MIMO wireless communications

Guodong Sun

8, Dec 2021

Table of Contents

Table of Contents

Wireless Communication

Foundations:

- Electromagnetic (EM) wave propagation
- Frequency spectrum
- Information theory

Directions to improve

- More access points (APs)
- More frequency band
- Higher spectrum efficiency
  - Multiple antenna technique

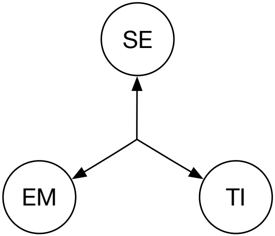
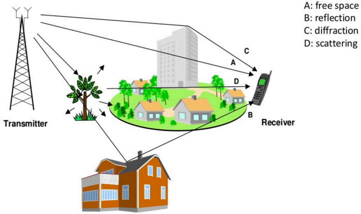


Figure: Communication

Multiple antenna technique



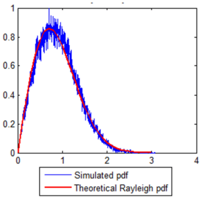


Figure: Multipath effect and the PDF of Rayleigh fading

Multipath fading

Multipath effect occurs in any practical environment because EM waves interact with all objects.

Motivation: using multiple antennas to exploit the **multipath** EM propagation.

- Diversity: multiple copy of one signal to avoid deep fading
- Beamforming: signals at particular angle experience constructive interference
- Multiplexing: multiple independent data streams

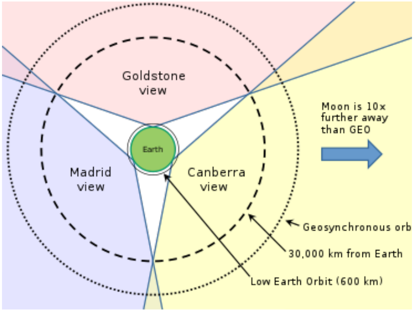
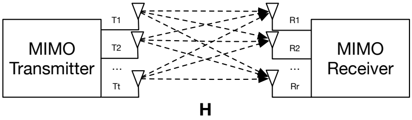


Figure: Diversity

Figure: Beamforming

Table of Contents

MIMO link model



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

- $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$
- $\mathbf{x} \in \mathbb{C}^{N_t}$
- $\mathbf{n}, \mathbf{y} \in \mathbb{C}^{N_r}$

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1t} \\ \vdots & \ddots & \vdots \\ h_{r1} & \cdots & h_{rt} \end{bmatrix}$$

MIMO channel of

- N_t transmit antennas (columns)
- N_r receive antennas (rows)

In particular, the channel can be

- Either deterministic, we use another notation $H \in \mathbb{C}^{N_r \times N_t}$
- Or stochastic, we use $\mathbf{H}$

Differential entropy of a random vector $\mathbf{x}$

Definition

The probability density of a circularly symmetric complex Gaussian with mean μ and covariance Q is given by

$$\gamma_{\mu,Q} = \det(\pi Q)^{-1} \exp\left(-(\mathbf{x}-\mu)^\dagger Q^{-1}(\mathbf{x}-\mu)\right)$$

Due to the invariance of translation, we consider the zero mean $\mu = 0$ case, and the entropy is given by

$$\begin{aligned}
\mathcal{H}(\gamma_{\mathcal{Q}}) &= \mathbb{E}[-\log \gamma_Q(\mathbf{x})] \\
&= \log\det(\pi Q) + \log(e)\mathbb{E}[\mathbf{x}^\dagger Q^{-1}\mathbf{x}] \\
&= \log\det(\pi Q) + \log(e)\mathrm{tr}(\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]Q^{-1}) \\
&= \log\det(\pi Q) + \log(e)\mathrm{tr}(I) \\
&= \log\det(\pi e Q)
\end{aligned} \tag{1}$$

Circularly symmetric complex Gaussian

Lemma

Circularly symmetric complex Gaussians are entropy maximizers

Proof.

Let p be any pdf satisfying $\int_{\mathbb{C}} p(x) x_i x_j * = Q_{ij}$, and recall that $\gamma_Q = \det(\pi Q)^{-1} \exp\left(-(\mathbf{x})^{\dagger} Q^{-1}(\mathbf{x})\right)$. So $\log(\gamma_{,Q})$ is a linear combination of the terms $x_i x_j^*$, $\mathbb{E}_{\gamma_Q}[\log \gamma_Q(x)] = \mathbb{E}_p[\log \gamma_Q(x)]$. Therefore

$$\begin{aligned}
\mathcal{H}(p) - \mathcal{H}(\gamma_Q) &= -\int_{\mathbb{C}} p(x) \log p(x) \mathrm{d}x + \int_{\mathbb{C}} \gamma_Q(x) \log \gamma_Q(x) \mathrm{d}x \\
&= -\int_{\mathbb{C}} p(x) \log p(x) \mathrm{d}x + \int_{\mathbb{C}} p(x) \log \gamma_Q(x) \mathrm{d}x \qquad (2) \\
&= \int_{\mathbb{C}} p(x) \log \frac{\gamma_Q(x)}{p(x)} \mathrm{d}x \leq 0.
\end{aligned}$$

since $\log(\cdot)$ is concave and apply the Jansen's equality. □

Mutual information [1]

$$\mathcal{I}(\mathbf{x};\mathbf{y}) = \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{y}|\mathbf{x}) = \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{n}), \tag{3}$$

Given the input variance Q and $\mathbf{y} = H\mathbf{x} + \mathbf{n}$, the output $\mathbf{y}$ has the variance $\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger] = HQH^\dagger + I_r$. Hence, maximal mutual entropy is realized when $\mathbf{x}, \mathbf{y}$ are circular Gaussian

$$\mathcal{I}(\mathbf{x};\mathbf{y}) = \log\det(I_r + HQH^\dagger) \tag{4}$$

The channel capacity is the maximum of $\mathcal{I}_Q(\mathbf{x};\mathbf{y})$, subjecting to the power constraint $\mathrm{tr}(Q) \leq P$.

Channel capacity: optimal condition

Diagonalizing $H^\dagger H = U^\dagger \Lambda U$, we have

$$\det(I_r + HQH^H) = \det\left(I_r + \Lambda^{1/2} UQU^\dagger (\Lambda^\dagger)^{1/2}\right) \tag{5}$$

Given that for any non-negative definite matrix A, $\det(A) \le \prod_i A_{ii}$, we have

$$\det(I_r + \Lambda^{1/2}\tilde{Q}(\Lambda^\dagger)^{1/2}) \le \prod_i (1 + \tilde{Q}_{ii}\lambda_i) \tag{6}$$

with equality when $\tilde{Q}$ is diagonal and the optimal entries can be found via water-filling

$$\tilde{Q}_{ii} = (\mu - \lambda_i^{-1})^+, \quad C(\mu) = \sum_i \left(\ln(\mu\lambda_i)\right)^+ \tag{7}$$

Channel capacity

Apply SVD $H = UDV$, where $U \in \mathbb{C}^{r \times r}, V \in \mathbb{C}^{t \times t}$ are unitary matrix and $D \in \mathbb{C}^{r \times t}$ is non-negative and diagonal.

$$\mathbf{y} = UDV\mathbf{x} + \mathbf{n} \to \tilde{\mathbf{y}} = \Sigma\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{8}$$

where $\tilde{\mathbf{y}} = U^H\mathbf{y}$, $\tilde{\mathbf{n}} = U^H\mathbf{n}$, and $\tilde{\mathbf{x}} = V^H\mathbf{x}$.

The component of D is denoted by $\lambda_i^{1/2}$, we can have orthogonal streams of signals

$$\tilde{y}_i = \lambda_i^{1/2}\tilde{x}_i + \tilde{n}_i, \quad \forall i \in [\min\{N_r, N_t\}]. \tag{9}$$

Channel capacity: optimal condition

Given

$$\tilde{y}_i = \lambda_i^{1/2} \tilde{x}_i + \tilde{n}_i,$$

the channel capacity is the sum of the orthogonal parallel channels, expressed by

$$C(H, P) = \sum_i \log(1 + \frac{P_i \lambda_i}{\sigma^2}). \tag{10}$$

$$\sum_i P_i = P \tag{11}$$

The channel capacity is the maximal mutual information, achieved by the water-filling algorithm

$$P(\mu) = \sum_i (\mu - \lambda_i^{-1})^+, \quad C(\mu) = \sum_i \Big(\ln(\mu \lambda_i) \Big)^+. \tag{12}$$

Example low-rank:
Let $H_{ij} = 1$ for all i, j. H is written

$$H = \begin{bmatrix} \sqrt{1/r} \\ \cdots \\ \sqrt{1/r} \end{bmatrix} (\sqrt{r}) \left[\sqrt{1/t} \cdots \sqrt{1/t} \right]. \tag{13}$$

Hence,

$$C = \log(1 + rP). \tag{14}$$

Example high-rank:
Take $r = t = n$ and $H = I_n$. Then

$$C = n \log(1 + P/n) \tag{15}$$

Gaussian channel with Rayleigh fading

Each entry of $\mathbf{H}$ has

- uniformly distributed phase
- Rayleigh distributed magnitude

Lemma

The distribution of $\mathbf{H}$ is invariant under unitary transformations.

Suppose the inputs are circularly symmetric complex Gaussian of covariance Q, the maximal mutual information, expressed by

$$\Psi(Q) = \mathbb{E}[\log\det(I_r + \mathbf{H}Q\mathbf{H}^\dagger)] \tag{16}$$

where the non-negative definite Q subjects to $\mathrm{tr}(Q) \leq P$.

Ergodic capacity: optimal condition

Lemma.3 $\rightarrow \Psi(Q) = \Psi(D)$, where $Q = UDU^\dagger$. We focus on the case of nonnegative diagonal D.

Consider the permutation matrix $\Pi : Q^\Pi = \Pi Q \Pi^\dagger$, we have

$$\tilde{Q} = \frac{1}{t!} \sum_{\Pi} Q^\Pi \tag{17}$$

satisfies $\Psi(\tilde{Q}) \geq \Psi(Q)$ and $\mathrm{tr}(\tilde{Q}) = \mathrm{tr}(Q)$.

Note that $\tilde{Q} = \alpha I$, the capacity, i.e. the maximal mutual information, is obtained when $\frac{P}{t} I_r$, and given by

$$\Psi(Q) = \mathbb{E}[\log\det(I_r + \frac{P}{t}\mathbf{H}\mathbf{H}^\dagger)]. \tag{18}$$

Evaluation of capacity

For $m = \min\{N_t, N_r\} > 1$, $n = \max\{N_t, N_r\})$, let

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^\dagger & N_r < N_t \\ \mathbf{H}^\dagger\mathbf{H} & N_r \geq N_t. \end{cases} \tag{19}$$

Then $\mathbf{W}$ is an $m \times m$ random non-negative definite matrix and has real, non-negative eigenvalues $\lambda_1, \cdots, \lambda_m$. The channel capacity is

$$\mathbb{E}[\sum_{i=1}^{m} \log(1 + \frac{P}{t}\lambda_i)] \tag{20}$$

$\mathbf{W}$ follows the Wishart distribution with parameters (m, n)

Joint Density

Ordered eigenvalues

$$P_{\lambda,\text{ordered}}(\lambda_1, \cdots, \lambda_m) = K_{m,n}^{-1} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2 \tag{21}$$

Unordered eigenvalues

$$P_{\lambda,\text{unordered}}(\lambda_1, \cdots, \lambda_m) = (m! K_{m,n})^{-1} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2 \tag{22}$$

$$\mathbb{E}[\log(1 + \frac{P}{t}\lambda_i)] = \sum_{i=1}^{m} \mathbb{E}[\log(1 + \frac{P}{t}\lambda_i)] \tag{23}$$

unordered distribution: $= m\mathbb{E}[\log(1 + \frac{P}{t}\lambda_1)]$

Evaluation of capacity

Hence, the channel capacity is

$$\Psi(\mathbf{H}, P) = m\mathbb{E}[\log(1 + \frac{P}{t}\lambda_1)]$$

where $P_\lambda(\lambda_1) = \int \cdots \int P_\lambda(\lambda_1, \cdots, \lambda_m)\mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_m$.

Theorem

$$\Psi(\mathbf{H}, P) = \int \log(1 + \frac{P}{t}\lambda) \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} \left[L_k^{n,m}(\lambda)\right]^2 \lambda^{n-m} e^{-\lambda} d\lambda. \quad (24)$$

where $\{m, n\}$ *is* min *or* max *of* $\{N_t, N_r\}$, *and* $L_k^{n,m}$ *are the associated Laguerre polynomials.*

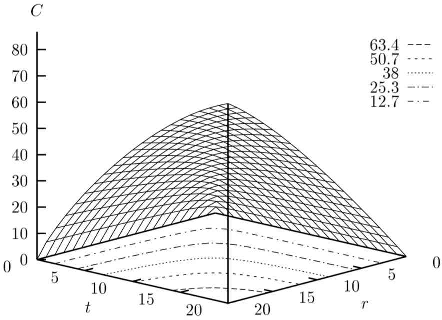


Figure 1: Capacity (in nats) vs. r and t for $P = 20$dB

Figure: Performance of diversity[1]

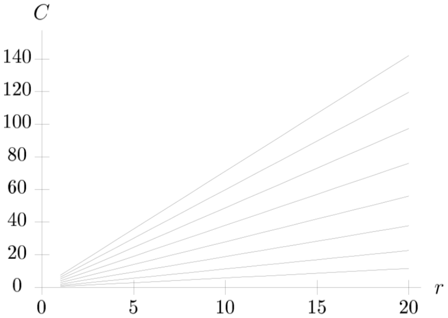


Figure: Performance of multiplexing[1]

Table of Contents

Constraints of point-to-point MIMO

Disadvantages:

- RF chain per antenna
- Low correlation among antenna array
- Rich scattering environment
- High SNR

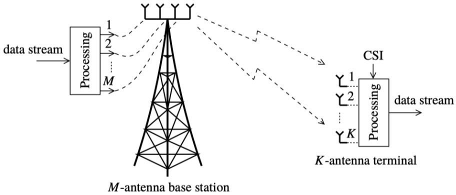


Figure: Point to point MIMO [2]

Multi-user MIMO

Advantages:

- Each terminal has a single antenna
- Less sensitive to the assumptions on propagation

Disadvantages

- Requiring global CSI, high cost for downlink
- Requiring complicated signal processing

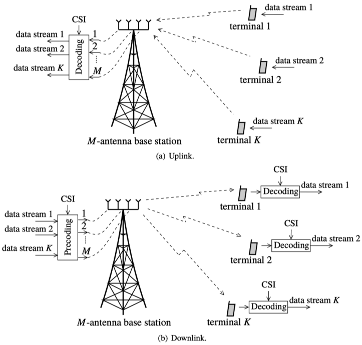


Figure: MU MIMO [2]

Massive MIMO

Scalable MU-MIMO [2] jointly exploit the beam-forming and spatial multiplexing.

- Only the BS learns CSI, via UL pilots
- BS has much more antennas than the amount of the UE
- Simple linear signal processing
  - Decoding SINR rather than SNR
  - Relying on spatial division

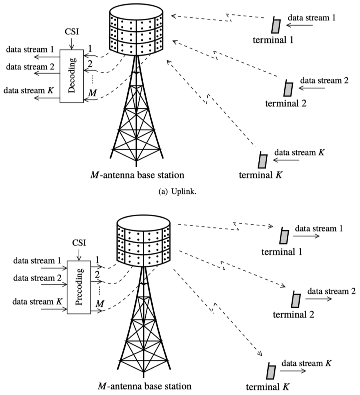


Figure: Massive MIMO [2]

Massive MIMO - MMSE or Zero forcing

To reduce the complexity, system sacrifice accuracy:
$\rightarrow$ Decoding SINR rather than SNR.

	zero forcing	MMSE
Uplink	$(M-K)\rho_{\mathrm{ul}}\beta_k\eta_k$	$\frac{M\rho_{\mathrm{ul}}\beta_k\eta_k}{1+\rho_{\mathrm{ul}}\sum_{k'=1}^{K}\beta_{k'}\eta_{k'}}$
Downlink	$(M-K)\rho_{\mathrm{dl}}\beta_k\eta_k$	$\frac{M\rho_{\mathrm{dl}}\beta_k\eta_k}{1+\rho_{\mathrm{dl}}\beta_k\sum_{k'=1}^{K}\eta_{k'}}$

Table: Effective SINR of kth UE for single cell system where perfect CSI at the BS is assumed.

where

- M is the number of antennas in the BS
- K is number of UEs
- ρ a dimensionless constant that scales the transmitted signal
- β_k is large-scale fading coefficients
- η_k is the power control coefficient

Performance

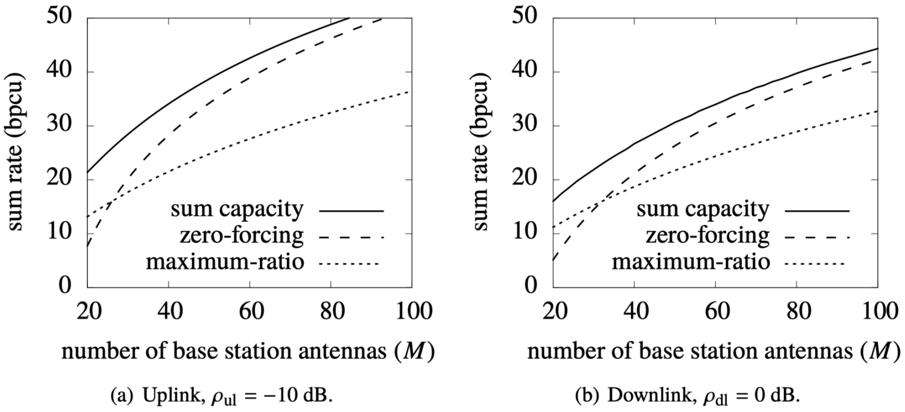


Figure: Sum capacity compared with zero-forcing and maximum-ratio as functions of number of BS antennas M, for number of UEs $K = 16$, and perfect CSI

Reference

E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

T. L. Marzetta, *Fundamentals of massive MIMO*.
Cambridge University Press, 2016.

END

Appendix: ergodic capacity under Rayleigh fading

Note that $\prod_{i<j}(\lambda_i, \lambda_j)$ is the determinant of the Vandermonde matrix

$$D(\lambda_1, \cdots, \lambda_m) = \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_m \\ \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \cdots & \lambda_m^{m-1} \end{bmatrix}$$

Hence, the PDF of unordered eigenvalues

$$P_{\lambda,}(\lambda_1, \cdots, \lambda_m) = (m! K_{m,n})^{-1} \det \left(D(\lambda_1, \cdots, \lambda_m) \right)^2 \prod_i \lambda_i^{n-m} e^{-\lambda_i} \quad (25)$$

To evaluate $D(\lambda_1, \cdots, \lambda_m)$, we apply row operation

$$\tilde{D}(\lambda_1, \cdots, \lambda_m) = \begin{bmatrix} \phi_1(\lambda_1) & \cdots & \phi_1(\lambda_m) \\ \vdots & \ddots & \vdots \\ \phi_m(\lambda_1) & \cdots & \phi_m(\lambda_m) \end{bmatrix} \tag{26}$$

where $\phi_1, \cdots, \phi_m$ is the results of applying the Gram-Schmidt orthogonalization to the sequence

$$1, \lambda, \cdots, \lambda^{m-1}$$

in the functional space

$$\langle f, g \rangle = \int_0^\infty f(\lambda) g(\lambda) \lambda^{n-m} e^{-\lambda} \mathrm{d}\lambda$$

Hence, $\int_0^\infty \phi_i(\lambda)\phi_j(\lambda)\lambda^{n-m}e^{-\lambda}\mathrm{d}\lambda = \delta_{i,j}$ and resulting in

$$\det\left(\tilde{D}(\lambda_1, \cdots, \lambda_m)\right) = \sum_{\alpha}(-1)^{\mathrm{per}(\alpha)}\prod_{i}\phi_{\alpha_i}(\lambda_i) \tag{27}$$

And

$$P_{\lambda,}(\lambda_1, \cdots, \lambda_m) = \frac{\sum_{\alpha,\beta}(-1)^{\mathrm{per}(\alpha)+\mathrm{per}(\beta)}\phi_{\alpha_1}(\lambda_1)\phi_{\beta_1}(\lambda_1)\lambda_1^{n-m}e^{-\lambda_1}}{(m!K_{m,n})} \tag{28}$$

Integrating over $\lambda_2, \cdots, \lambda_m$,

$$P_{\lambda}(\lambda_1) = \frac{1}{m}\sum_{i=1}^{m}\phi_i(\lambda_i)^2\lambda_1^{n-m}e^{-\lambda_1} \tag{29}$$

Given that the Gram-Schmidt orthonormalization yields

$$\phi_{k+1}(\lambda) = \left[\frac{k!}{(k+n-m)!}\right]^{1/2} L_k^{n-m}(\lambda), \quad k = 0, 1, \cdots, k-1. \tag{30}$$

where the Laguerre polynomial with order k is

$$L_k^{n-m}(x) = \frac{1}{k!} e^x x^{m-n} \frac{\mathrm{d}^k}{\mathrm{d}x^k}(e^{-x} x^{n+k-m}).$$

To summarize, the channel capacity is

$$\Psi(\mathbf{H}, P) = \int \log(1 + \frac{P}{t}\lambda) \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} \left[L_k^{n,m}(\lambda)\right]^2 \lambda^{n-m} e^{-\lambda} \mathrm{d}\lambda.$$