- **Remark (2).** If $\mathbf{Q} = \mathbf{I}$, the filtered noise energy is given by

$$\frac{1}{L}\mathbb{E}\left\|\mathbf{F}_{MMSE}^{*}\mathbf{V}\right\|_F^2 = \sigma^2\left\|\mathbf{F}_{MMSE}\right\|_F^2$$

$$= \sigma^2\sum_{i=1}^{r}\left(\frac{\sigma_i(\mathbf{H})}{\sigma_i(\mathbf{H})^2+\sigma^2}\right)^2$$

The noise energy cannot be unbounded for a fixed value of σ^2 if e.g. $\sigma_r(\mathbf{H}) \to 0$.

The SIC decoder (1)

- **Principle (1).** The idea is to perform Successive Interference Cancellation (SIC) technique based on a QR decomposition of the channel matrix $\mathbf{H}$: if we assume $M \geq N$, then

$$\mathbf{H} = \mathbf{QR}$$

where $\mathbf{Q}$ is a $M \times N$ isometric matrix, and $\mathbf{R}$ is a $N \times N$ upper triangular matrix.

The SIC decoder (2)

- **Principle (2).** We compute

$$
\begin{aligned}
\mathbf{Z} &= \mathbf{Q}^*\mathbf{Y} \\
&= \mathbf{R}\mathbf{X} + \mathbf{Q}^*\mathbf{V} \\
&= \begin{pmatrix} r_{1,1} & \dots & r_{1,N} \\ & \ddots & \vdots \\ & & r_{N,N} \end{pmatrix} \begin{pmatrix} x_{1,1} & \dots & x_{1,L} \\ \vdots & & \vdots \\ x_{N,1} & \dots & x_{N,L} \end{pmatrix} + \mathbf{W}
\end{aligned}
$$

where $\mathbf{W} = \mathbf{Q}^*\mathbf{V}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ entries (i.e. the transformed noise remains temporally and spatially white).

The SIC decoder (3)

SIC decoder

The SIC decoder is given by $\hat{\mathbf{X}} = (\hat{x}_{n,\ell})_{\substack{n=1,\dots,N \\ \ell=1,\dots,L}}$ with

$$\hat{x}_{N,\ell} \in \operatorname*{argmin}_{s \in \mathcal{A}} \left| z_{N,\ell} - r_{N,N} s \right|^2 \quad \text{for } \ell = 1, \dots, L$$

and for all $n = N-1, \dots, 1$ and $\ell = 1, \dots, L$

$$\hat{x}_{n,\ell} \in \operatorname*{argmin}_{s \in \mathcal{A}} \left| z_{n,\ell} - \sum_{k=n+1}^{N} r_{n,k} \hat{x}_{k,\ell} - r_{n,n} s \right|^2,$$

with $\mathbf{Z} = (z_{n,\ell})_{\substack{n=1,\dots,N \\ \ell=1,\dots,L}} = \mathbf{Q}^* \mathbf{Y}$.

The SIC decoder (4)

- **Remark (1).** In general, the SIC decoder outperforms the ZF and MMSE decoder in terms of error probability, in the high SNR regime.
- **Remark (2).** The SIC performance may be degraded due to error propagation.
- **Remark (3).** The SIC idea may also be used in conjunction with ZF and MMSE decoders, giving new decoders SIC-ZF and SIC-MMSE.

Performance of standard decoders (1)

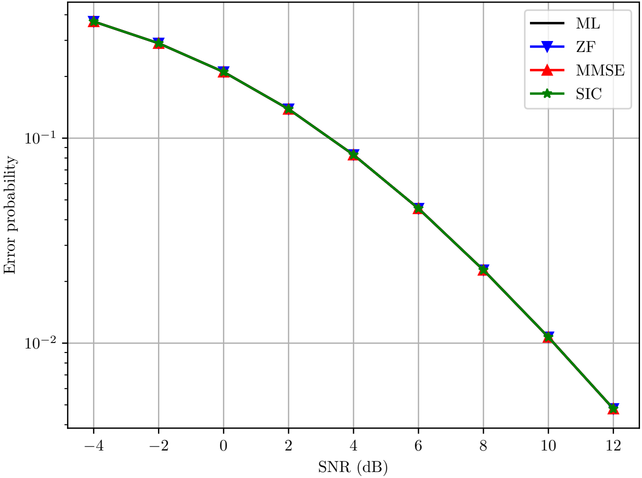


Figure: Error probability vs. SNR ($M = 2$, $N = 1$, $L = 1$, $\mathcal{C} = \mathcal{A}^{N \times L}$, $\mathcal{A} =$ QPSK)

Performance of standard decoders (2)

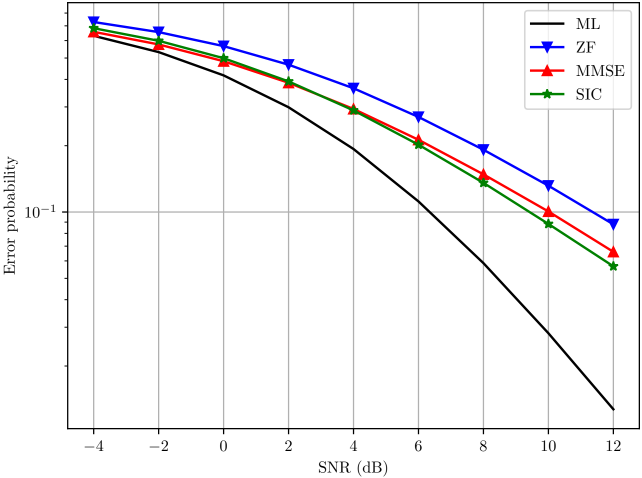


Figure: Error probability vs. SNR ($M = N = 2$, $L = 1$, $\mathcal{C} = \mathcal{A}^{N \times L}$, $\mathcal{A} =$ QPSK)

Performance of standard decoders (3)

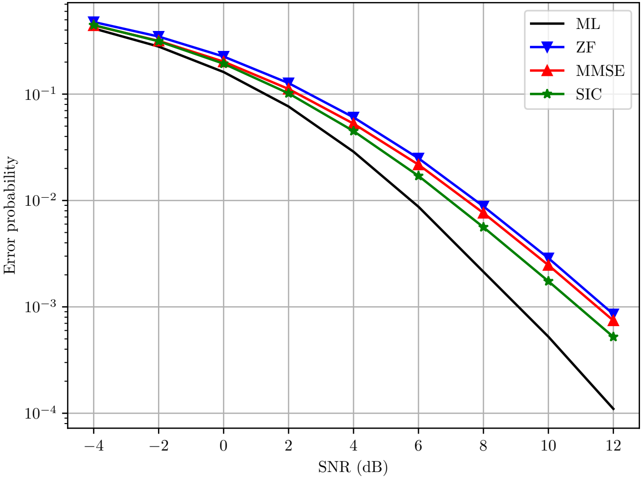


Figure: Error probability vs. SNR ($M = 4$, $N = 2$, $L = 1$, $\mathcal{C} = \mathcal{A}^{N\times L}$, $\mathcal{A} =$ QPSK)

Performance of standard decoders (4)

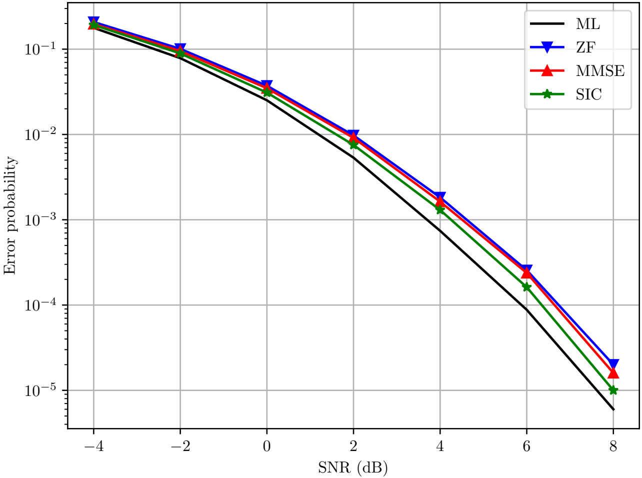


Figure: Error probability vs. SNR ($M = 8$, $N = 2$, $L = 1$, $\mathcal{C} = \mathcal{A}^{N \times L}$, $\mathcal{A} =$ QPSK)

Performance of standard decoders (5)

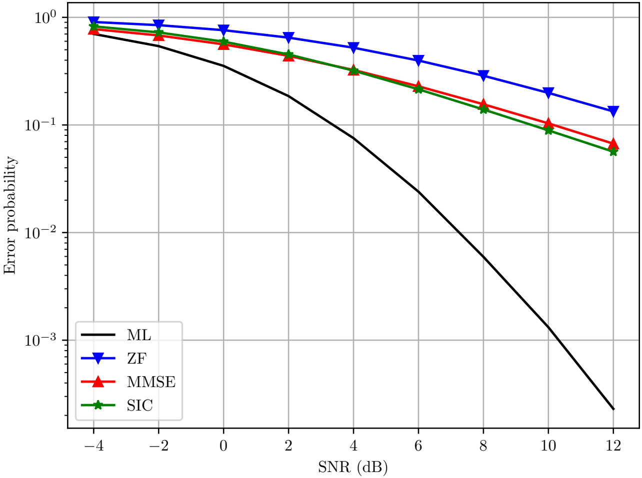


Figure: Error probability vs. SNR ($M = 4$, $N = 4$, $L = 1$, $\mathcal{C} = \mathcal{A}^{N \times L}$, $\mathcal{A} =$ QPSK)

PEP (1)

- **Error probability (1).** When the random codeword $\mathbf{X}$ is uniformly distributed in the code $\mathcal{C}$, we have

$$P_e = \mathbb{P}\left(\hat{\mathbf{X}} \neq \mathbf{X}\right)$$

$$= \frac{1}{|\mathcal{C}|} \sum_{\mathbf{C} \in \mathcal{C}} \mathbb{P}\left(\hat{\mathbf{X}} \neq \mathbf{X} | \mathbf{X} = \mathbf{C}\right).$$

$$= \frac{1}{|\mathcal{C}|} \sum_{\mathbf{C} \in \mathcal{C}} \sum_{\substack{\mathbf{C}' \in \mathcal{C} \\ \mathbf{C}' \neq \mathbf{C}}} \mathbb{P}\left(\hat{\mathbf{X}} = \mathbf{C}' | \mathbf{X} = \mathbf{C}\right).$$

PEP (2)

- **Error probability (2).** Using the ML decoder (which minimizes P_e), we get

$$P_e = \frac{1}{|\mathcal{C}|} \sum_{\mathbf{C} \in \mathcal{C}} \sum_{\substack{\mathbf{C}' \in \mathcal{C} \\ \mathbf{C}' \neq \mathbf{C}}} \mathbb{P}\left(\hat{\mathbf{X}} = \mathbf{C}' | \mathbf{X} = \mathbf{C}\right)$$

$$= \frac{1}{|\mathcal{C}|} \sum_{\mathbf{C} \in \mathcal{C}} \sum_{\substack{\mathbf{C}' \in \mathcal{C} \\ \mathbf{C}' \neq \mathbf{C}}} \mathbb{P}\left(\bigcap_{\mathbf{C}'' \in \mathcal{C}} \left\{\|\mathbf{Y} - \mathbf{H}\mathbf{C}'\|_F^2 \leq \|\mathbf{Y} - \mathbf{H}\mathbf{C}''\|_F^2\right\} | \{\mathbf{X} = \mathbf{C}\}\right)$$

$$\leq \frac{1}{|\mathcal{C}|} \sum_{\mathbf{C} \in \mathcal{C}} \sum_{\substack{\mathbf{C}' \in \mathcal{C} \\ \mathbf{C}' \neq \mathbf{C}}} \underbrace{\mathbb{P}\left(\|\mathbf{Y} - \mathbf{H}\mathbf{C}'\|_F^2 \leq \|\mathbf{Y} - \mathbf{H}\mathbf{C}\|_F^2 \ | \ \mathbf{X} = \mathbf{C}\right)}_{:= P_e(\mathbf{C} \to \mathbf{C}')}.$$

PEP (3)

- **PEP.** The Pairwise Error Probability (PEP) is defined as

$$P_e\left(\mathbf{C} \to \mathbf{C}'\right) = \mathbb{P}\left(\|\mathbf{Y} - \mathbf{H}\mathbf{C}'\|_F^2 \leq \|\mathbf{Y} - \mathbf{H}\mathbf{C}\|_F^2 \;\mid \mathbf{X} = \mathbf{C}\right)$$

- **Remark.** The PEP represents the probability of decoding $\mathbf{C}'$ knowing that $\mathbf{C}$ is the transmit codeword, and assuming that $\mathbf{C}$ and $\mathbf{C}'$ are the only codewords in the code.

Tarokh's criteria (1)

A bound on the PEP

We have

$$P_e\left(\mathbf{C} \to \mathbf{C}'\right) \leq \prod_{j=1}^{N} \left(\frac{1}{1+\frac{\lambda_j}{4\sigma^2}}\right)^M$$

where $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$ denote the eigenvalues of matrix $(\mathbf{C}-\mathbf{C}')(\mathbf{C}-\mathbf{C}')^*$.

- **Remark.** The eigenvalues $\lambda_1, \ldots, \lambda_N$ intuitively measures the distance between codewords $\mathbf{C}$ and $\mathbf{C}'$, so that if $\lambda_1 \to \infty$, we have

$$P_e\left(\mathbf{C} \to \mathbf{C}'\right) \to 0.$$

Tarokh's criteria (2)

- **Large SNR behaviour.** Letting $\rho = \frac{1}{\sigma^2}$ and $r = \text{rank}(\mathbf{C} - \mathbf{C}')$, we have

$$P_e(\mathbf{C} \to \mathbf{C}') \leq \left(\frac{4}{\rho}\right)^{Mr} \left(\prod_{j=1}^{r} \frac{1}{\lambda_j}\right)^{M} + \mathcal{O}\left(\frac{1}{\rho^{Mr+1}}\right)$$

as $\rho \to \infty$.

The STBC can be designed in terms of parameters r and $\lambda_1, \ldots, \lambda_r$ to minimize the bound on the PEP for large SNR.

Tarokh's criteria (3)

The Rank Criterion

The STBC should be designed to *maximize* the *minimum rank of the code*, i.e. choose the code $\mathcal{C}$ such that

$$r_{min}(\mathcal{C}) = \min_{\substack{\mathbf{C},\mathbf{C}' \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{C}'}} \text{rank}\left(\mathbf{C} - \mathbf{C}'\right)$$

is maximal.

- **Remark (1).** The quantity $d(\mathcal{C}) = M \times r_{min}(\mathcal{C})$ is called the *diversity advantage* of the code $\mathcal{C}$, and corresponds to the exponent of the factor $\frac{1}{\rho}$ in the error probability.
- **Remark (2).** The maximum diversity advantage is $d(\mathcal{C}) = MN$ (achieved when $r_{min}(\mathcal{C}) = N$), and coincides with the number of independent paths in the $M \times N$ MIMO channel.

Tarokh's criteria (4)

The Determinant Criterion

The STBC should be designed to *maximize*

$$\kappa_{min}(\mathcal{C}) = \min_{\substack{\mathbf{C},\mathbf{C}'\in\mathcal{C}\\ \mathbf{C}\neq\mathbf{C}'}} \prod_{j=1}^{\text{rank}(\mathbf{C}-\mathbf{C}')} \lambda_j.$$

- **Remark (1).** The quantity $\kappa_{min}(\mathcal{C})$ is called the *coding advantage* of the code $\mathcal{C}$, and acts as a secondary criterion to compare two STBC having the same diversity advantage.

- **Remark (2).** When $r_{min}(\mathcal{C}) = N$, then

$$\kappa_{min}(\mathcal{C}) = \max_{\substack{\mathbf{C},\mathbf{C}'\in\mathcal{C}\\ \mathbf{C}\neq\mathbf{C}'}} \det\left((\mathbf{C}-\mathbf{C}')(\mathbf{C}-\mathbf{C}')^*\right).$$

Tarokh's criteria (5)

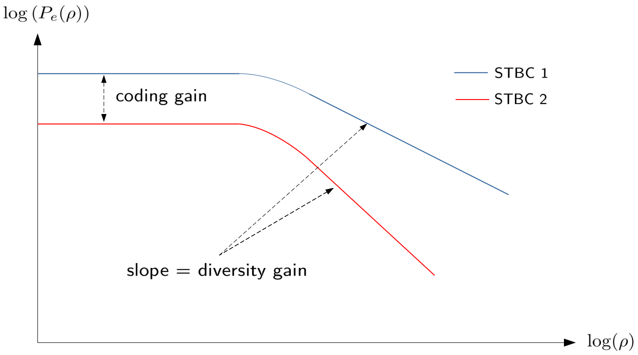


Figure: Illustration of the diversity and coding advantages.

STBC

- **Principles.** Space-Time Block Coding consists in organizing the sequence of symbols
  - across the transmit antennas (space)
  - across channel uses (time)

$$\mathbf{C} = \begin{array}{c} \\ N \text{ antennas} \updownarrow \\ \\ \end{array} \underbrace{\begin{pmatrix} c_{1,1} & \cdots & c_{1,L} \\ \vdots & & \vdots \\ c_{N,1} & \cdots & c_{N,L} \end{pmatrix}}_{\longleftrightarrow \ L \text{ channel uses}}$$

- **Challenge.** Compromise between the error probability (diversity advantage) and the rate (number of effective transmit symbol per codeword).

Layered codes: the V-BLAST (1)

- **V**ertical **B**ell **LA**bs **S**pace-**T**ime.
- **Idea.** Transmit N streams of symbols $(s_\ell^{(0)})_{\ell\in\mathbb{Z}}, \ldots, (s_\ell^{(N-1)})_{\ell\in\mathbb{Z}}$, with each stream attached to one antenna, so that
- **Codeword.**

$$\mathbf{C} = \begin{pmatrix} s_0^{(0)} & s_1^{(0)} & \ldots & s_{L-1}^{(0)} \\ \vdots & \vdots & & \vdots \\ s_0^{(N-1)} & s_1^{(N-1)} & \ldots & s_{L-1}^{(N-1)} \end{pmatrix}$$

Layered codes: the V-BLAST (2)

- **Code size.** The code is given by

$$\mathcal{C} = \mathcal{S}^N$$

where $\mathcal{S}$ is the set of substreams ($\mathcal{S} \subset \mathcal{A}^L$).

- **Rate.**

$$R(\mathcal{C}) = \frac{\log|\mathcal{C}|}{L} = \frac{N}{L}\log|\mathcal{S}|.$$

If $\mathcal{S} = \mathcal{A}^L$, $R(\mathcal{C}) = N$ is the maximal rate achievable by V-BLAST.

Layered codes: the V-BLAST (3)

- **Diversity advantage.**

$$r_{min} = \min_{\substack{\mathbf{C},\mathbf{C}' \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{C}'}} \mathsf{rank}\left(\mathbf{C} - \mathbf{C}'\right) = 1.$$

V-BLAST offers the minimal diversity advantage $d(\mathcal{C}) = M$, whatever the choice of substream set $\mathcal{S}$.

- **Remark.** V-BLAST offers a poor diversity advantage due to the absence of coding *across* antennas, since each substream is associated with a fixed antenna.

Layered codes: the D-BLAST (1)

- **D**iagonal **B**ell **LA**bs **S**pace-**T**ime.

- **Idea.** Transmit K streams of symbols $(s_\ell^{(0)})_{\ell\in\mathbb{Z}}, \dots, (s_\ell^{(K-1)})_{\ell\in\mathbb{Z}}$, with each stream shifting "circularly" across the antennas, so that

- **Codeword.**

$$\mathbf{C} = \underbrace{\begin{pmatrix} s_0^{(0)} & \dots & \dots & s_0^{(K-1)} & 0 & \dots & \dots & 0 \\ 0 & s_1^{(0)} & \dots & \dots & s_1^{(K-1)} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & & \ddots & 0 \\ 0 & \dots & \dots & 0 & s_{N-1}^{(0)} & \dots & \dots & s_{N-1}^{(K-1)} \end{pmatrix}}_{L = N+K-1}$$

Layered codes: the D-BLAST (2)

- **Code size.** The code is given by

$$\mathcal{C} = \mathcal{S}^K$$

where $\mathcal{S}$ is the set of substreams ($\mathcal{S} \subset \mathcal{A}^N$).

- **Rate.**

$$R(\mathcal{C}) = \frac{\log|\mathcal{C}|}{L} = \frac{K}{N+K-1} \log|\mathcal{S}|.$$

If $\mathcal{S} = \mathcal{A}^N$, $R(\mathcal{C}) = \frac{KN}{N+K-1}$ is the maximum rate achievable by D-BLAST.

Layered codes: the D-BLAST (3)

- **Diversity advantage.**

$$r_{min} = \min_{\substack{\mathbf{C},\mathbf{C}' \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{C}'}} \operatorname{rank}\left(\mathbf{C} - \mathbf{C}'\right) = \min_{\substack{\mathbf{s},\mathbf{s}' \in \mathcal{S} \\ \mathbf{s} \neq \mathbf{s}'}} \left\|\mathbf{s} - \mathbf{s}'\right\|_0 ,$$

where $\|.\|_0$ is the ℓ^0-norm (number of non-zero entries).

D-BLAST may achieve a diversity advantage $d(\mathcal{C})$ up to N, depending on the choice of the substream set $\mathcal{S}$.

- **Remark.** With D-BLAST, each substream is sent across the N antennas, allowing to achieve a better diversity advantage than V-BLAST, at the cost of a lower rate.

Orthogonal codes (1)

- **Principles.** Orthogonal codes are STBC such that every codeword matrix $\mathbf{C}$ can be represented as

$$\mathbf{C} = \sum_{k=1}^{K} \left(s_k \boldsymbol{\Phi}_k + \overline{s}_k \boldsymbol{\Psi}_k \right)$$

with $s_1, \ldots, s_K \in \mathcal{A}$, and where $\boldsymbol{\Phi}_1, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Phi}_K, \boldsymbol{\Psi}_K$ are $N \times L$ matrices with entries in the set $\{-1, 0, 1\}$ chosen so that

$$\mathbf{C}\mathbf{C}^* = \alpha \left(\sum_{k=1}^{K} |s_k|^2 \right) \mathbf{I}$$

for some $\alpha > 0$.

Orthogonal codes (2)

- **Decoding.** Orthogonal codes can be decoded via ML with reduced computational cost: the minimization problem

$$\hat{\mathbf{C}} = \underset{\mathbf{C}\in\mathcal{C}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{H}\mathbf{C}\|_F^2$$

is equivalent to estimate $s_1, \ldots, s_K$ as

$$\hat{s}_k = \underset{s\in\mathcal{A}}{\operatorname{argmin}} \left|z_k - \alpha\|\mathbf{H}\|_F^2 s\right|^2,$$

where $z_k = \overline{\operatorname{tr}\left(\mathbf{Y}^*\mathbf{H}\mathbf{\Phi}_k\right)} + \operatorname{tr}\left(\mathbf{Y}^*\mathbf{H}\mathbf{\Psi}_k\right)$.

Using orthogonal codes breaks the exponential computational cost of ML.

Orthogonal codes (3)

- **Example 1.** For the case where $N = L = 2$, orthogonal codes are called *Alamouti* codes, and codewords are represented (up to some permutation) by matrices

$$\mathbf{C} = \begin{pmatrix} s_1 & s_2 \\ -\overline{s_2} & \overline{s_1} \end{pmatrix}$$

In that case, $K = 2$ and

$$\mathbf{\Phi}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \mathbf{\Psi}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$
$$\mathbf{\Phi}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \mathbf{\Psi}_2 = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}$$

- **Remark.** The Alamouti code achieves a maximum diversity advantage of $2M$ and a rate of 1 symbol per channel use.

Orthogonal codes (4)

- **Example 2.** For $N = 3, L = 4$, the following codewords

$$\mathbf{C} = \begin{pmatrix} s_1 & -\overline{s_2} & \overline{s_3} & 0 \\ s_2 & \overline{s_1} & 0 & \overline{s_3} \\ s_3 & 0 & -\overline{s_1} & -\overline{s_2} \end{pmatrix}$$

for $s_1, s_2, s_3 \in \mathcal{A}$ define an orthogonal code, with basis matrices

$$\mathbf{\Phi}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \; \mathbf{\Phi}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \; \mathbf{\Phi}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{\Psi}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix} \; \mathbf{\Psi}_2 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \; \mathbf{\Psi}_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This code has a diversity advantage of $3M$ and a rate $\frac{3}{4}$.

A new look at diversity (1)

- **Reminder.** Let us recall that the ergodic capacity, given by

$$C(\rho) = \mathbb{E}\left[\operatorname{logdet}\left(\mathbf{I} + \rho\mathbf{H}\mathbf{H}^*\right)\right]$$

behaves at large SNR ρ as

$$C(\rho) \underset{\rho\to+\infty}{\sim} \min(M,N)\log(\rho),$$

that is, as the capacity of $\min(M,N)$ independent SISO Gaussian channels.

To take fully benefits from this behaviour, it is relevant to consider STBC parametrized by SNR ρ with rates increasing logarithmically with ρ.

A new look at diversity (2)

Diversity and multiplexing gains

Consider a family of STBC $(\mathcal{C}(\rho))_{\rho>0}$ with rate $R(\rho)$, error probability $P_e(\rho)$ and length L fixed w.r.t. ρ. We say that this family achieves

- a multiplexing gain of $r > 0$ if

$$\lim_{\rho\to\infty} \frac{R(\rho)}{\log(\rho)} = r,$$

- a diversity gain of $d(r) > 0$ if

$$\lim_{\rho\to\infty} \frac{\log\left(P_e(\rho)\right)}{\log(\rho)} = -d(r).$$

A new look at diversity (3)

- **Remark 1.** The notion of diversity defined here is related to the actual error probability P_e, while the one defined previously was related to the PEP.

- **Remark 2.** The case $r = 0$ includes the situation where the code rate R is kept fixed as SNR $\rho \to \infty$, and in that case, $d(0)$ coincides with the diversity gain on the PEP.

- **Remark 3.** Intuitively, in the case where the multiplexing gain $r = \min(M, N)$, the code rate R is close to the ergodic capacity and one expects a diversity gain $d(\min(M, N)) = 0$.

Diversity-Multiplexing Trade-off (1)

Zheng & Tse Theorem [17]

Let $d^*(r)$ denotes the supremum of the diversity gains achieved over all family of STBC $(\mathcal{C}(\rho))_{\rho>0}$ having multiplexing gain r. Then if $L \geq M + N - 1$, the function

$$r \mapsto d^*(r)$$

is piecewise linear between the points $k \in \{0, 1, \ldots, \min(M, N)\}$, with

$$d^*(k) = (M - k)(N - k).$$

Diversity-Multiplexing Trade-off (2)

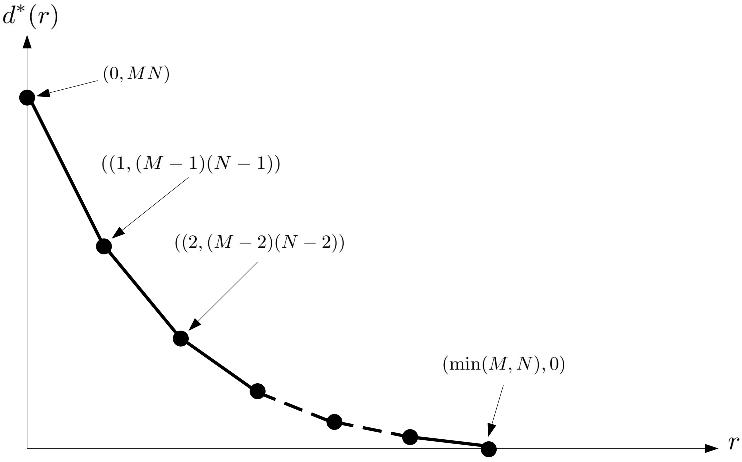


Figure: Illustration of DMT

Diversity-Multiplexing Trade-off (3)

- **Remark 1.** Increasing the STBC length L above $M + N - 1$ to increase diversity is useless; a length $L = M + N - 1$ is sufficient to extract the maximal diversity.

- **Remark 2.** DMT can be seen as an extension of Shannon's theorem in a different asymptotic regime ($\rho \to \infty$ instead of $L \to \infty$)

- **Remark 3.** In this framework, the Rank and Determinant criteria are no longer relevant to design efficient STBC.

DMT for standard codes (1)

- **Symbol constellation.** To ensure a rate $R = R(\rho) \sim r\log(\rho)$ as $\rho \to \infty$, we consider a QAM constellation $\mathcal{A}(\rho)$ such that

$$|\mathcal{A}(\rho)| \underset{\rho\to\infty}{\sim} \rho^{\frac{r}{N}}.$$

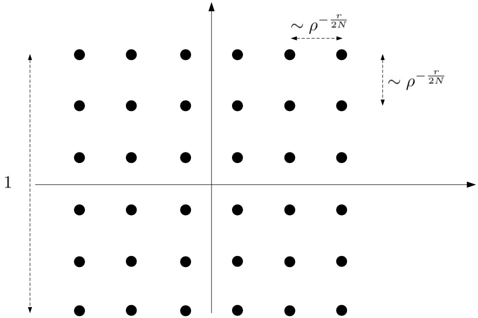


Figure: QAM constellation with size behaving as $\rho^{\frac{r}{N}}$ as $\rho \to \infty$

DMT for standard codes (2)

V-BLAST [11]

When $M \geq N$ and using ML decoder, the DMT for the V-BLAST code is given by

$$d_{\mathrm{V-BLAST}}(r) = M\left(1 - \frac{r}{N}\right)^{+}$$

- V-BLAST has a maximal diversity gain of M and maximal multiplexing gain of N.
- If $M = N$, V-BLAST achieves optimal DMT only for multiplexing gain $r = N - 1$.

DMT for standard codes (3)

D-BLAST [11]

For $M, N \geq 2$, $K = 2$ streams, and using ML decoder, the DMT for the D-BLAST code is given by

$$d_{\mathrm{D-BLAST}}(r) = M \left(N - \frac{N+1}{2} r \right)^+$$

- For $M = 2$, D-BLAST achieves the optimal DMT for $0 \leq r \leq 1$.

DMT for standard codes (4)

Alamouti [17]

Using ML decoder, the DMT for the Alamouti code is given by

$$d_{\text{ALAMOUTI}}(r) = 2M\left(1-r\right)^{+}.$$

- If $M = 1$, then

$$d_{\text{ALAMOUTI}}(r) = d^{*}(r)$$

  and the Alamouti code achieves the optimal DMT.

- If $M > 1$, then

$$d_{\text{ALAMOUTI}}(r) < d^{*}(r) \quad \forall\, 0 < r \leq 2.$$

  Alamouti code is suboptimal when the receiver has more than one antenna.

DMT for standard codes (4)

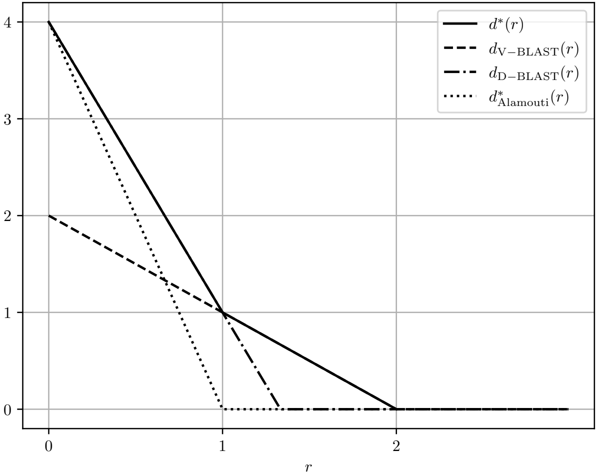


Figure: DMT for $M = N = 2$.

DMT for standard codes (5)

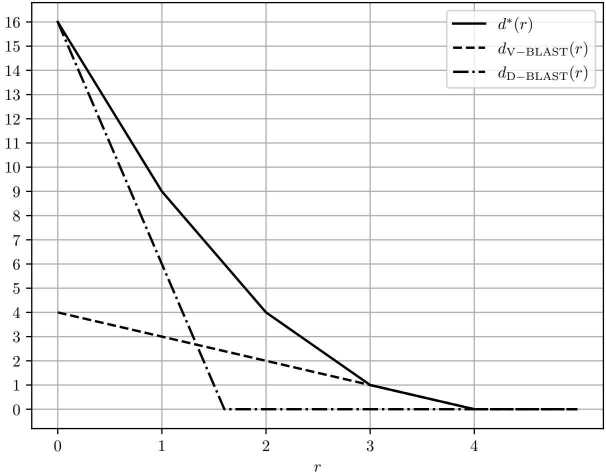


Figure: DMT for $M = N = 4$.

Further references on DMT

- **Optimal DMT achieving codes.** [1, 10, 4, 5]
- **DMT with linear decoders.** [9, 7, 3]
- **DMT for MAC.** [13]

Outline

References I

[1] J-C Belfiore, G. Rekaya, and E. Viterbo.
The golden code: a 2×2 full-rate space-time code with nonvanishing determinants.
IEEE Trans. Inf. Theory, 51(4):1432–1436, 2005.

[2] G. Caire and S. Shamai.
On the achievable throughput of a multiantenna gaussian broadcast channel.
IEEE Trans. Inf. Theory, 49(7):1691–1706, 2003.

[3] P. Dupuy, F.and Loubaton.
Diversity of the mmse receiver in flat fading and frequency selective mimo channels at fixed rate.
In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 932–936. IEEE, 2011.

[4] P. Elia, R. Kumar, S. Pawar, P. Kumar, and H. Lu.
Explicit space–time codes achieving the diversity–multiplexing gain tradeoff.
IEEE Trans. Inf. Theory, 52(9):3869–3884, 2006.

[5] P. Elia, B. Sethuraman, and P. Kumar.
Perfect space–time codes for any number of antennas.
IEEE Trans. Inf. Theory, 53(11):3853–3868, 2007.

References II

[6] G. Foschini and M. Gans.
On limits of wireless communications in a fading environment when using multiple antennas.
Wirel. Pers. Commun., 6(3):311–335, 1998.

[7] Y. Jiang, M. Varanasi, and J. Li.
Performance analysis of zf and mmse equalizers for mimo systems: An in-depth study of the high snr regime.
IEEE Trans. Inf. Theory, 57(4):2008–2026, 2011.

[8] E. Jorswieck and H. Boche.
Channel capacity and capacity-range of beamforming in mimo wireless systems under correlated fading with covariance feedback.
IEEE Trans. Wirel. Commun., 3(5):1543–1553, 2004.

[9] R. Kumar, G. Caire, and A. Moustakas.
Asymptotic performance of linear receivers in mimo fading channels.
IEEE Trans. Inf. Theory, 55(10):4398–4418, 2009.

References III

[10] F. Oggier, G. Rekaya, J-C Belfiore, and E. Viterbo.
Perfect space-time block codes.
IEEE Trans. Inf. Theory, 52(9):3885–3902, 2006.

[11] S. Tavildar and P. Viswanath.
Approximately universal codes over slow-fading channels.
IEEE Trans. Inf. Theory, 52(7):3233–3258, 2006.

[12] E. Telatar.
Capacity of Multi-antenna Gaussian Channels.
Eur. Trans. Telecommun., 10(6):585–595, 1999.

[13] D. Tse, P. Viswanath, and L. Zheng.
Diversity-multiplexing tradeoff in multiple-access channels.
IEEE Trans. Inf. Theory, 50(9):1859–1874, 2004.

[14] P. Viswanath and D. Tse.
Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality.
IEEE Trans. Inf. Theory, 49(8):1912–1921, 2003.

References IV

[15] P. Viswanath, D. Tse, and V. Anantharam.
Asymptotically optimal water-filling in vector multiple-access channels.
IEEE Trans. Inf. Theory, 47(1):241–267, 2001.

[16] L. Zheng and D. Tse.
Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel.
IEEE Trans. Inf. Theory, 48(2):359–383, 2002.

[17] L. Zheng and D. Tse.
Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels.
IEEE Trans. Inf. Theory, 49(5):1073–1096, 2003.