Advanced Wireless Communications

Bruno Clerckx

Department of Electrical and Electronic Engineering, Imperial College London
School of Electrical Engineering, Korea University

August 2014

Course Ojectives

- Advanced course on wireless communication and communication theory
  - Provides the fundamentals of wireless communications from a 4G and beyond perspective
  - At the cross-road between information theory, coding theory, signal processing and antenna/propagation theory
- Major focus of the course is on MIMO (Multiple Input Multiple Output) and multi-user/multi-cell communications
  - Includes as special cases SISO (Single Input Single Output), MISO (Multiple Input Single Output), SIMO (Single Input Multiple Output)
  - Applications: everywhere in wireless communication networks: 3G, 4G(LTE,LTE-A), (5G?), WiMAX(IEEE 802.16e, IEEE 802.16m), WiFi(IEEE 802.11n), satellite,...+ in other fields, e.g. radar, medical devices, speech and sound processing, ...
- Valuable for those who want to either pursue a PhD in communication or a career in a high-tech telecom company (research centres, R&D branches of telecom manufacturers and operators,...).
- Skills
  - Mathematical modelling and analysis of (MIMO-based) wireless communication systems
  - Design (transmitters and receivers) of multi-cell multi-user MIMO wireless communication systems
  - Practical understanding of MIMO applications and performance evaluations

Content

Central question: How to deal with fading and interference in wireless networks?

- Some fundamentals/revision (matrix analysis, probability, information theory)
- **Single link: point to point communications**
  - Fading and Diversity
  - MIMO Channels - Modelling and Propagation
  - Capacity of point-to-point MIMO Channels
  - Space-Time Coding/Decoding over I.I.D. Rayleigh Flat Fading Channels
  - Space-Time Coding in Real-World MIMO Channels
  - Partial Channel State Information at the Transmitter (CSIT)
  - Frequency-Selective MIMO Channels - MIMO-OFDM
- **Multiple links: multiuser communications**
  - Multi-User MIMO - Capacity of Multiple Access Channels (Uplink)
  - Multi-User MIMO - Capacity of Broadcast Channels (Downlink)
  - Multi-User MIMO - Scheduling, Linear and Non-Linear Precoding, DPC (Downlink)
  - Multi-User MIMO (Downlink) with/for Imperfect CSIT
- **Multiple cells: multiuser multicell communications**
  - Introduction to Multi-Cell MIMO
  - Capacity of Interference Channel
  - Coordinated Scheduling and Power Control
  - Coordinated Beamforming and Interference Alignment
  - Network MIMO

Content

- **Massive MIMO**
- **Real-World MIMO Wireless Networks**
  - MIMO and Interference Management in 4G and beyond (LTE, LTE-Advanced, WiMAX)
  - Homogeneous and Heterogeneous Networks
  - System-Level Performance Evaluations
- **Additional topics** (if time permits)
  - Modeling of Wireless Networks: Stochastic Geometry
  - Wireless Communication and Power Networks - Energy Harvesting

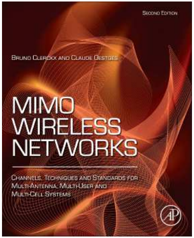
Important Information

- Course webpage: http://www.ee.ic.ac.uk/bruno.clerckx/Teaching.html
- Prerequisite: Good background on Communication Theory
- Lectures
  - Week 1: Wednesday, Thursday, Friday from 09.30 till 12.15
  - Week 2: Monday, Tuesday, Wednesday, Thursday from 09.30 till 12.15
  - Week 3: Monday, Tuesday, Wednesday, Thursday from 09.30 till 12.15
  - Week 4: Monday, Tuesday, Wednesday, Thursday from 09.30 till 12.15
- Problem sheets from Imperial College EE4-65/EE9-SO27 Wireless Communications course available on course webpage (2 types: 1. paper/pencil, 2. matlab)
- Matlab project also available on course webpage
  - Encourage students to work on it if time permits.

Important Information

- Reference book

  Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.

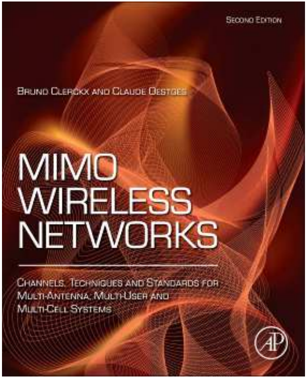


- Another interesting reference on wireless communications (more introductory) "Fundamentals of Wireless Communication," by D. Tse and P. Viswanath, Cambridge University Press, May 2005

Some fundamentals/revisions (matrix analysis, probability, information theory)

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Appendix A, B

- T. Cover and J. Thomas, "Elements of Information Theory," Second Edition, Wiley, 2006.

Matrix properties

- *Vector Orthogonality*: $\mathbf{a}^H\mathbf{b} = 0$ (^{H} stands for Hermitian, i.e. conjugate transpose)
- *Hermitian matrix*: $\mathbf{A} = \mathbf{A}^H$
- *Unitary matrix*: $\mathbf{A}^H\mathbf{A} = \mathbf{I}$
- *Singular Value Decomposition* (SVD) of a matrix $\mathbf{H}$ $[n_r \times n_t]$: $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$
  - $\mathbf{U}$ $[n_r \times r(\mathbf{H})]$: unitary matrix of left singular vectors
  - $\mathbf{\Sigma} = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\}$: diagonal matrix containing the singular values of $\mathbf{H}$
  - $\mathbf{V}$ $[n_t \times r(\mathbf{H})]$: unitary matrix of left singular vectors
  - $r(\mathbf{H})$: the rank of $\mathbf{H}$

  We will often look at Hermitian matrices of the form $\mathbf{A} = \mathbf{H}^H\mathbf{H}$ whose *Eigenvalue Value Decomposition* (EVD) writes as $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ with $\mathbf{\Lambda} = \mathbf{\Sigma}^2$.
- *Trace* of a matrix $\mathbf{A}$: $\text{Tr}\{\mathbf{A}\} = \sum_i \mathbf{A}(i,i)$.
- *Frobenius norm* of a matrix $\mathbf{A}$: $\|\mathbf{A}\|_F^2 = \sum_i \sum_j |A(i,j)|^2$
- $\|\mathbf{A}\|_F^2 = \text{Tr}\left\{\mathbf{A}\mathbf{A}^H\right\} = \text{Tr}\left\{\mathbf{A}^H\mathbf{A}\right\}$
- $\text{Tr}\{\mathbf{A}\mathbf{B}\} = \text{Tr}\{\mathbf{B}\mathbf{A}\}$
- *Hadamard's inequality*: $\det(\mathbf{A}) \leq \prod_{k=1}^{n} \mathbf{A}(k,k)$ if $\mathbf{A} > 0$ of size $n \times n$

Matrix properties

- *Kronecker product*: $\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}(1,1)\mathbf{B} & \dots & \mathbf{A}(1,n)\mathbf{B} \\ \vdots & \dots & \vdots, \\ \mathbf{A}(m,1)\mathbf{B} & \dots & \mathbf{A}(m,n)\mathbf{B} \end{bmatrix}$
- $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$
- $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$
- $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ if $\mathbf{A}, \mathbf{B}$ square and non singular.
- $\det(\mathbf{A}_{m\times m} \otimes \mathbf{B}_{n \times n}) = \det(\mathbf{A})^n \det(\mathbf{B})^m$
- $\mathrm{Tr}\{\mathbf{A} \otimes \mathbf{B}\} = \mathrm{Tr}\{\mathbf{A}\}\,\mathrm{Tr}\{\mathbf{B}\}$
- $\mathrm{Tr}\{\mathbf{AB}\} \geq \mathrm{Tr}\{\mathbf{A}\}\,\sigma_{min}^2(\mathbf{B})$ with $\sigma_{min}(\mathbf{B})$ the smallest singular value of $\mathbf{B}$
- $\mathrm{vec}(\mathbf{A})$ converts $[m \times n]$ matrix into $mn \times 1$ vector by stacking the columns of $\mathbf{A}$ on top of one another.
  - $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$
- $\det(\mathbf{I} + \epsilon\mathbf{A}) = 1 + \epsilon \mathrm{Tr}\{\mathbf{A}\}$ if $\epsilon << 1$

Gaussian random variable

- *Real Gaussian random variable* x with mean $\mu = \mathcal{E}\{x\}$ and variance σ^2

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

  Standard Gaussian random variable: $\mu = 0$ and $\sigma^2 = 1$
- *Real Gaussian random vector* $\mathbf{x}$ of dimension n with mean vector $\boldsymbol{\mu} = \mathcal{E}\{\mathbf{x}\}$ and covariance matrix $\mathbf{R} = \mathcal{E}\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right\}$:

$$p(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(\mathbf{R})}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{R}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}\right).$$

  Standard Gaussian random vector $\mathbf{x}$ of dimension n: entries are independent and identically distributed (i.i.d.) standard Gaussian random variables $x_1, \ldots, x_n$

$$p(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right).$$

Gaussian random variable

- *Complex Gaussian random variable* $x = x_r + jx_i$: $[x_r, x_i]^T$ is a real Gaussian random vector.
- Important case: $x = x_r + jx_i$ is such that its real and imaginary parts are i.i.d. zero mean Gaussian variables of variance σ^2 (circularly symmetric complex Gaussian random variable).
- $s = |x| = \sqrt{x_r^2 + x_i^2}$ is Rayleigh distributed

$$p(s) = \frac{s}{\sigma^2} \exp\left(-\frac{s^2}{2\sigma^2}\right).$$

- $y = s^2 = |x|^2 = x_r^2 + x_i^2$ is χ_2^2 (i.e. exponentially) distributed (with two degrees of freedom)

$$p_y(y) = \frac{1}{2\sigma^2} \exp\left(-\frac{y}{2\sigma^2}\right).$$

  Hence, $\mu = \mathcal{E}\left\{y\right\} = 2\sigma^2$.
- More generally, χ_n^2 is the sum of the square of n i.i.d. zero-mean Gaussian random variables.
- Assume n i.i.d. zero mean complex Gaussian variables $h_1, \ldots, h_n$ (real and imaginary parts with variance σ^2). Defining $u = \sum_{k=1}^{n} |h_k|^2$, the MGF of u is given by

$$\mathcal{M}_u(\tau) = \mathcal{E}\{e^{\tau u}\} = \left[\frac{1}{1 - 2\sigma^2\tau}\right]^n,$$

Discrete Memoryless Channel

Definition

A *discrete* channel is defined as a system consisting of an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and a probability transition matrix $p(y|x)$ that expresses the probability of observing the output symbols y given that the symbol x is sent.

Definition

The channel is *memoryless* if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs, i.e. if $x_1,...,x_n$ are inputs, and $y_1,...,y_n$ denote the corresponding outputs, for n channel uses, then

$$p(y_1, ..., y_n|x_1, ...x_n) = p(y_1|x_1)...p(y_n|x_n)$$

Example

Binary Symmetric Channel (BSC): x and y take values in 0,1 such that

$$p(y|x) = \begin{cases} 1-p, & y = x, \\ p, & y = 1-x. \end{cases}$$

Entropy

- Entropy is a measure of the average uncertainty of a random variable

Definition

For a discrete random variable X, the entropy $H(X)$ is defined as

$$H(X) = \mathcal{E}\left\{\frac{1}{\log_2 p(X)}\right\} = -\mathcal{E}\left\{\log_2 p(X)\right\} = -\sum_x p(x)\log_2 p(x),$$

where $p(x)$ is the probability mass function of X.

- It is the number of bits on average required to describe the random variable.

Example

Let X be a Bernoulli random variable

$$X = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p. \end{cases}$$

Then $H(X) = -p\log_2 p - (1-p)\log_2(1-p)$. For $p = 0, 1$, there is no uncertainty on the value of the RV, so no information gained. For $p = 1/2$, $H(X)$ (uncertainty/information) is maximized.

Entropy

Lemma

$H(X) \geq 0$

Proof: $0 \leq p(x) \leq 1$ such that $\frac{1}{\log_2 p(x)} \geq 0$ □

Definition

The joint entropy $H(X, Y)$ of a pair of discrete random variables X and Y with a joint pmf $p(x, y)$ is defined as

$$H(X, Y) = -\mathcal{E}\{\log_2 p(X, Y)\} = -\sum_x \sum_y p(x, y) \log_2 p(x, y)$$

Conditional Entropy

- The conditional entropy of a random variable given another is the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable

Definition

The conditional entropy $H(Y|X)$ is defined as

$$\begin{aligned} H(Y|X) &= \sum_x p(x) H(Y|X=x) \\ &= -\sum_x p(x) \sum_y p(y|x) \log_2 p(y|x) \\ &= -\sum_x \sum_y p(x,y) \log_2 p(y|x) \\ &= -\mathcal{E}\left\{\log_2 p(Y|X)\right\} \end{aligned}$$

Joint Entropy

Theorem

Chain rule

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

Proof:

$$\begin{aligned}
H(X,Y) &= -\sum_x \sum_y p(x,y) \log_2 p(x,y) = -\sum_x \sum_y p(x,y) \log_2 p(x)p(y|x) \\
&= -\sum_x \sum_y p(x,y) \log_2 p(x) - \sum_x \sum_y p(x,y) \log_2 p(y|x) \\
&= -\sum_x p(x) \log_2 p(x) - \sum_x \sum_y p(x,y) \log_2 p(y|x) \\
&= H(X) + H(Y|X)
\end{aligned}$$

Alternatively,

$$\log_2 p(X,Y) = \log_2 p(X) + \log_2 p(Y|X)$$
$$\mathcal{E}\left\{\log_2 p(X,Y)\right\} = \mathcal{E}\left\{\log_2 p(X)\right\} + \mathcal{E}\left\{\log_2 p(Y|X)\right\}$$

Relative Entropy

- The relative entropy is a measure of the distance between two distributions.

Definition

The relative entropy between two pmf $p(x)$ and $q(x)$ is defined as

$$D(p||q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} = \mathcal{E}_p \left\{ \log_2 \frac{p(X)}{q(X)} \right\}$$

Theorem

The relative entropy is always nonnegative $D(p||q) \geq 0$ *and is zero if and only if* $p = q$.

Mutual Information

- The mutual information is a measure of the amount of information that one RV contains about another RV. It is a measure of the dependence between the two RVs.

Definition

For a pair of discrete random variables X and Y with a joint pmf $p(x, y)$ and marginal pmf $p(x)$ and $p(y)$, the mutual information $I(X; Y)$ is the relative entropy between $p(x, y)$ and $p(x)p(y)$

$$I(X;Y) = D(p(x,y)||p(x)p(y)) = \mathcal{E}_{p(x,y)}\left\{\log_2 \frac{p(X,Y)}{p(X)p(Y)}\right\}$$

$$= \sum_x \sum_y p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Mutual Information

- The mutual information $I(X;Y)$ is the reduction in the uncertainty of one random variable due to the knowledge of the other

$$
\begin{aligned}
I(X;Y) &= \sum_{x,y} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x,y} p(x,y)\log_2 \frac{p(x|y)}{p(x)} \\
&= -\sum_{x,y} p(x,y)\log_2 p(x) + \sum_{x,y} p(x,y)\log_2 p(x|y) \\
&= -\sum_{x} p(x)\log_2 p(x) - \left(-\sum_{x,y} p(x,y)\log_2 p(x|y)\right) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) = I(Y;X)
\end{aligned}
$$

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

Mutual Information

Theorem

Nonnegativity of mutual information: For any two random variables X,Y

$$I(X;Y) \geq 0$$

with equality if and only if X *and* Y *are independent*

Theorem

Conditioning reduces entropy: For any two random variables X,Y

$$H(X|Y) \leq H(X)$$

with equality if and only if X *and* Y *are independent*

Proof: $0 \leq I(X;Y) = H(X) - H(X|Y)$ □

Knowing another RV Y can only reduce on the average the uncertainty in X.

Channel Coding Theorem

Theorem

(a) For a DMC with channel transition pmf $p(y|x)$, we can use i.i.d. inputs with pmf $p(x)$ to communicate reliably, as long as the code rate satisfies

$$R < I(X;Y).$$

(b) The achievable rate can be maximized over the input density $p(x)$ to obtain the channel capacity

$$C = \max_{p(x)} I(X;Y).$$

Differential Entropy

Definition

For a continuous random variable X, the differential entropy $h(X)$ is defined as

$$h(X) = \mathcal{E}\left\{\frac{1}{\log_2 p(x)}\right\} = -\mathcal{E}\left\{\log_2 p(x)\right\} = -\int p(x)\log_2 p(x)dx,$$

where $p(x)$ is the probability density function of X.

Caution: $h(X)$ can be negative.

Example

For $X \sim N(\mu, \sigma^2)$, $-\log_2 p(x) = \frac{(x-\mu)^2}{2\sigma^2}\log_2(e) + \frac{1}{2}\log_2(2\pi\sigma^2)$. Thus, $h(X) = -\mathcal{E}\left\{\log_2 p(x)\right\} = \frac{1}{2}\log_2(e) + \frac{1}{2}\log_2(2\pi\sigma^2) = \frac{1}{2}\log_2(2\pi e\sigma^2)$. The mean does not affect the differential entropy.

Theorem

Consider a RV with zero mean and variance σ^2. Then $h(X) \leq \frac{1}{2}\log_2(2\pi e\sigma^2)$, with equality iff $X \sim N(0, \sigma^2)$.

AWGN Channel

- Real discrete-time AWGN channel

$$Y = X + N, \qquad N \sim N(0, \sigma^2)$$

  where X is power-constrained input $\mathcal{E}\left\{X^2\right\} \leq E_s$

- The channel transition density is given by

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$$

AWGN Channel Capacity

Theorem

The capacity of the real AWGN channel is

$$C = \max_{p(x):\mathcal{E}\{X^2\}\leq E_s} I(X;Y) = \frac{1}{2}\log_2(1+\frac{E_s}{\sigma^2}).$$

Proof: Consider $Y = X + N$, with $N \sim N(0,\sigma^2)$ and $\mathcal{E}\{X^2\} \leq E_s$. Given $X = x$, $h(Y|X=x) = h(N)$, so that $h(Y|X) = h(N)$ and

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(N).$$

Maximizing $I(X;Y)$ comes to maximize $h(Y)$. Since X and N are independent, $\mathcal{E}\{Y^2\} = \mathcal{E}\{X^2\} + \mathcal{E}\{N^2\} \leq E_s + \sigma^2$. We now know that

$$h(Y) \leq \frac{1}{2}\log_2(2\pi e(E_s+\sigma^2))$$

and equality is achieved iff $Y \sim N(0, E_s+\sigma^2)$. $Y \sim N(0, E_s+\sigma^2)$ is achieved if the input distribution is $X \sim N(0, E_s)$, independent of the noise. We then get

$$I(X;Y) = h(Y) - h(Z) = \frac{1}{2}\log_2(2\pi e(E_s+\sigma^2)) - \frac{1}{2}\log_2(2\pi e\sigma^2) = \frac{1}{2}\log_2(1+\frac{E_s}{\sigma^2}).$$

Jensen's inequality

Theorem

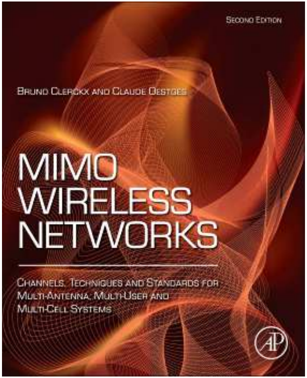
If f is a convex function and X is a random variable,

$$\mathcal{E}\{f(X)\} \geq f(\mathcal{E}\{X\}).$$

Fading and Diversity

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Chapter 1
    - Section: 1.2, 1.3, 1.4, 1.5
    - Appendix A, B

Space-Time Wireless Channels:Discrete Time Representation

- *channel*: the impulse response of the linear time-varying communication system between one (or more) transmitter(s) and one (or more) receiver(s).
- Assume a SISO transmission where the digital signal is defined in discrete-time by the complex time series $\{c_l\}_{l\in}$ and is transmitted at the symbol rate T_s.
- The transmitted signal is then represented by

$$c(t) = \sum_{l=-\infty}^{\infty} \sqrt{E_s} c_l \delta(t - lT_s),$$

where E_s is the transmitted symbol energy, assuming that the average energy constellation is normalized to unity.
- Define a function $\mathsf{h}_B(t, \tau)$ as the time-varying (along variable t) impulse response of the channel (along τ) over the system bandwidth $B = 1/T_s$, i.e. $\mathsf{h}_B(t, \tau)$ is the response at time t to an impulse at time $t - \tau$.
- The received signal $\mathsf{y}(t)$ is given by

$$\begin{aligned} \mathsf{y}(t) &= \mathsf{h}_B(t, \tau) \star c(t) + \mathsf{n}(t) \\ &= \int_0^{\tau_{max}} \mathsf{h}_B(t, \tau) c(t - \tau) d\tau + \mathsf{n}(t) \end{aligned}$$

where $\star$ denotes the convolution product, $\mathsf{n}(t)$ is the additive noise of the system and τ_{max} is the maximal length of the impulse response.

Discrete Time Representation

- h_B is a scalar quantity, which can be further decomposed into three main terms,

$$\mathsf{h}_B(t,\tau) = f_r \star \mathsf{h}(t,\tau) \star f_t,$$

  where
  - f_t is the pulse-shaping filter,
  - $\mathsf{h}(t,\tau)$ is the electromagnetic propagation channel (including the transmit and receive antennas) at time t,
  - f_r is the receive filter.
- Nyquist criterion: the cascade $f = f_r \star f_t$ does not create inter-symbol interference when $\mathsf{y}(t)$ is sampled at rate T_s.
- In practice,
  - difficult to model $\mathsf{h}(t,\tau)$ (infinite bandwidth is required).
  - $\mathsf{h}_B(t,\tau)$ is usually the modeled quantity, but is written as $\mathsf{h}(t,\tau)$ (abuse of notation).
  - Same notational approximation: the channel impulse response writes as $\mathsf{h}(t,\tau)$ or $\mathsf{h}_t[\tau]$.
- The input-output relationship reads thereby as

$$\mathsf{y}(t) = \mathsf{h}(t,\tau) \star c(t) + \mathsf{n}(t) = \sum_{l=-\infty}^{\infty} \sqrt{E_s} c_l \mathsf{h}_t[t - lT_s] + \mathsf{n}(t).$$

Discrete Time Representation

- Sampling the received signal at the symbol rate T_s ($\mathsf{y}_k = \mathsf{y}(t_0 + kT_s)$, using the epoch t_0) yields

$$\begin{aligned}\mathsf{y}_k &= \sum_{l=-\infty}^{\infty} \sqrt{E_s} c_l \mathsf{h}_{t_0+kT_s}[t_0 + (k-l)T_s] + \mathsf{n}(t_0 + kT_s) \\ &= \sum_{l=-\infty}^{\infty} \sqrt{E_s} c_l \mathsf{h}_k[k-l] + \mathsf{n}_k\end{aligned}$$

Example

At time $k = 0$, the channel has two taps: $\mathsf{h}_0[0]$, $\mathsf{h}_0[1]$

$$\mathsf{y}_0 = \sqrt{E_s}\left[c_0 \mathsf{h}_0[0] + c_{-1}\mathsf{h}_0[1]\right] + \mathsf{n}_0$$

- If $T_s >> \tau_{max}$,
  - $\mathsf{h}_B(t,\tau)$ is modeled by a single dependence on t: write simply as $\mathsf{h}_B(t)$ (or $\mathsf{h}(t)$ using the same abuse of notation). In the sampled domain, $\mathsf{h}_k = \mathsf{h}(t_0 + kT_s)$.
  - the channel is then said to be *flat fading* or narrowband

$$\mathsf{y}_k = \sqrt{E_s}\mathsf{h}_k c_k + \mathsf{n}_k$$

- Otherwise the channel is said to be *frequency selective*.

Path-Loss and Shadowing

- Assuming narrowband channels and given specific Tx and Rx locations, h_k is modeled as
$$\mathsf{h}_k = \frac{1}{\sqrt{\Lambda_0\, S}}\, h_k,$$
where
  - *path-loss* Λ_0: a real-valued deterministic attenuation term modeled as $\Lambda_0 \propto R^\eta$ where η designates the path-loss exponent and R the distance between Tx and Rx.
  - *shadowing* S: a real-valued random additional attenuation term, which, for a given range, depends on the specific location of the transmitter and the receiver and modeled as a lognormal variable, i.e., $10\log_{10}(S)$ is a zero-mean normal variable of given standard deviation σ_S.
  - *fading* h_k: caused by the combination of non coherent multipaths. By definition of Λ_0 and S, $\mathcal{E}\left\{\, |h|^2 \,\right\} = 1$.

- Alternatively, $\mathsf{h}_k = \Lambda^{-1/2}\, h_k$ with Λ modeled on a logarithm scale
$$\Lambda|_{\mathsf{dB}} = \Lambda_0|_{\mathsf{dB}} + S|_{\mathsf{dB}} = L_0|_{\mathsf{dB}} + 10\eta\log_{10}\left(\frac{R}{R_0}\right) + S|_{\mathsf{dB}},$$
where $|_{\mathsf{dB}}$ indicates the conversion to dB, and L_0 is the deterministic path-loss at a reference distance R_0, and Λ is generally known as the path-loss.

Path-Loss and Shadowing

- Path loss models are identical for both single- and multi-antenna systems.

- For point to point systems, it is common to discard the path loss and shadowing and only investigate the effect due to fading, i.e. the classical model for narrowband channels

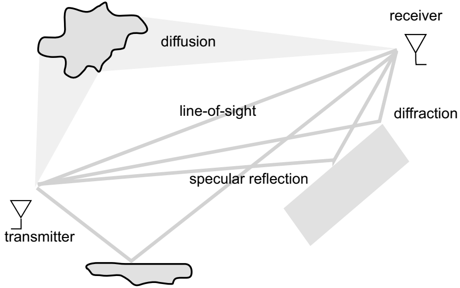
$$y = \sqrt{E_s} hc + n,$$

  where the time index is removed for better legibility and n is usually taken as white Gaussian distributed, $\mathcal{E}\{n_k n_l^*\} = \sigma_n^2 \delta(k-l)$.

- E_s can then be seen as an average received symbol energy. The average SNR is then defined as $\rho \triangleq E_s / \sigma_n^2$.

Fading

- Multipaths



- Assuming that the signal reaches the receiver via a large number of paths of similar energy,
  - h is modeled such that its real and imaginary parts are i.i.d. zero mean Gaussian variables of variance σ^2 (circularly symmetric complex Gaussian variable).
  - Recall $\mathcal{E}\{\,|h|^2\,\} = 2\sigma^2 = 1$.

Fading

- The channel *amplitude* $s \triangleq |h|$ follows a *Rayleigh* distribution,

$$p_s(s) = \frac{s}{\sigma^2} \exp\left(-\frac{s^2}{2\sigma^2}\right),$$

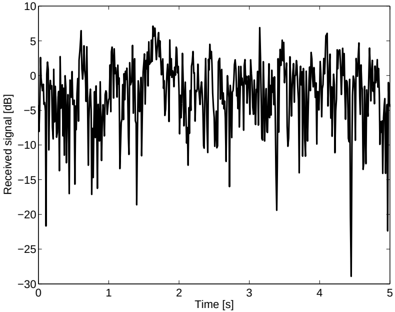
whose first two moments are

$$\mathcal{E}\{s\} = \sigma\sqrt{\frac{\pi}{2}}$$
$$\mathcal{E}\{s^2\} = 2\sigma^2 = \mathcal{E}\{\,|h|^2\,\} = 1.$$

- The *phase* of h is uniformly distributed over $[0, 2\pi)$

Fading

- Illustration of the typical received signal strength of a Rayleigh fading channel over a certain time interval



  - The signal level randomly fluctuates, with some sharp declines of power and instantaneous received SNR known as *fades*.
  - When the channel is in a deep fade, a reliable decoding of the transmitted signal may not be possible anymore, resulting in an error.
  - How to recover the signal? Use of diversity techniques

Maximum likelihood detection

- Decision rule: choose the hypothesis that maximizes the conditional density

$$\arg\max_x p(y|x) = \arg\max_x \log p(y|x)$$

- If real AWGN $y = x + n$ with $n \sim N(0, \sigma_n^2)$,

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y-x)^2}{2\sigma_n^2}\right)$$

  and

$$\arg\max_x p(y|x) = \arg\max_x \,(y-x)^2$$

- If $y = \sqrt{E_s}hc + n$, the ML decision rule becomes

$$\arg\max_c \left\| y - \sqrt{E_s}hc \right\|^2$$

Diversity in Multiple Antennas Wireless Systems

- What is the impact of fading on system performance?
- Consider the simple case of BPSK transmission through an AWGN channel and a SISO Rayleigh fading channel:
  - In the absence of fading ($h = 1$), the symbol-error rate (SER) in an additive white Gaussian noise (AWGN) channel is given by

$$\bar{P} = \mathcal{Q}\left(\sqrt{\frac{2E_s}{\sigma_n^2}}\right) = \mathcal{Q}(\sqrt{2\rho}),$$

    where $\mathcal{Q}(x)$ is the Gaussian $\mathcal{Q}$-function defined as

$$\mathcal{Q}(x) \triangleq P(y \geq x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{y^2}{2}\right) dy.$$

  - In the presence of (Rayleigh) fading, the received signal level fluctuates as $s\sqrt{E_s}$, and the SNR varies as ρs^2. As a result, the SER

$$\begin{aligned} \bar{P} &= \int_0^{\infty} \mathcal{Q}(\sqrt{2\rho}s) p_s(s)\, ds \\ &= \frac{1}{2}\left(1 - \sqrt{\frac{\rho}{1+\rho}}\right) \\ &\overset{(\rho\nearrow)}{\cong} \frac{1}{4\rho} \end{aligned}$$

    although the average SNR $\bar{\rho} = \int_0^{\infty} \rho s^2\, p_s(s)\, ds$ remains equal to ρ.

Diversity in Multiple Antennas Wireless Systems

- How to combat the impact of fading? Use diversity techniques
- The principle of diversity is to provide the receiver with multiple versions (called diversity branch) of the same transmitted signal.
  - Independent fading conditions across branches needed.
  - Diversity stabilizes the link through channel hardening which leads to better error rate.
  - Multiple domains: time (coding and interleaving), frequency (equalization and multi-carrier modulations) and space (multiple antennas/polarizations).
- *Array Gain*: increase in average output SNR (i.e., at the input of the detector) relative to the single-branch average SNR ρ

$$g_a \triangleq \frac{\bar{\rho}_{out}}{\bar{\rho}} = \frac{\bar{\rho}_{out}}{\rho}$$

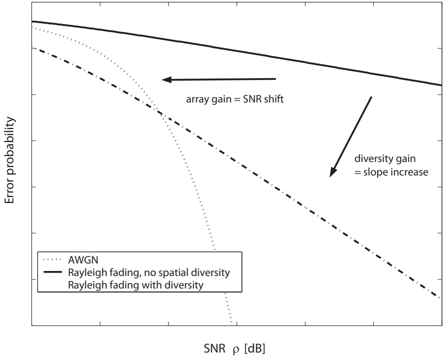
- *Diversity Gain*: increase in the error rate slope as a function of the SNR. Defined as the negative slope of the log-log plot of the average error probability $\bar{P}$ versus SNR

$$g_d^o(\rho) \triangleq -\frac{\log_2\left(\bar{P}\right)}{\log_2\left(\rho\right)}.$$

The diversity gain is commonly taken as the asymptotic slope, i.e., for $\rho \to \infty$.

Diversity in Multiple Antennas Wireless Systems

- Illustration of diversity and array gains



Careful that the curves have been plotted against the single-branch average SNR $\bar{\rho} = \rho$!
If plotted against the output average SNR $\bar{\rho}_{out}$, the array gain disappears.

- *Coding Gain*: a shift of the error curve (error rate vs. SNR) to the left, similarly to the array gain.
  - If the error rate vs. the average receive SNR $\bar{\rho}_{out}$, any variation of the array gain is invisible but any variation of the coding gain is visible: for a given SNR level $\bar{\rho}_{out}$ at the input of the detector, the error rates will differ.

SIMO Systems

- Receive diversity may be implemented via two rather different combining methods:
  - *selection combining*: the combiner selects the branch with the highest SNR among the n_r receive signals, which is then used for detection,
  - *gain combining*: the signal used for detection is a linear combination of all branches, $z = \mathbf{g}\mathbf{y}$, where $\mathbf{g} = [g_1, \ldots, g_{n_r}]$ is the combining vector.
    1. Equal Gain Combining
    2. Maximal Ratio Combining
    3. Minimum Mean Square Error Combining
- Space antennas sufficiently far apart from each other so as to experience independent fading on each branch.
- We assume that the receiver is able to acquire the perfect knowledge of the channel.

Receive Diversity via Selection Combining

- Assume that the n_r channels are independant and identically Rayleigh distributed (i.i.d.) with unit energy and that the noise levels are equal on each antenna.
- Choose the branch with the largest amplitude $s_{max} = \max\{s_1, \ldots, s_{n_r}\}$.
- The probability that s falls below a certain level S is given by its CDF

$$P[s < S] = 1 - e^{-S^2/2\sigma^2}.$$

- The probability that s_{max} falls below a certain level S is given by

$$P\left[s_{max} < S\right] = P[s_1, \ldots, s_{n_r} \leq S] = \left[1 - e^{-S^2}\right]^{n_r}.$$

- The PDF of s_{max} is then obtained by derivation of its CDF

$$p_{s_{max}}(s) = n_r\, 2s\, e^{-s^2} \left[1 - e^{-s^2}\right]^{n_r - 1}.$$

- The average SNR at the output of the combiner $\bar{\rho}_{out}$ is eventually given by

$$\bar{\rho}_{out} = \int_0^\infty \rho s^2 p_{s_{max}}(s)\, ds = \rho \sum_{n=1}^{n_r} \frac{1}{n} \overset{n_r \nearrow}{\approx} \rho \left[\gamma + \log(n_r) + \frac{1}{2n_r}\right].$$

where $\gamma \approx 0.57721566$ is Euler's constant. We observe that the array gain g_a is of the order of $\log(n_r)$.

Receive Diversity via Selection Combining

- For BPSK and a two-branch diversity, the SER as a function of the average SNR per channel ρ writes as

$$\begin{aligned}\bar{P} &= \int_0^\infty \mathcal{Q}\left(\sqrt{2\rho s}\right) p_{s_{max}}(s)\, ds \\ &= \frac{1}{2} - \sqrt{\frac{\rho}{1+\rho}} + \frac{1}{2}\sqrt{\frac{\rho}{2+\rho}} \\ &\overset{\rho \nearrow}{\cong} \frac{3}{8\rho^2}.\end{aligned}$$

  The slope of the bit error rate curve is equal to 2.

- In general, the diversity gain g_d^o of a n_r-branch selection diversity scheme is equal to n_r. Selection diversity extracts all the possible diversity out of the channel.

Receive Diversity via Gain Combining

- In gain combining, the signal z used for detection is a linear combination of all branches,

$$z = \mathbf{g}\mathbf{y} = \sum_{n=1}^{n_r} g_n y_n = \sqrt{E_s}\mathbf{g}\mathbf{h}c + \mathbf{g}\mathbf{n}$$

  where
  - g_n's are the combining weights and $\mathbf{g} \triangleq [g_1, \ldots, g_{n_r}]$
  - the data symbol c is sent through the channel and received by n_r antennas
  - $\mathbf{h} \triangleq [h_1, \ldots, h_{n_r}]^T$
- Assume Rayleigh distributed channels $h_n = |h_n| e^{j\phi_n}$, $n = 1, \ldots, n_r$, with unit energy, all the channels being independent.
- *Equal Gain Combining*: fixes the weights as $g_n = e^{-j\phi_n}$.
  - Mean value of the output SNR $\bar{\rho}_{out}$ (averaged over the Rayleigh fading):

$$\bar{\rho}_{out} = \frac{\mathcal{E}\left\{\left[\sum_{n=1}^{n_r} \sqrt{E_s}\,|h_n|\right]^2\right\}}{n_r \sigma_n^2} = \ldots = \rho\left[1 + (n_r - 1)\frac{\pi}{4}\right],$$

    where the expectation is taken over the channel statistics. The array gain grows linearly with n_r, and is therefore larger than the array gain of selection combining.
  - The diversity gain of equal gain combining is equal to n_r analogous to selection.

Receive Diversity via Gain Combining

- *Maximal Ratio Combining*:the weights are chosen as $g_n = h_n^*$.
  - It maximizes the average output SNR $\bar{\rho}_{out}$

$$\bar{\rho}_{out} = \frac{E_s}{\sigma_n^2}\mathcal{E}\left\{\frac{\|\mathbf{h}\|^4}{\|\mathbf{h}\|^2}\right\} = \rho\mathcal{E}\left\{\|\mathbf{h}\|^2\right\} = \rho n_r.$$

    The array gain g_a is thus always equal to n_r, or equivalently, the output SNR is the sum of the SNR levels of all branches (holds true irrespective of the correlation between the branches).
  - For BPSK transmission, the symbol error rate reads as

$$\bar{P} = \int_0^\infty \mathcal{Q}\left(\sqrt{2\rho u}\right) p_u(u)\, du$$

    where $u = \|\mathbf{h}\|^2$ is χ^2 distribution with $2n_r$ degrees of freedom when the different channels are i.i.d. Rayleigh

$$p_u(u) = \frac{1}{(n_r - 1)!}\, u^{n_r - 1} e^{-u}.$$

    At high SNR, $\bar{P}$ becomes

$$\bar{P} = (4\rho)^{-n_r}\begin{pmatrix} 2n_r - 1 \\ n_r \end{pmatrix}.$$

    The diversity gain is again equal to n_r.

Receive Diversity via Gain Combining

- For alternative constellations, the error probability is given, assuming ML detection, by

$$\bar{P} \approx \int_0^\infty \bar{N}_e \mathcal{Q}\left(d_{min}\sqrt{\frac{\rho u}{2}}\right) p_u(u)\, du,$$
$$\leq \bar{N}_e \mathcal{E}\left\{ e^{-\frac{d_{min}^2 \rho u}{4}} \right\} \qquad \text{(using Chernoff bound } \mathcal{Q}(x) \leq \exp\left(-\frac{x^2}{2}\right)\text{)}$$

where $\bar{N}_e$ and d_{min} are respectively the number of nearest neighbors and minimum distance of separation of the underlying constellation.

Since u is a χ^2 variable with $2n_r$ degrees of freedom, the above average upper-bound is given by

$$\bar{P} \leq \bar{N}_e \left(\frac{1}{1+\rho d_{min}^2/4}\right)^{n_r}$$
$$\overset{\rho\nearrow}{\leq} \bar{N}_e \left(\frac{\rho d_{min}^2}{4}\right)^{-n_r}.$$

The diversity gain g_d^o is equal to the number of receive branches in i.i.d. Rayleigh channels.

Receive Diversity via Gain Combining

- *Minimum Mean Square Error Combining*
  - So far noise was white Gaussian. When the noise (and interference) is colored, MRC is not optimal anymore.
  - Let us denote the combined noise plus interference signal as $\mathbf{n}_i$ such that $\mathbf{y} = \sqrt{E_s}\mathbf{h}c + \mathbf{n}_i$.
  - An optimal gain combining technique is the minimum mean square error (MMSE) combining, where the weights are chosen in order to minimize the mean square error between the transmitted symbol c and the combiner output z, i.e.,
  $$\mathbf{g}^{\star} = \arg\min_{\mathbf{g}} \mathcal{E}\left\{ |\mathbf{g}\mathbf{y} - c|^2 \right\}.$$
  - The optimal weight vector $\mathbf{g}^{\star}$ is given by
  $$\mathbf{g}^{\star} = \mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1},$$
  where $\mathbf{R}_{\mathbf{n}_i} = \mathcal{E}\left\{\mathbf{n}_i \mathbf{n}_i^H\right\}$ is the correlation matrix of the combined noise plus interference signal $\mathbf{n}_i$.
  - Such combiner can be thought of as first whitening the noise plus interference by multiplying $\mathbf{y}$ by $\mathbf{R}_{\mathbf{n}_i}^{-1/2}$ and then match filter the effective channel $\mathbf{R}_{\mathbf{n}_i}^{-1/2}\mathbf{h}$ using $\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-H/2}$.
  - The Signal to Interference plus Noise Ratio (SINR) at the output of the MMSE combiner simply writes as
  $$\rho_{out} = E_s \mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h}.$$
  - In the absence of interference and the presence of white noise, MMSE combiner reduces to MRC filter up to a scaling factor.

Receive Diversity via Gain Combining

Example

Question: Assume a transmission of a signal c from a single antenna transmitter to a multi-antenna receiver through a SIMO channel $\mathbf{h}$. The transmission is subject to the interference from another transmitter sending signal x through the interfering SIMO channel $\mathbf{h}_i$.
The received signal model writes as

$$\mathbf{y} = \mathbf{h}c + \mathbf{h}_i x + \mathbf{n}$$

where $\mathbf{n}$ is the zero mean complex additive white Gaussian noise (AWGN) vector with $\mathcal{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_{n_r}$.
We apply a combiner $\mathbf{g}$ at the receiver to obtain the observation $z = \mathbf{g}\mathbf{y}$.
Derive the expression of the MMSE combiner and the SINR at the output of the combiner.

Receive Diversity via Gain Combining

Example

Answer: The MMSE combiner $\mathbf{g}$ is given by

$$\mathbf{g} = \mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1}$$

where $\mathbf{R}_{\mathbf{n}_i} = \mathcal{E}\left\{\mathbf{n}_i \mathbf{n}_i^H\right\}$ with $\mathbf{n}_i = \mathbf{h}_i x + \mathbf{n}$.
Hence $\mathbf{R}_{\mathbf{n}_i} = \mathbf{h}_i P_x \mathbf{h}_i^H + \sigma_n^2 \mathbf{I}_{n_r}$ with $P_x = \mathcal{E}\{\,|x|^2\,\}$, the power of the interfering signal.
Hence,

$$\mathbf{g} = \mathbf{h}^H \left(\mathbf{h}_i P_x \mathbf{h}_i^H + \sigma_n^2 \mathbf{I}_{n_r}\right)^{-1}.$$

At the receiver, we obtain

$$z = \mathbf{g}\mathbf{y} = \mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h} c + \mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{n}_i.$$

Receive Diversity via Gain Combining

Example

Answer: The output SINR writes

$$
\begin{aligned}
\rho_{out} &= \frac{\left|\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h}\right|^2 P_c}{\mathcal{E}\left\{\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{n}_i \left(\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{n}_i\right)^H\right\}} \\
&= \frac{\left|\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h}\right|^2 P_c}{\mathcal{E}\left\{\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{n}_i \mathbf{n}_i^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h}\right\}} \\
&= \frac{\left|\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h}\right|^2 P_c}{\mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h}} \\
&= \mathbf{h}^H \mathbf{R}_{\mathbf{n}_i}^{-1} \mathbf{h} P_c \\
&= P_c \mathbf{h}^H \left(\mathbf{h}_i P_x \mathbf{h}_i^H + \sigma_n^2 \mathbf{I}_{n_r}\right)^{-1} \mathbf{h} \\
&= \mathrm{SNR}\, \mathbf{h}^H \left(\mathrm{INR}\, \mathbf{h}_i \mathbf{h}_i^H + \mathbf{I}_{n_r}\right)^{-1} \mathbf{h}
\end{aligned}
$$

with $P_c = \mathcal{E}\{\,|c|^2\,\}$, SNR $= P_c/\sigma_n^2$ (the average SNR), INR $= P_x/\sigma_n^2$ (the average INR - Interference to Noise Ratio). □

MISO Systems

- MISO systems exploit diversity at the transmitter through the use of n_t transmit antennas in combination with pre-processing or precoding.

- A significant difference with receive diversity is that the transmitter might not have the knowledge of the MISO channel.
  - At the receiver, the channel is easily estimated.
  - At the transmit side, feedback from the receiver is required to inform the transmitter.

- There are basically two different ways of achieving *direct transmit diversity*:
  - when Tx has a *perfect channel knowledge*, beamforming can be performed to achieve both diversity and array gains,
  - when Tx has a *partial or no channel knowledge of the channel*, space-time coding is used to achieve a diversity gain (but no array gain in the absence of any channel knowledge).

- *Indirect transmit diversity* techniques convert spatial diversity to time or frequency diversity.

Transmit Diversity via Matched Beamforming

- The actual transmitted signal is a vector $\mathbf{c}'$ that results from the multiplication of the signal c by a weight vector $\mathbf{w}$.
- At the receiver, the signal reads as

$$y = \sqrt{E_s}\mathbf{h}\mathbf{c}' + n = \sqrt{E_s}\mathbf{h}\mathbf{w}c + n,$$

  where $\mathbf{h} \triangleq [h_1, \ldots, h_{n_t}]$ represents the MISO channel vector, and $\mathbf{w}$ is also known as the precoder.
- The choice that maximizes the receive SNR is given by

$$\mathbf{w} = \frac{\mathbf{h}^H}{\|\mathbf{h}\|}.$$

- Transmit along the direction of the matched channel, hence it is also known as *matched beamforming* or *transmit MRC*.
- The array gain is equal to the number of transmit antennas, i.e. $\bar{\rho}_{out} = n_t \rho$.
- The diversity gain equal to n_t as the symbol error rate is upper-bounded at high SNR by

$$\bar{P} \le \bar{N}_e \left(\frac{\rho d_{min}^2}{4}\right)^{-n_t}.$$

- Matched beamforming presents the same performance as receive MRC, but *requires a perfect transmit channel knowledge*.

Transmit Diversity via Space-Time Coding

- *Alamouti scheme* is an ingenious transmit diversity scheme for two transmit antennas which does not require transmit channel knowledge.
  - Assume that the flat fading channel remains constant over the two successive symbol periods, and is denoted by $\mathbf{h} = [h_1 \ h_2]$.
  - Two symbols c_1 and c_2 are transmitted simultaneously from antennas 1 and 2 during the first symbol period, followed by symbols $-c_2^*$ and c_1^*, transmitted from antennas 1 and 2 during the next symbol period:

$$y_1 = \sqrt{E_s} h_1 \frac{c_1}{\sqrt{2}} + \sqrt{E_s} h_2 \frac{c_2}{\sqrt{2}} + n_1, \qquad \text{(first symbol period)}$$

$$y_2 = -\sqrt{E_s} h_1 \frac{c_2^*}{\sqrt{2}} + \sqrt{E_s} h_2 \frac{c_1^*}{\sqrt{2}} + n_2. \qquad \text{(second symbol period)}$$

    The two symbols are spread over two antennas and over two symbol periods.
  - Equivalently

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{E_s} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\mathbf{H}_{eff}} \underbrace{\begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix}}_{\mathbf{c}} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}.$$

  - Applying the matched filter $\mathbf{H}_{eff}^H$ to the received vector $\mathbf{y}$ effectively decouples the transmitted symbols as shown below

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{H}_{eff}^H \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{E_s} \left[|h_1|^2 + |h_2|^2 \right] \mathbf{I}_2 \begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix} + \mathbf{H}_{eff}^H \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}$$

Transmit Diversity via Space-Time Coding

– The mean output SNR (averaged over the channel statistics) is thus equal to

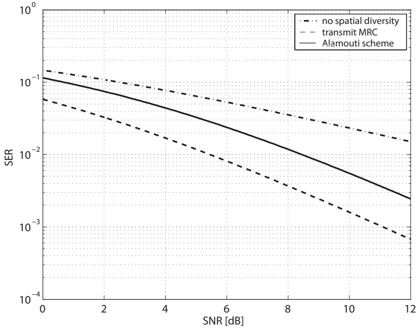
$$\bar{\rho}_{out} = \frac{E_s}{\sigma_n^2}\,\mathcal{E}\left\{\frac{\left[\,\|\mathbf{h}\|^2\,\right]^2}{2\,\|\mathbf{h}\|^2}\right\} = \rho.$$

No array gain owing to the lack of transmit channel knowledge.

– The average symbol error rate at high SNR can be upper-bounded according to

$$\bar{P} \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{8}\right)^{-2}.$$

The diversity gain is equal to $n_t = 2$ despite the lack of transmit channel knowledge.



Transmit MRC vs. Alamouti with 2 transmit antennas in i.i.d. Rayleigh fading channels (BPSK).

Observations:

– At high SNR, any increase in the SNR by 10dB leads to a decrease of SER by 10^{-n} for diversity order n.

- Alamouti, transmit MRC: 2
- No spatial diversity: 1

– Transmit MRC has 3 dB gain over Alamouti

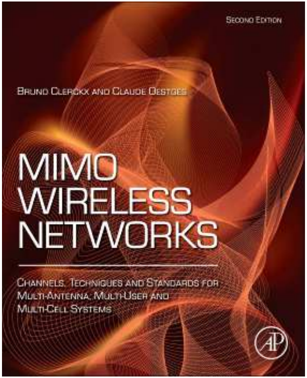
Indirect Transmit Diversity

- It is also possible to convert spatial diversity to time or frequency diversity, which are then exploited using well-known SISO techniques.
- Assume that $n_t = 2$ and that the signal on the second transmit branch is
  - either delayed by one symbol period: the spatial diversity is converted into frequency diversity (delay diversity)
  - either phase-rotated: the spatial diversity is converted into time diversity
  - The effective SISO channel resulting from the addition of the two branches seen by the receiver now fades over frequency or time. This selective fading can be exploited by conventional diversity techniques, e.g. FEC/interleaving.

MIMO Systems - Transmission

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Chapter 1
    - Section: 1.2.4, 1.3.2, 1.6

Introduction - Previous Lectures

- Discrete Time Representation
  - SISO: $y = \sqrt{E_s} hc + n$
  - SIMO: $\mathbf{y} = \sqrt{E_s}\mathbf{h}c + \mathbf{n}$
  - MISO (with perfect CSIT): $y = \sqrt{E_s}\mathbf{hw}c + n$
- h is fading
  - amplitude Rayleigh distributed
  - phase uniformly distributed
- Diversity
  - Diversity gain: $g_d^o(\rho) \triangleq -\frac{\log_2(\bar{P})}{\log_2(\rho)}$
  - Array gain: $g_a \triangleq \frac{\bar{\rho}_{out}}{\bar{\rho}} = \frac{\bar{\rho}_{out}}{\rho}$
- SIMO
  - selection combining
  - gain combining
- MISO
  - with perfect channel knowledge at Tx: Matched Beamforming
  - without channel knowledge at Tx: Space-Time Coding (Alamouti Scheme), indirect (time, frequency) transmit diversity

MIMO Systems

- In MIMO systems, the fading channel between each transmit-receive antenna pair can be modeled as a SISO channel.
- For uni-polarized antennas and small inter-element spacings (of the order of the wavelength), path loss and shadowing of all SISO channels are identical.
- Stacking all inputs and outputs in vectors $\mathbf{c}_k = [c_{1,k}, \ldots, c_{n_t,k}]^T$ and $\mathbf{y}_k = [y_{1,k}, \ldots, y_{n_r,k}]^T$, the input-output relationship at any given time instant k reads as

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}'_k + \mathbf{n}_k,$$

  where
  - $\mathbf{c}'_k$ is a precoded version of $\mathbf{c}_k$ that depends on the channel knowledge at the Tx.
  - $\mathbf{H}_k$ is defined as the $n_r \times n_t$ MIMO channel matrix, $\mathbf{H}_k(n,m) = h_{nm,k}$, with h_{nm} denoting the narrowband channel between transmit antenna m $(m = 1, \ldots, n_t)$ and receive antenna n $(n = 1, \ldots, n_r)$,
  - $\mathbf{n}_k = [n_{1,k}, \ldots, n_{n_r,k}]^T$ is the sampled noise vector, containing the noise contribution at each receive antenna, such that the noise is white in both time and spatial dimensions, $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_l^H\} = \sigma_n^2\mathbf{I}_{n_r}\delta(k-l)$.
- Using the same channels normalization as for SISO channels, $\mathcal{E}\{\,\|\mathbf{H}\|_F^2\,\} = n_t n_r$.
- when Tx has a *perfect channel knowledge*: (dominant and multiple) eigenmode transmission
- when Tx has *no knowledge of the channel*: space-time coding (with $\mathbf{c}'_k = \mathbf{c}_k$)

Space-Time Coding

- MIMO without Transmit Channel Knowledge
- Array/diversity/coding gains are exploitable in SIMO, MISO and ... MIMO
- *Alamouti scheme* can easily be applied to 2×2 MIMO channels

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

- Received signal vector (make sure the channel remains constant over two symbol periods!)

$$\mathbf{y}_1 = \sqrt{E_s}\mathbf{H}\begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix} + \mathbf{n}_1, \qquad \text{(first symbol period)}$$

$$\mathbf{y}_2 = \sqrt{E_s}\mathbf{H}\begin{bmatrix} -c_2^*/\sqrt{2} \\ c_1^*/\sqrt{2} \end{bmatrix} + \mathbf{n}_2. \qquad \text{(second symbol period)}$$

- Equivalently

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2^* \end{bmatrix} = \sqrt{E_s}\underbrace{\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{12}^* & -h_{11}^* \\ h_{22}^* & -h_{21}^* \end{bmatrix}}_{\mathbf{H}_{eff}}\underbrace{\begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix}}_{\mathbf{c}} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2^* \end{bmatrix}.$$

Space-Time Coding

- Apply the matched filter $\mathbf{H}_{eff}^H$ to $\mathbf{y}$ ($\mathbf{H}_{eff}^H \mathbf{H}_{eff} = \|\mathbf{H}\|_F^2 \mathbf{I}_2$)

$$\mathbf{z} = \left[\begin{array}{c} z_1 \\ z_2 \end{array} \right] = \sqrt{E_s} \mathbf{H}_{eff}^H \mathbf{y} = \sqrt{E_s} \|\mathbf{H}\|_F^2 \mathbf{I}_2 \, \mathbf{c} + \mathbf{n}'$$

where $\mathbf{n}'$ is such that $\mathcal{E}\{\mathbf{n}'\} = \mathbf{0}_{2\times 1}$ and $\mathcal{E}\{\mathbf{n}'\mathbf{n}'^H\} = \|\mathbf{H}\|_F^2 \; \sigma_n^2 \mathbf{I}_2$.

- Average output SNR

$$\bar{\rho}_{out} = \frac{E_s}{\sigma_n^2} \mathcal{E}\left\{ \frac{\left[\|\mathbf{H}\|_F^2 \right]^2}{2 \|\mathbf{H}\|_F^2} \right\} = 2\rho,$$

Receive array gain ($g_a = n_r = 2$) but no transmit array gain!

- Average symbol error rate

$$\bar{P} \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{8} \right)^{-4}.$$

Full diversity ($g_d^o = n_t n_r = 4$)

Dominant Eigenmode Transmission

- MIMO with Perfect Transmit Channel Knowledge
- Extension of Matched Beamforming to MIMO

$$\mathbf{y} = \sqrt{E_s}\mathbf{H}\mathbf{c}' + \mathbf{n} = \sqrt{E_s}\mathbf{H}\mathbf{w}c + \mathbf{n},$$
$$z = \mathbf{g}\mathbf{y} = \sqrt{E_s}\mathbf{g}\mathbf{H}\mathbf{w}c + \mathbf{g}\mathbf{n}.$$

- Decompose

$$\mathbf{H} = \mathbf{U}_{\mathbf{H}}\mathbf{\Sigma}_{\mathbf{H}}\mathbf{V}_{\mathbf{H}}^{H},$$
$$\mathbf{\Sigma}_{\mathbf{H}} = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\}.$$

- Received SNR is maximized by matched filtering, leading to

$$\mathbf{w} = \mathbf{v}_{max}$$
$$\mathbf{g} = \mathbf{u}_{max}^{H}$$

  where $\mathbf{v}_{max}$ and $\mathbf{u}_{max}$ are respectively the right and left singular vectors corresponding to the maximum singular value of $\mathbf{H}$, $\sigma_{max} = \max\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\}$. Note the generalization of matched beamforming (MISO) and MRC (SIMO)!
- Equivalent channel: $z = \sqrt{E_s}\sigma_{max}c + \tilde{n}$ where $\tilde{n} = \mathbf{g}\mathbf{n}$ has a variance equal to σ_n^2.

Dominant Eigenmode Transmission

- Array gain: $\mathcal{E}\{\sigma_{max}^2\} = \mathcal{E}\{\lambda_{max}\}$ where λ_{max} is the largest eigenvalue of $\mathbf{H}\mathbf{H}^H$. Commonly, $\max\{n_t, n_r\} \leq g_a \leq n_t n_r$.

Example

Array gain changes depending on the channel properties and distribution

- Line of Sight: $\mathbf{H} = \mathbf{1}_{n_r \times n_t}$. Only one singular value is non-zero and equal to $\sqrt{n_t n_r}$: $g_a = n_t n_r$.
- In the i.i.d. Rayleigh case: for large n_t, n_r, $g_a = \left(\sqrt{n_t} + \sqrt{n_r}\right)^2$.

- Diversity gain: the dominant eigenmode transmission extracts a full diversity gain of $n_t n_r$ in i.i.d. Rayleigh channels.

Dominant Eigenmode Transmission

Example

Question: Show that the optimum (in the sense of SNR maximization) transmit precoder and combiner in dominant eigenmode transmission is given by the dominant right and left singular vector of the channel matrix, respectively.

Answer: Let us write

$$\mathbf{y} = \sqrt{E_s}\mathbf{H}\mathbf{c}' + \mathbf{n} = \sqrt{E_s}\mathbf{H}\mathbf{w}c + \mathbf{n},$$
$$z = \mathbf{g}\mathbf{y} = \sqrt{E_s}\mathbf{g}\mathbf{H}\mathbf{w}c + \mathbf{g}\mathbf{n}.$$

where $\|\mathbf{w}\|^2 = 1$ (power constraint). We decompose

$$\mathbf{H} = \mathbf{U}_{\mathbf{H}}\mathbf{\Sigma}_{\mathbf{H}}\mathbf{V}_{\mathbf{H}}^H, \qquad \mathbf{\Sigma}_{\mathbf{H}} = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\}.$$

In order to maximize the SNR, we choose $\mathbf{g}$ as a matched filter, i.e. $\mathbf{g} = (\mathbf{H}\mathbf{w})^H$ such that

$$\mathbf{g}\mathbf{H}\mathbf{w} = \mathbf{w}^H\mathbf{H}^H\mathbf{H}\mathbf{w} = \mathbf{w}^H\mathbf{V}_{\mathbf{H}}\mathbf{\Sigma}_{\mathbf{H}}^2\mathbf{V}_{\mathbf{H}}^H\mathbf{w} = \sum_{i=1}^{r(\mathbf{H})} \sigma_i^2 \left|\mathbf{v}_i^H\mathbf{w}\right|^2 \leq \sigma_{max}^2$$

where $\mathbf{v}_i$ is the i column of $\mathbf{V}_{\mathbf{H}}$ and $\sigma_{max} = \max_{i=1,\ldots,r(\mathbf{H})} \sigma_i$.

Dominant Eigenmode Transmission

Example

Answer: The inequality is replaced by an equality if $\mathbf{w} = \mathbf{v}_{max}$. By choosing $\mathbf{w} = \mathbf{v}_{max}$,

$$\begin{aligned}\mathbf{g} = \mathbf{w}^H \mathbf{H}^H &= \mathbf{v}_{max}^H \mathbf{V}_\mathbf{H} \mathbf{\Sigma}_\mathbf{H} \mathbf{U}_\mathbf{H}^H \\ &= \sigma_{max} \mathbf{u}_{max}^H\end{aligned}$$

where $\mathbf{u}_{max}$ is the column of $\mathbf{U}_\mathbf{H}$ corresponding to the dominant singular value σ_{max} of $\mathbf{H}$. If we normalize $\mathbf{g}$ such that $\|\mathbf{g}\|^2 = 1$, we can write $\mathbf{g} = \mathbf{u}_{max}$. □

Multiple Eigenmode Transmission

- Assume $n_r \geq n_t$ an that $r(\mathbf{H}) = n_t$, i.e. n_t singular values in $\mathbf{H}$. Hence, what about spreading symbols over all non-zero eigenmodes of the channel:
  - Tx side: multiply the input vector $\mathbf{c}$ $(n_t \times 1)$ using $\mathbf{V_H}$, i.e. $\mathbf{c}' = \mathbf{V_H c}$.
  - Rx side: multiply the received vector $\mathbf{y}$ by $\mathbf{G} = \mathbf{U_H^H}$.
  - Overall,

$$\begin{aligned}\mathbf{z} &= \sqrt{E_s}\mathbf{GHc}' + \mathbf{Gn} \\ &= \sqrt{E_s}\mathbf{U_H^H H V_H c} + \mathbf{U}^H\mathbf{n} \\ &= \sqrt{E_s}\mathbf{\Sigma_H c} + \tilde{\mathbf{n}}.\end{aligned}$$

  The channel has been decomposed into n_t parallel SISO channels given by $\{\sigma_1, \ldots, \sigma_{n_t}\}$.
- The rate achievable in the MIMO channel is the sum of the SISO channel capacities

$$R = \sum_{k=1}^{n_t} \log_2(1 + \rho s_k \sigma_k^2),$$

  where $\{s_1, \ldots, s_{n_t}\}$ is the power allocation on each of the channel eigenmodes.
- The capacity scales linearly in n_t. By contrast, this transmission does not necessarily achieve the full diversity gain of $n_t n_r$ but does at least provide n_r-fold array and diversity gains (still assuming $n_t \leq n_r$).
- In general, the rate scales linearly with the rank of $\mathbf{H}$.

Multiple Eigenmode Transmission

Example

Question: Is the rate achievable in a MIMO channel with multiple eigenmode transmission and uniform power allocation across modes always larger than that achievable with dominant eigenmode transmission?

Answer: No! The achievable rate with multiple eigenmode transmission in the MIMO channel is the sum of the SISO channel achievable rates

$$R = \sum_{k=1}^{r(\mathbf{H})} \log_2(1 + \rho s_k \sigma_k^2),$$

where $\{s_1, \ldots, s_{r(\mathbf{H})}\}$ is the power allocation on each of the channel eigenmodes.

Two strategies (for a total power constraint $\sum_{k=1}^{r(\mathbf{H})} s_k = 1$):

- Uniform power allocation: $R_u = \sum_{k=1}^{r(\mathbf{H})} \log_2(1 + \rho 1/r(\mathbf{H})\sigma_k^2)$
- Dominant eigenmode transmission: $R_d = \log_2(1 + \rho\sigma_{max}^2)$

R_u could be either greater or smaller than R_d. For instance, if $\sigma_1 >> 0$ and $\sigma_k \approx \epsilon$ for $k > 1$, $R_u \approx \log_2(1 + \rho\sigma_1^2/r(\mathbf{H})) \leq R_d$ for small values of ρ. At very high SNR, despite the little contributions of $\sigma_k \approx \epsilon$, R_u will become higher than R_d. □

Multiplexing gain

- Array/diversity/coding gains are exploitable in SIMO, MISO and MIMO but MIMO can offer much more than MISO and SIMO.

- MIMO channels offer *multiplexing gain*: measure of the number of independent streams that can be transmitted in parallel in the MIMO channel. Defined as

$$g_s \triangleq \lim_{\rho \to \infty} \frac{R(\rho)}{\log_2(\rho)}$$

  where $R(\rho)$ is the transmission rate.

- The multiplexing gain is the pre-log factor of the rate at high SNR, i.e.

$$R \approx g_s \log_2(\rho)$$

- Modeling only the individual SISO channels from one Tx antenna to one Rx antenna not enough:
  - MIMO performance depends on the channel matrix properties
  - characterize all statistical correlations between all matrix elements necessary!

Interference Management

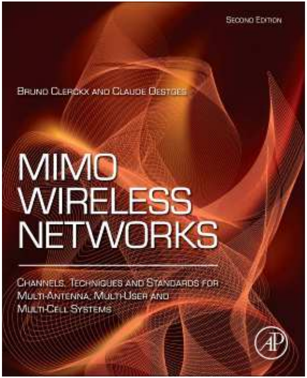
- In wireless networks, co-channel interference is caused by the necessary frequency re-use.

- With multiple antennas, it is possible to exploit the difference between the spatial signatures of the desired vs. the interfering channels to reduce the intra-cell and inter-cell interference:
  - In a single-cell multi-user context, Multi-user MIMO (MU-MIMO)
  - In a multi-cell context, Multi-Cell MIMO (MC-MIMO)

Channel Modelling

Reference Book

- Bruno Clerckx and Claude Oestges, “MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems,” Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 2
    - Section: 2.1.1, 2.1.2, 2.1.3, 2.1.5, 2.2, 2.3.1
- Chapter 3
    - Section: 3.2.1, 3.2.2, 3.4.1

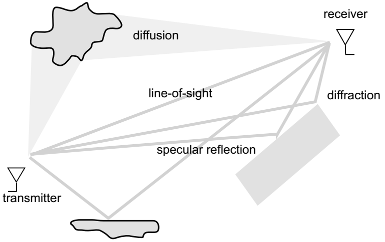
Double-Directional Channel Modeling

- Space comes as an additional dimension
  - *directional*: model the angular distribution of the energy at the antennas
  - *double*: there are multiple antennas at transmit and receive sides
- Neglecting path-loss and shadowing, the time-variant double-directional channel

$$h\left(t, \mathbf{p}_t, \mathbf{p}_r, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) = \sum_{k=0}^{n_s-1} h_k\left(t, \mathbf{p}_t, \mathbf{p}_r, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right),$$

  - $\mathbf{p}_t$, $\mathbf{p}_r$: location of Tx and Rx, respectively
  - n_s contributions
  - time t: variation with time (with the motion of the receiver)
  - delay τ: each contribution arrives with a delay proportional to its path length
  - $\mathbf{\Omega}_t$, $\mathbf{\Omega}_r$: direction of departure (DoD), directions of arrival (DoA). In spherical coordinates (i.e., the azimuth Θ_t and elevation ψ_t) on a sphere of unit radius

$$\mathbf{\Omega}_t = [\cos\Theta_t \sin\psi_t, \sin\Theta_t \sin\psi_t, \cos\psi_t]^T$$



Double-Directional Channel Modeling

- In the case of a plane wave, and considering a fixed transmitter and a mobile receiver,

$$h_k\left(t, \mathbf{p}_t, \mathbf{p}_r, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\right) \triangleq \alpha_k\, e^{j\phi_k}\, e^{-j\Delta\omega_k t}\, \delta(\tau - \tau_k)\, \delta(\boldsymbol{\Omega}_t - \boldsymbol{\Omega}_{t,k})\, \delta(\boldsymbol{\Omega}_r - \boldsymbol{\Omega}_{r,k}),$$

  where
  - α_k is the amplitude of the k^{th} contribution,
  - ϕ_k is the phase of the k^{th} contribution,
  - $\Delta\omega_k$ is the Doppler shift of the k^{th} contribution,
  - τ_k is the time delay of the k^{th} contribution,
  - $\Omega_{t,k}$ is the DoD of the k^{th} contribution,
  - $\Omega_{r,k}$ is the DoA of the k^{th} contribution.
- A more compact notation (all temporal variations are grouped into t)

$$h\left(t, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\right) = \sum_{k=0}^{n_s-1} h_k\left(t, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\right)$$

- Impulse response of the channel (as in Lecture 1, without path loss/shadowing)

$$h(t, \tau) = \iint h\big(t, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\big)\, d\boldsymbol{\Omega}_t\, d\boldsymbol{\Omega}_r$$

- Narrowband transmission (the channel is not frequency selective)

$$h(t) = \iiint h\big(t, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\big)\, d\tau\, d\boldsymbol{\Omega}_t\, d\boldsymbol{\Omega}_r$$

Wide-Sense Stationary Uncorrelated Scattering Homogeneous

- Assumption: Wide-Sense Stationary Uncorrelated Scattering Homogeneous (WSSUSH) channels
- Wide-Sense Stationary:
  - Time correlations only depend on the time difference
  - Signals arriving with different Doppler frequencies are uncorrelated
- Uncorrelated Scattering:
  - Frequency correlations only depend on the frequency difference
  - Signals arriving with different delays are uncorrelated
- Homogeneous:
  - Spatial correlation only depends on the spatial difference at both transmit and receive sides
  - Signals departing/arriving with different directions are uncorrelated

Spectra

- Doppler spectrum and coherence time
- Power delay spectrum and delay spread
- Power direction spectrum and angle spread
  - the power-delay joint direction spectrum

$$\mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big) = \mathcal{E}\big\{ \big|h\big(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\big|^2 \big\},$$

  - the joint direction power spectrum

$$\mathcal{A}(\mathbf{\Omega}_t, \mathbf{\Omega}_r) = \int \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau,$$

  - the transmit direction power spectrum

$$\mathcal{A}_t(\mathbf{\Omega}_t) = \int\int \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau\, d\mathbf{\Omega}_r,$$

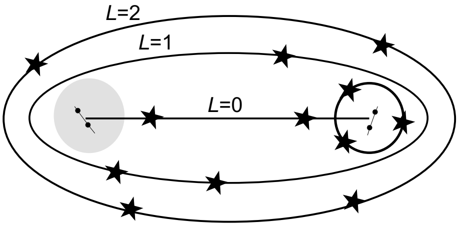
  - the receive direction power spectrum

$$\mathcal{A}_r(\mathbf{\Omega}_r) = \int\int \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau\, d\mathbf{\Omega}_t.$$

Angular Spread

- The channel angle-spreads are defined similarly to the delay-spread
  - delay-spread $\Longleftrightarrow$ channel frequency selectivity
  - angle-spread $\Longleftrightarrow$ channel spatial selectivity

$$
\begin{aligned}
\boldsymbol{\Omega}_{t,M} &= \frac{\int \boldsymbol{\Omega}_t \mathcal{A}_t(\boldsymbol{\Omega}_t)\, d\boldsymbol{\Omega}_t}{\int \mathcal{A}_t(\boldsymbol{\Omega}_t)\, d\boldsymbol{\Omega}_t} \\
\Omega_{t,RMS} &= \sqrt{\frac{\int \left\|\boldsymbol{\Omega}_t - \boldsymbol{\Omega}_{t,M}\right\|^2 \mathcal{A}_t(\boldsymbol{\Omega}_t)\, d\boldsymbol{\Omega}_t}{\int \mathcal{A}_t(\boldsymbol{\Omega}_t)\, d\boldsymbol{\Omega}_t}}
\end{aligned}
$$



The MIMO Channel Matrix

- Convert the double-directional channel to a $n_r \times n_t$ MIMO channel

$$\mathbf{H}(t,\tau) = \begin{bmatrix} h_{11}(t,\tau) & h_{12}(t,\tau) & \dots & h_{1n_t}(t,\tau) \\ h_{21}(t,\tau) & h_{22}(t,\tau) & \dots & h_{2n_t}(t,\tau) \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r1}(t,\tau) & h_{n_r2}(t,\tau) & \dots & h_{n_rn_t}(t,\tau) \end{bmatrix},$$

  where

$$h_{nm}(t,\tau) \triangleq \iint h_{nm}(t,\tau,\mathbf{\Omega}_t,\mathbf{\Omega}_r)\, d\mathbf{\Omega}_t\, d\mathbf{\Omega}_r$$

- For narrowband (i.e. same delay for all antennas) balanced (i.e. $|h_{nm}| = |h_{11}|$) arrays and plane wave incidence, $h_{nm}(t,\tau,\mathbf{\Omega}_t,\mathbf{\Omega}_r)$ is a phase shifted version of $h_{11}(t,\tau,\mathbf{\Omega}_t,\mathbf{\Omega}_r)$

$$h_{nm}(t,\tau) = \int\int h_{11}(t,\tau,\mathbf{\Omega}_t,\mathbf{\Omega}_r) e^{-j\mathbf{k}_r^T(\mathbf{\Omega}_r)\left[\mathbf{p}_r^{(n)}-\mathbf{p}_r^{(1)}\right]} e^{-j\mathbf{k}_t^T(\mathbf{\Omega}_t)\left[\mathbf{p}_t^{(m)}-\mathbf{p}_t^{(1)}\right]} d\mathbf{\Omega}_t d\mathbf{\Omega}_r$$

  where $\mathbf{k}_t(\mathbf{\Omega}_t)$ and $\mathbf{k}_r(\mathbf{\Omega}_r)$ are the transmit and receive wave propagation 3×1 vectors.

Steering Vectors

- For a transmit ULA oriented broadside to the link axis,
$$e^{-j\mathbf{k}_t^T(\mathbf{\Omega}_t)\cdot\left[\mathbf{p}_t^{(m)}-\mathbf{p}_t^{(1)}\right]} = e^{-j(m-1)\varphi_t(\theta_t)},$$
where $\varphi_t\left(\theta_t\right) = 2\pi(d_t/\lambda)\cos\theta_t$, and $d_t = \left\|\mathbf{p}_t^{(m)} - \mathbf{p}_t^{(m-1)}\right\|$ denotes the inter-element spacing of the transmit array.
  - θ_t is defined relatively to the array orientation (so $\theta_t = \pi/2$ corresponds to the link axis for a broadside array).

- Steering vector (expressed here for a ULA)
  - At the transmitter in the relative direction θ_t:
$$\mathbf{a}_t(\theta_t) = \begin{bmatrix} 1 & e^{-j\varphi_t(\theta_t)} & \ldots & e^{-j(n_t-1)\varphi_t(\theta_t)} \end{bmatrix}^T.$$
  - At the receiver in the relative direction θ_r:
$$\mathbf{a}_r\left(\theta_r\right) = \begin{bmatrix} 1 & e^{-j\varphi_r(\theta_r)} & \ldots & e^{-j(n_r-1)\varphi_r(\theta_r)} \end{bmatrix}^T.$$

- Under the plane wave and balanced narrowband array assumptions, the MIMO channel matrix can be rewritten as a function of steering vectors as
$$\mathbf{H}(t,\tau) = \int\int h\left(t, \mathbf{p}_t^{(1)}, \mathbf{p}_r^{(1)}, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) \mathbf{a}_r(\mathbf{\Omega}_r)\, \mathbf{a}_t^T(\mathbf{\Omega}_t)\, d\mathbf{\Omega}_t\, d\mathbf{\Omega}_r.$$

A Finite Scatterer MIMO Channel Representation

- The transmitter and receiver are coupled via a finite number of scattering paths with $n_{s,t}$ DoDs at the transmitter and $n_{s,r}$ DoAs at the receiver.
  $\longrightarrow$ Replace the integral by a summation (assume for simplicity 2-D azimuthal propagation)

$$\begin{aligned}\mathbf{H}(t,\tau) &= \sum_{l=1}^{n_{s,t}} \sum_{p=1}^{n_{s,r}} h_{11}^{(l,p)}(t,\tau)\mathbf{a}_r(\theta_r^{(p)})\mathbf{a}_t^T(\theta_t^{(l)}) \\ &= \mathbf{A}_r \mathbf{H}_s(t,\tau)\mathbf{A}_t^T\end{aligned}$$

  where
  - $\mathbf{A}_r$ and $\mathbf{A}_t$ represent the $n_r \times n_{s,r}$ and $n_t \times n_{s,t}$ matrices whose columns are the steering vectors related to the directions of each path observed at Rx and Tx
  - $\mathbf{H}_s(t,\tau)$ is a $n_{s,r} \times n_{s,t}$ matrix whose elements are the complex path gains between all DoDs and DoAs at time instant t and delay τ
- Assume the columns of $\mathbf{A}_t$ are written as $\mathbf{a}_t(\theta_t^{(l)})$, $l = 1, ..., n_{s,t}$. Let us write

$$\mathbf{H} = \underbrace{\mathbf{A}_r \mathbf{H}_s}_{\tilde{\mathbf{H}}_s} \mathbf{A}_t^T = \sum_{l=1}^{n_{s,t}} \tilde{\mathbf{H}}_s(:,l)\mathbf{a}_t^T(\theta_t^{(l)}) = \sum_{l=1}^{n_{s,t}} \mathbf{H}^{(l)},$$

  where $\mathbf{H}^{(l)}$ can be viewed as the channel matrix corresponding to the l^{th} scatterer located in the direction of departure $\theta_t^{(l)}$.

Statistical Properties of the MIMO Channel Matrix

- Assume narrowband channels, the spatial correlation matrix of the MIMO channel

$$\mathbf{R} = \mathcal{E}\{\text{vec}(\mathbf{H}^H)\text{vec}(\mathbf{H}^H)^H\}$$

  This is a $n_t n_r \times n_t n_r$ positive semi-definite Hermitian matrix.
- It describes the correlation between all pairs of transmit-receive channels:
  - $\mathcal{E}\left\{\mathbf{H}(n,m)\mathbf{H}^*(n,m)\right\}$: the average energy of the channel between antenna m and antenna n,
  - $r_m^{(nq)} = \mathcal{E}\left\{\mathbf{H}(n,m)\mathbf{H}^*(q,m)\right\}$: the receive correlation between channels originating from transmit antenna m and impinging upon receive antennas n and q,
  - $t_n^{(mp)} = \mathcal{E}\left\{\mathbf{H}(n,m)\mathbf{H}^*(n,p)\right\}$: the transmit correlation between channels originating from transmit antennas m and p and arriving at receive antenna n,
  - $\mathcal{E}\left\{\mathbf{H}(n,m)\mathbf{H}^*(q,p)\right\}$: the cross-channel correlation between channels (m,n) and (q,p).

Example

2x2 MIMO

$$\mathbf{R} = \begin{bmatrix} 1 & t_1^* & r_1^* & s_1^* \\ t_1 & 1 & s_2^* & r_2^* \\ r_1 & s_2 & 1 & t_2^* \\ s_1 & r_2 & t_2 & 1 \end{bmatrix}$$

$$t_1 = \mathcal{E}\left\{\mathbf{H}(1,1)\mathbf{H}^*(1,2)\right\}$$

$$r_1 = \mathcal{E}\left\{\mathbf{H}(1,1)\mathbf{H}^*(2,1)\right\}$$

Spatial Correlation

- How are these correlations related to the propagation channel?
- Let us consider the case of ULAs and 2-D azimuthal propagation

$$h_{nm}(t) = \int\int h_{11}\left(t, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) e^{-j(m-1)\varphi_t(\theta_t)} e^{-j(n-1)\varphi_r(\theta_t)} \, d\theta_t \, d\theta_r$$

where
  - $\varphi_{r,t}\left(\theta_{r,t}\right) = 2\pi(d_{r,t}/\lambda)\cos\theta_{r,t}$,
  - d_r and d_t are the inter-element spacing at the receive/transmit arrays
  - $h_{11}\left(t, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) \triangleq \int h_{11}\left(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) d\tau$.
- Correlation between channels h_{nm} and h_{qp}

$$\begin{aligned}\mathcal{E}\left\{h_{nm}h_{qp}^*\right\} &= \mathcal{E}\left\{\int_0^{2\pi}\int_0^{2\pi}\left|h_{11}\left(t, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right)\right|^2 e^{-j(m-p)\varphi_t(\theta_t)} e^{-j(n-q)\varphi_r(\theta_t)} \, d\theta_t \, d\theta_r\right\} \\ &= \int_0^{2\pi}\int_0^{2\pi}\mathcal{E}\left\{\left|h_{11}\left(t, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right)\right|^2\right\} e^{-j(m-p)\varphi_t(\theta_t)} e^{-j(n-q)\varphi_r(\theta_t)} \, d\theta_t \, d\theta_r, \\ &= \int_0^{2\pi}\int_0^{2\pi}\mathcal{A}\left(\theta_t, \theta_r\right) e^{-j(m-p)\varphi_t(\theta_t)} e^{-j(n-q)\varphi_r(\theta_t)} \, d\theta_t \, d\theta_r,\end{aligned}$$

where $\mathcal{A}\left(\theta_t, \theta_r\right)$ is the joint direction power spectrum restricted to the azimuth angles.
- The channel correlation is related to both the antenna spacings and the joint direction power spectrum!

Spatial Correlation

- When the energy spreading is very large at both sides and d_t/d_r are sufficiently large, elements of $\mathbf{H}$ become uncorrelated, and $\mathbf{R}$ becomes diagonal.

Example

Consider two transmit antennas spaced by d_t. The transmit correlation writes as

$$t = \int_0^{2\pi} e^{j2\pi(d_t/\lambda)\cos\theta_t} \mathcal{A}_t(\theta_t) d\theta_t,$$

which only depends on the transmit antenna spacing and the transmit direction power spectrum.

- *isotropic scattering*: very rich scattering environment around the transmitter with a uniform distribution of the energy, i.e. $\mathcal{A}_t(\theta_t) \cong 1/2\pi$

$$t = \frac{1}{2\pi} \int_0^{2\pi} e^{j\varphi_t(\theta_t)} d\theta_t = \frac{1}{2\pi} \int_0^{2\pi} e^{j2\pi(d_t/\lambda)\cos\theta_t} d\theta_t$$
$$= \mathrm{J}_0\left(2\pi \frac{d_t}{\lambda}\right).$$

The transmit correlation only depends on the spacing between the two antennas.

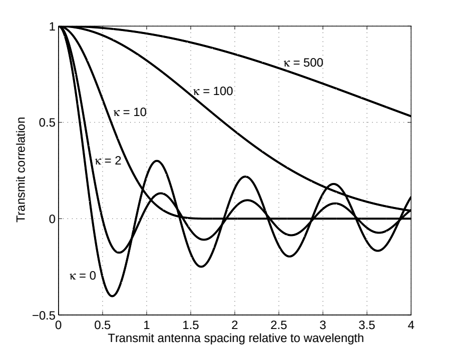
Spatial Correlation

Example

- *highly directional scattering*: scatterers around the transmit array are concentrated along a narrow direction $\theta_{t,0}$, i.e., $\mathcal{A}_t(\theta_t) \rightarrow \delta(\theta_t - \theta_{t,0})$

$$t \rightarrow e^{j\varphi_t(\theta_{t,0})} = e^{j2\pi(d_t/\lambda)\cos\theta_{t,0}}.$$

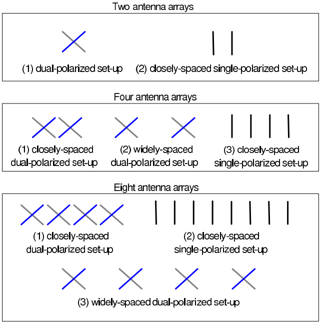
Very high transmit correlation approaching one. The scattering direction is directly related to the phase of the transmit correlation.



- $\mathcal{A}_t(\theta_t)$ in real-world channels: neither uniform nor a delta.
- isotropic scattering ($\kappa = 0$): first minimum for $d_t = 0.38\lambda$
- directional scattering ($\kappa = \infty$): correlation never reaches 0
- in practice, decorrelation in rich scattering is reached for $d_t \approx 0.5\lambda$
- The more directional the azimuthal dispersion (i.e. for κ increasing), the larger the antenna spacing required to obtain a null correlation.

Analytical Representation of Rayleigh MIMO Channels

- Independent and Identically Distributed (I.I.D.) Rayleigh fading
  - $\mathbf{R} = \mathbf{I}_{n_t n_r}$
  - $\mathbf{H} = \mathbf{H}_w$ is a random fading matrix with unit variance and i.i.d. circularly symmetric complex Gaussian entries.
- Realistic in practice only if both conditions are satisfied:
  - the antenna spacings and/or the angle spreads at Tx and Rx are large enough,
  - all individual channels characterized by the same average power (i.e., balanced array).
- What about real-world channels? Sometimes significantly deviate from this ideal channel:
  - *limited angular spread and/or reduced array sizes* cause the channels to become *correlated* (channels are not independent anymore)
  - a *coherent contribution* may induce the channel statistics to become *Ricean* (channels are not Rayleigh distributed anymore),
  - the use of multiple *polarizations* creates gain imbalances between the various elements of the channel matrix (channels are not identically distributed anymore).



Correlated Rayleigh Fading Channels

- For identically distributed Gaussian channels, $\mathbf{R}$ constitutes a sufficient description of the stochastic behavior of the MIMO channel.
- Any channel realization is obtained by

$$\text{vec}\left(\mathbf{H}^H\right) = \mathbf{R}^{1/2}\, \text{vec}(\mathbf{H}_w),$$

  where $\mathbf{H}_w$ is one realization of an i.i.d. channel matrix.
- Complicated to use because
  - cross-channel correlation not intuitive and not easily tractable
  - Too many parameters: dimensions of $\mathbf{R}$ rapidly become large as the array sizes increase
  - vec operation complicated for performance analysis
- Kronecker model: use a separability assumption

$$\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t,$$

$$\mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$$

  where $\mathbf{R}_t$ and $\mathbf{R}_r$ are respectively the transmit and receive correlation matrices.
- Strictly valid only if $r_1 = r_2 = r$ and $t_1 = t_2 = t$ and $s_1 = rt$ and $s_2 = rt^*$ (for 2×2)

$$\mathbf{R} = \begin{bmatrix} 1 & t_1^* & r_1^* & s_1^* \\ t_1 & 1 & s_2^* & r_2^* \\ r_1 & s_2 & 1 & t_2^* \\ s_1 & r_2 & t_2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & t^* & r^* & r^*t^* \\ t & 1 & r^*t & r^* \\ r & rt^* & 1 & t^* \\ rt & r & t & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & r^* \\ r & 1 \end{bmatrix}}_{\mathbf{R}_r} \otimes \underbrace{\begin{bmatrix} 1 & t^* \\ t & 1 \end{bmatrix}}_{\mathbf{R}_t}$$

Correlated Rayleigh Fading Channels

Example

Question: Assume a MISO system with two transmit antennas. The channel gains are identically distributed circularly symmetric complex Gaussian but can be correlated and are denoted as h_1 and h_2. Write the expression of the transmit correlation matrix $\mathbf{R}_t$ and derive the eigenvalues and eigenvectors of $\mathbf{R}_t$ as a fonction of the transmit correlation coefficient t.

Answer: We write

$$\mathbf{R}_t = \mathcal{E}\left\{\left[\begin{array}{c} h_1^* \\ h_2^* \end{array}\right]\left[\begin{array}{cc} h_1 & h_2 \end{array}\right]\right\} = \left[\begin{array}{cc} \mathcal{E}\left\{\left|h_1\right|^2\right\} & \mathcal{E}\left\{h_1^* h_2\right\} \\ \mathcal{E}\left\{h_1 h_2^*\right\} & \mathcal{E}\left\{\left|h_2\right|^2\right\} \end{array}\right] = \left[\begin{array}{cc} 1 & t^* \\ t & 1 \end{array}\right]$$

where $t = \mathcal{E}\left\{h_1 h_2^*\right\}$ is the transmit correlation coefficient. The SVD leads to

$$\mathbf{R}_t = \left[\begin{array}{cc} 1 & 1 \\ t/|t| & -t/|t| \end{array}\right]\left[\begin{array}{cc} 1+|t| & 0 \\ 0 & 1-|t| \end{array}\right]\left[\begin{array}{cc} 1 & 1 \\ t/|t| & -t/|t| \end{array}\right]^H.$$

The eigenvalues are only function of the magnitude of t while the eigenvectors are only function of the phase of t.

□

Correlated Rayleigh Fading Channels

Example

Question: Assume the previous example with $|t| \to 1$. Compute the weights of the matched beamformer (or maximum ratio transmission/transmit MRC).
Answer: With matched beamforming, $\mathbf{w} = \mathbf{h}^H / \|\mathbf{h}\|$ where

$$\begin{aligned}\mathbf{h} &= \mathbf{h}_w \mathbf{R}_t^{1/2} \\ &= \mathbf{h}_w \begin{bmatrix} 1 & 1 \\ t/|t| & -t/|t| \end{bmatrix} \begin{bmatrix} \sqrt{1+|t|} & 0 \\ 0 & \sqrt{1-|t|} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ t/|t| & -t/|t| \end{bmatrix}^H \\ &= 2\mathbf{h}_w \begin{bmatrix} 1 \\ t/|t| \end{bmatrix} \begin{bmatrix} 1 \\ t/|t| \end{bmatrix}^H\end{aligned}$$

where the last equality comes from the fact that $|t| = 1$. This shows that for high correlation, the channel direction $(\mathbf{h}/\|\mathbf{h}\|)$ is aligned with $\begin{bmatrix} 1 & t^*/|t| \end{bmatrix}$. Hence

$$\mathbf{w} = \mathbf{h}^H / \|\mathbf{h}\| = \begin{bmatrix} 1 \\ t/|t| \end{bmatrix}.$$

Transmission is performed in the direction where all scatterers are located. □

Analytical Representation of Ricean MIMO Channels

- In the presence of a strong coherent component which does not experience any fading over time
  - e.g. a line-of-sight field, one or several specular contributions, coherent addition of reflected and diffracted contributions (in fixed wireless access only).
- All these situations lead to a Ricean distribution of the received field amplitude.
  - The relative strength of the dominant coherent component is characterized by the K-factor K. As the channel contains a coherent component, its amplitude can be written as

$$|h(t)| = \left|\bar{h} + \tilde{h}(t)\right|,$$

    where $\bar{h}$ is the coherent component, and $\tilde{h}(t)$ is the non coherent part, whose energy is denoted as $2\sigma_s^2$.
  - The K-factor is defined as

$$K = \frac{\left|\bar{h}\right|^2}{2\sigma_s^2}.$$

  - $|h(t)| \triangleq s'$ is Ricean distributed, and its distribution is given, as a function of K, as

$$p_{s'}(s') = \frac{2s'K}{\left|\bar{h}\right|^2} \exp\left[-K\left(\frac{s'^2}{\left|\bar{h}\right|^2}+1\right)\right] \mathrm{I}_0\left(\frac{2s'K}{\left|\bar{h}\right|}\right).$$

  - For $K = 0$, the Ricean distribution boils down to the Rayleigh distribution while for $K = \infty$, the channel becomes deterministic (no fading).

Ricean MIMO Channels

- Common Ricean MIMO channel model

$$\mathbf{H} = \sqrt{\frac{K}{1+K}}\,\bar{\mathbf{H}} + \sqrt{\frac{1}{1+K}}\,\tilde{\mathbf{H}}$$

- The matrix $\tilde{\mathbf{H}}$ relates to the Rayleigh component (non-coherent part). It can be modeled and characterized using

$$\mathbf{R} = \mathcal{E}\{\mathrm{vec}(\tilde{\mathbf{H}}^H)\mathrm{vec}(\tilde{\mathbf{H}}^H)^H\}$$

- The matrix $\bar{\mathbf{H}}$ corresponding to the coherent component(s) has fixed phase-shift-only entries (strongly related to the array configuration and orientation)

$$\bar{\mathbf{H}} = \begin{bmatrix} e^{j\alpha_{11}} & e^{j\alpha_{12}} \\ e^{j\alpha_{21}} & e^{j\alpha_{22}} \end{bmatrix}$$

  - With only one coherent contribution with given DoD and DoA ($\mathbf{\Omega}_{t,c}$ and $\mathbf{\Omega}_{r,c}$),

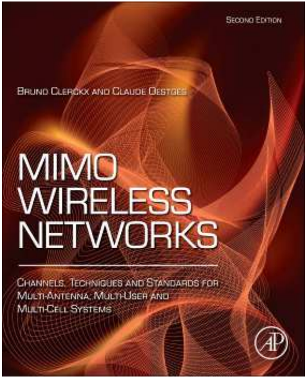
$$\bar{\mathbf{H}} = \mathbf{a}_r(\mathbf{\Omega}_{r,c})\,\mathbf{a}_t^T(\mathbf{\Omega}_{t,c})$$

  - For broadside arrays with a pure line-of-sight component, $\bar{\mathbf{H}} = \mathbf{1}_{n_r \times n_t}$.

Capacity of point-to-point MIMO Channels

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 5
  - Section: 5.1, 5.2, 5.3, 5.4.1, 5.4.2 (except "Antenna Selection Schemes"), 5.5.1 - "Kronecker Correlated Rayleigh Channels", 5.5.2, 5.7, 5.8.1 (except Proof of Proposition 5.9 and Example 5.4)

Introduction - Previous Lectures

- Transmission strategies
  - Space-Time Coding when no Tx channel knowledge
  - Multiple (including dominant) eigenmode transmission when Tx channel knowledge

$$
\begin{aligned}
\mathbf{z} &= \sqrt{E_s}\mathbf{G}\mathbf{H}\mathbf{c}' + \mathbf{G}\mathbf{n} \\
&= \sqrt{E_s}\mathbf{U}_{\mathbf{H}}^H\mathbf{H}\mathbf{V}_{\mathbf{H}}\mathbf{c} + \mathbf{U}^H\mathbf{n} \\
&= \sqrt{E_s}\mathbf{\Sigma}_{\mathbf{H}}\mathbf{c} + \tilde{\mathbf{n}}.
\end{aligned}
$$

  Multiple parallel data pipes $\rightarrow$ Spatial multiplexing gain!

- Performance highly depends on the channel matrix properties
  - Angle spread and inter-element spacing
  - Spatial Correlation: spread antennas far apart to decrease spatial correlation
  - Rayleigh and Ricean distribution

System Model

- A single-user MIMO system with n_t transmit and n_r receive antennas over a frequency flat-fading channel.
- The transmit and received signals in a MIMO channel are related by

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}'_k + \mathbf{n}_k$$

  where

  - $\mathbf{y}_k$ is the $n_r \times 1$ received signal vector,
  - $\mathbf{H}_k$ is the $n_r \times n_t$ channel matrix
  - $\mathbf{n}_k$ is a $n_r \times 1$ zero mean complex additive white Gaussian noise (AWGN) vector with $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_l^H\} = \sigma_n^2\mathbf{I}_{n_r}\delta(k-l)$.
  - $\rho = E_s/\sigma_n^2$ represents the SNR.

- The input covariance matrix is defined as the covariance matrix of the transmit signal $\mathbf{c}'$ (we drop the time index) and writes as $\mathbf{Q} = \mathcal{E}\left\{\mathbf{c}'\mathbf{c}'^H\right\}$.
- Short-term power constraint: Tr$\{\mathbf{Q}\} \leq 1$.
- Long-term power constraint (over a duration $T_p >> T$): $\mathcal{E}\left\{\text{Tr}\{\mathbf{Q}\}\right\} \leq 1$ where the expectation refers here to the averaging over successive codeword of length T.
- Channel time variation: T_{coh} coherence time
  - *slow fading*: T_{coh} is so long that coding is performed over a single channel realization.
  - *fast fading*: T_{coh} is so short that coding over multiple channel realizations is possible.

Capacity of Deterministic MIMO Channels

Proposition

For a deterministic MIMO channel $\mathbf{H}$, *the mutual information* $\mathcal{I}$ *is written as*

$$\mathcal{I}(\mathbf{H},\mathbf{Q}) = \log_2 \det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}\mathbf{Q}\mathbf{H}^H\right]$$

where $\mathbf{Q}$ *is the input covariance matrix whose trace is normalized to unity.*

Definition

The capacity of a deterministic $n_r \times n_t$ MIMO channel with perfect channel state information at the transmitter is

$$C\left(\mathbf{H}\right) = \max_{\mathbf{Q}\geq 0:\mathrm{Tr}\{\mathbf{Q}\}=1} \log_2 \det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}\mathbf{Q}\mathbf{H}^H\right].$$

Note the difference with SISO capacity.

Proof: Denoting the entropy by $H(.)$, the mutual information between input and output is given by

$$\begin{aligned}\mathcal{I}(\mathbf{H},\mathbf{Q}) &= I(\mathbf{c}';\mathbf{y}\,|\mathbf{H}\,),\\ &= H(\mathbf{y}\,|\mathbf{H}\,) - H(\mathbf{y}\,\big|\mathbf{c}',\mathbf{H}\,)\\ &= H(\mathbf{y}\,|\mathbf{H}\,) - H(\mathbf{n}\,\big|\mathbf{c}',\mathbf{H}\,).\end{aligned}$$

When the input vector has a covariance $\mathbf{Q} = \mathcal{E}\{\mathbf{c}'\mathbf{c}'^H\}$, we have that the covariance of $\mathbf{y}$ is given by

$$\mathcal{E}\{\mathbf{y}\mathbf{y}^H\} = \sigma_n^2\mathbf{I}_{n_r} + E_s\mathbf{H}\mathbf{Q}\mathbf{H}^H,$$

since the noise is an additive white Gaussian noise (AWGN). Following similar steps as in SISO, $H(\mathbf{y}\,|\mathbf{H}\,)$ is largest when $\mathbf{y}$ is zero-mean circularly symmetric complex Gaussian, which is achieved when $\mathbf{c}'$ is zero-mean circularly symmetric complex Gaussian. Because the differential entropy $H(\mathbf{c}')$ of a zero-mean circularly symmetric complex Gaussian input vector $\mathbf{c}'$ with covariance matrix $\mathbf{Q}$ is given by $\log_2 \det(\pi e\mathbf{Q})$, we get

$$\begin{aligned}\mathcal{I}(\mathbf{H},\mathbf{Q}) &= \log_2\det\left(\pi e\left[\sigma_n^2 I_{n_r} + E_s\mathbf{H}\mathbf{Q}\mathbf{H}^H\right]\right) - \log_2\det(\pi e\sigma_n^2\mathbf{I}_{n_r}),\\ &= \log_2\det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}\mathbf{Q}\mathbf{H}^H\right].\end{aligned}$$

Capacity and Water-Filling Algorithm

- What is the best transmission strategy, i.e. the optimum input covariance matrix $\mathbf{Q}$?
- *First*, create $n = \min\{n_t, n_r\}$ parallel data pipes (Multiple Eigenmode Transmission)
  - Decouple the channel along the individual channel modes (in the directions of the singular vectors of the channel matrix $\mathbf{H}$ at both the transmitter and the receiver)

$$\mathbf{H} = \mathbf{U_H}\mathbf{\Sigma_H}\mathbf{V_H}^H,$$

$$\mathbf{U_H}^H\mathbf{H}\mathbf{V_H} = \mathbf{U_H}^H\mathbf{U_H}\mathbf{\Sigma_H}\mathbf{V_H}^H\mathbf{V_H} = \mathbf{\Sigma_H}$$

  - Optimum input covariance matrix $\mathbf{Q}^\star$ writes as

$$\mathbf{Q}^\star = \mathbf{V_H}\mathrm{diag}\left\{s_1^\star, \ldots, s_n^\star\right\}\mathbf{V_H}^H,$$

- *Second*, allocate power to data pipes
  - $\mathbf{\Sigma_H} = \mathrm{diag}\left\{\sigma_1, \ldots, \sigma_n\right\}$, and $\sigma_k^2 \triangleq \lambda_k$
  - Capacity: $C(\mathbf{H}) = \max_{\{s_k\}_{k=1}^n} \sum_{k=1}^n \log_2\left[1 + \rho s_k \lambda_k\right] = \sum_{k=1}^n \log_2\left[1 + \rho s_k^\star \lambda_k\right]$

Proposition

The power allocation strategy $\{s_1, \ldots, s_n\} = \{s_1^\star, \ldots, s_n^\star\}$ *that maximizes* $\sum_{k=1}^n \log_2\left(1 + \rho\lambda_k s_k\right)$ *under the power constraint* $\sum_{k=1}^n s_k = 1$*, is given by the water-filling solution,*

$$s_k^\star = \left(\mu - \frac{1}{\rho\lambda_k}\right)^+, \; k = 1, \ldots, n$$

where μ *is chosen so as to satisfy the power constraint* $\sum_{k=1}^n s_k^\star = 1$.

Water-Filling Algorithm

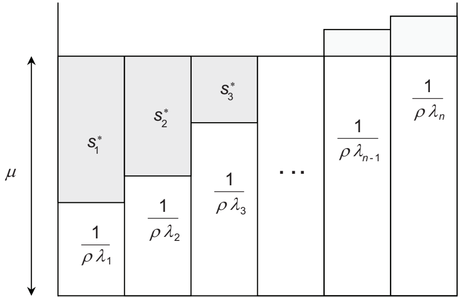
- Iterative power allocation
  - Order eigenvalues λ_k in decreasing order of magnitude
  - At iteration i, evaluate the constant μ from the power constraint

$$\mu(i) = \frac{1}{n-i+1}\left(1 + \sum_{k=1}^{n-i+1} \frac{1}{\rho\lambda_k}\right)$$

  - Calculate power

$$s_k(i) = \mu(i) - \frac{1}{\rho\lambda_k}, \qquad k = 1, \dots, n-i+1.$$

    If $s_{n-i+1} < 0$, set to 0
  - Iterate till the power allocated on each mode is non negative.



Water-Filling Algorithm

Example

Question: Consider the transmission $\mathbf{y} = \mathbf{H}\mathbf{c}' + \mathbf{n}$ with perfect CSIT over a deterministic point to point MIMO channel whose matrix is given by

$$\mathbf{H} = \begin{bmatrix} a & 0 & a & 0 \\ 0 & b & 0 & b \end{bmatrix}$$

where a and b are complex scalars with $|a| \geq |b|$. The input covariance matrix is given by $\mathbf{Q} = \mathcal{E}\left\{\mathbf{c}'\mathbf{c}'^H\right\}$ and is subject to the transmit power constraint $\text{Tr}\left\{\mathbf{Q}\right\} \leq P$.

1. Compute the capacity with perfect CSIT of that deterministic channel. Particularize to the case $a = b$. Explain your reasoning.
2. Explain how to achieve that capacity.
3. In which deployment scenario, could such channel matrix structure be encountered?

Water-Filling Algorithm

Example

Answer:

1. Let us write $\mathbf{Q} = \mathbf{V}\mathbf{P}\mathbf{V}^H$ with the diagonal element of $\mathbf{P}$, denoted as P_k (satisfying $\sum_{k=1}^{n_t} P_k = P$), refers to the power allocated to stream k. The capacity with perfect CSIT over the deterministic channel $\mathbf{H}$ is given by

$$C(\mathbf{H}) = \max_{P_1,\ldots,P_k} \sum_{k=1}^{\min\{2,4\}} \log_2\left(1 + \frac{P_k}{\sigma_n^2}\lambda_k\right)$$

where λ_k refers the non-zero eigenvalue of $\mathbf{H}^H\mathbf{H}$, respectively equal to $2|a|^2$ and $2|b|^2$. Hence,

$$C(\mathbf{H}) = \max_{P_1,P_2}\left(\log_2\left(1 + \frac{P_1}{\sigma_n^2}2|a|^2\right) + \log_2\left(1 + \frac{P_2}{\sigma_n^2}2|b|^2\right)\right).$$

The optimal power allocation is given by the water-filling solution

$$P_1^\star = \left(\mu - \frac{\sigma_n^2}{2|a|^2}\right)^+, \qquad P_2^\star = \left(\mu - \frac{\sigma_n^2}{2|b|^2}\right)^+$$

with μ computed such that $P_1^\star + P_2^\star = P$.

Water-Filling Algorithm

Example

Answer:
Assuming $P_1^\star$ and $P_2^\star$ are positive, $\mu = \frac{P}{2} + \frac{\sigma_n^2}{4}\left(\frac{1}{|a|^2} + \frac{1}{|b|^2}\right)$. If $\mu - \frac{\sigma_n^2}{2|b|^2} \leq 0$, i.e. $\frac{P}{2} + \frac{\sigma_n^2}{4|a|^2} - \frac{\sigma_n^2}{4|b|^2} \leq 0$, $P_2^\star = 0$ and $P_1^\star = P$. The capacity writes as

$$C(\mathbf{H}) = \log_2\left(1 + \frac{P}{\sigma_n^2} 2\left|a\right|^2\right).$$

If $\frac{P}{2} + \frac{\sigma_n^2}{4|a|^2} - \frac{\sigma_n^2}{4|b|^2} > 0$, $P_1^\star = \frac{P}{2} - \frac{\sigma_n^2}{4|a|^2} + \frac{\sigma_n^2}{4|b|^2}$ and $P_2^\star = \frac{P}{2} + \frac{\sigma_n^2}{4|a|^2} - \frac{\sigma_n^2}{4|b|^2}$. The capacity writes as

$$C(\mathbf{H}) = \log_2\left(1 + \frac{P_1^\star}{\sigma_n^2} 2\left|a\right|^2\right) + \log_2\left(1 + \frac{P_2^\star}{\sigma_n^2} 2\left|b\right|^2\right).$$

In the particular case where $a = b$, uniform power allocation $P_1^\star = P_2^\star = \frac{P}{2}$ is optimal and

$$C(\mathbf{H}) = 2\log_2\left(1 + \frac{P}{\sigma_n^2}\left|a\right|^2\right).$$

Water-Filling Algorithm

Example

Answer:

2. Transmit along $\mathbf{V}$, given by the two dominant eigenvector of $\mathbf{H}^H\mathbf{H}$. They are easily computed given the orthogonality of the channel matrix $\mathbf{H}$ as

$$\mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

   The power allocated to the two streams is given by $P_1^\star$ and $P_2^\star$. At the receiver, the precoded channel is already decoupled and no further combiner is necessary. Each stream can be decoded using the corresponding SISO decoder.

3. Dual-polarized antenna deployment (e.g. VHVH-VH) with LoS and good antenna XPD.

□

Capacity Bounds and Suboptimal Power Allocations

- Low SNR: power allocated to the dominant eigenmode

$$C\left(\mathbf{H}\right) \overset{\rho \to 0}{\to} \log_2\left(1+\rho\lambda_{max}\right).$$

- High SNR: power is uniformly allocated among the non-zero modes

$$C\left(\mathbf{H}\right) \overset{\rho \to \infty}{\to} \sum_{k=1}^{n} \log_2\left(1+\frac{\rho}{n}\lambda_k\right).$$

- At any SNR
  - lower bound

$$C\left(\mathbf{H}\right) \geq \log_2\left(1+\rho\lambda_{max}\right),$$

$$C\left(\mathbf{H}\right) \geq \sum_{k=1}^{n} \log_2\left(1+\frac{\rho}{n}\lambda_k\right).$$

  - upper bound (use Jensen's inequality $\mathcal{E}_x\left\{\mathcal{F}\left(x\right)\right\} \leq \mathcal{F}\left(\mathcal{E}_x\left\{x\right\}\right)$ if $\mathcal{F}$ concave)

$$C_{CSIT}\left(\mathbf{H}\right) = \sum_{k=1}^{n} \log_2\left[1+\rho s_k^{\star}\lambda_k\right] \overset{(a)}{\leq} n\log_2\left(1+\frac{\rho}{n}\left[\sum_{k=1}^{n} s_k^{\star}\lambda_k\right]\right),$$

$$\leq n\log_2\left[1+\frac{\rho}{n}\lambda_{max}\right].$$

Ergodic Capacity of Fast Fading Channels

- Fast fading:
  - Doppler frequency sufficiently high to allow for coding over many channel realizations/coherence time periods
  - The transmission capability is represented by a single quantity known as the ergodic capacity
- MIMO Capacity with Perfect Transmit Channel Knowledge
  - similar strategy as in deterministic channels: transmit along eigenvectors of channel matrix and allocate power following water-filling
  - short term power constraint: water-filling solution applied over space as in deterministic channels

$$\bar{C}_{CSIT,ST} = \mathcal{E}\left\{ \max_{\mathbf{Q}\geq 0:\mathrm{Tr}\{\mathbf{Q}\}=1} \log_2 \det\left[\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right]\right\}$$
$$= \sum_{k=1}^{n} \mathcal{E}\left\{ \log_2\left[1+\rho s_k^{\star}\lambda_k\right]\right\}.$$

  - long term power constraint: water-filling solution applied over both time and space

$$\bar{C}_{CSIT,LT} = \sum_{k=1}^{n} \mathcal{E}\left\{ \log_2\left[1+\rho s_k^{\star}\lambda_k\right]\right\}.$$

  - Impact on coding strategy? Use a variable-rate code (family of codes of different rates) adapted as a function of the water-filling allocation. No need for the codeword to span many coherence time periods.

MIMO Capacity with Partial Transmit Channel Knowledge

- $\mathbf{H}$ is not known to the transmitter $\rightarrow$ we cannot adapt $\mathbf{Q}$ at all time instants
- Rate of information flow between Tx and Rx at time instant k over channels $\mathbf{H}_k$

$$\log_2 \det \left[\mathbf{I}_{n_r} + \rho \mathbf{H}_k \mathbf{Q} \mathbf{H}_k^H\right].$$

Such a rate varies over time according to the channel fluctuations. The average rate of information flow over a time duration $T >> T_{coh}$ is

$$\frac{1}{T} \sum_{k=0}^{T-1} \log_2 \det \left[\mathbf{I}_{n_r} + \rho \mathbf{H}_k \mathbf{Q} \mathbf{H}_k^H\right].$$

Definition

The ergodic capacity of a $n_r \times n_t$ MIMO channel with channel distribution information at the transmitter (CDIT) is given by

$$\bar{C}_{CDIT} \triangleq \bar{C} = \max_{\mathbf{Q} \geq 0 : \mathrm{Tr}\{\mathbf{Q}\} = 1} \mathcal{E}\left\{ \log_2 \det \left[\mathbf{I}_{n_r} + \rho \mathbf{H} \mathbf{Q} \mathbf{H}^H\right]\right\},$$

where $\mathbf{Q}$ is the input covariance matrix optimized as to maximize the ergodic mutual information.

- $T >> T_c$ to average out the noise and the channel fluctuations

I.I.D. Rayleigh Fast Fading Channels: Perfect Transmit Channel Knowledge

- Low SNR: allocate all the available power to the strongest or dominant eigenmode. Use $\log_2(1+x) \approx x \log_2(e)$ for x small and get

$$\begin{aligned}
\bar{C}_{CSIT,ST} &= \mathcal{E}\left\{ \log_2\left[1+\rho\lambda_{max}\right] \right\} \\
&\cong \rho\mathcal{E}\left\{\lambda_{max}\right\} \log_2(e) \\
&\cong \rho n \log_2(e), \qquad N, n \to \infty, N/n >> 0. \\
\bar{C}_{CSIT,LT} &= \mathcal{E}\left\{ \log_2\left[1+\rho s^{\star}_{max}\lambda_{max}\right] \right\} \\
&\cong \rho\mathcal{E}\left\{s^{\star}_{max}\lambda_{max}\right\} \log_2(e)
\end{aligned}$$

Observations: $\bar{C}_{CSIT}$ grows linearly in the minimum number of antennas n.

- High SNR: uniform power allocation on all non-zeros eigenmodes

$$\bar{C}_{CSIT} \cong \sum_{k=1}^{n} \mathcal{E}\left\{ \log_2\left[1+\frac{\rho}{n}\lambda_k\right] \right\} \cong n\log_2\left(\frac{\rho}{n}\right) + \mathcal{E}\left\{ \sum_{k=1}^{n} \log_2(\lambda_k) \right\}.$$

Observations: $\bar{C}_{CSIT}$ also scales linearly with n. The spatial multiplexing gain is $g_s = n$. MISO fading channels do not offer any multiplexing gain.

I.I.D. Rayleigh Fast Fading Channels: Partial Transmit Channel Knowledge

- Optimal covariance matrix

Proposition

In i.i.d. Rayleigh fading channels, the ergodic capacity with CDIT is achieved under an equal power allocation scheme $\mathbf{Q} = \mathbf{I}_{n_t}/n_t$*, i.e.,*

$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = \mathcal{E}\left\{\log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}_w\mathbf{H}_w^H\right]\right\} = \mathcal{E}\left\{\sum_{k=1}^{n}\log_2\left[1 + \frac{\rho}{n_t}\lambda_k\right]\right\}.$$

Encoding requires a fixed-rate code (whose rate is given by the ergodic capacity) with encoding spanning many channel realizations.

- Low SNR:

$$\bar{C}_{CDIT} \geq \mathcal{E}\left\{\log_2\left[1 + \frac{\rho}{n_t}\|\mathbf{H}_w\|_F^2\right]\right\} \approx \frac{\rho}{n_t}\mathcal{E}\left\{\|\mathbf{H}_w\|_F^2\right\}\log_2(e) = n_r\rho\log_2(e)$$

Observations:

- $\bar{C}_{CDIT}$ is only determined by the energy of the channel.
- A MIMO channel only yields a n_r gain over a SISO channel. Increasing the number of transmit antennas is not useful (contrary to perfect CSIT). SIMO and MIMO channels reach the same capacity for a given n_r.

I.I.D. Rayleigh Fast Fading Channels: Partial Transmit Channel Knowledge

- High SNR:

$$\bar{C}_{CDIT} \approx \mathcal{E}\left\{\sum_{k=1}^{n} \log_2\left[\frac{\rho}{n_t}\lambda_k\right]\right\} = n\log_2\left(\frac{\rho}{n_t}\right) + \mathcal{E}\left\{\sum_{k=1}^{n} \log_2(\lambda_k)\right\}$$

Observations:

  - $\bar{C}_{CDIT}$ at high SNR scales linearly with n (by contrast to the low SNR regime).
  - The multiplexing gain g_s is equal to n, similarly to the CSIT case.
  - $\bar{C}_{CDIT}$ and $\bar{C}_{CSIT}$ are not equal: constant gap equal to $n \log_2(n_t/n)$ at high SNR.
- Expressions can be particularized to SISO, SIMO, MISO cases. At high SNR,
  - SISO ($N = n = 1$):

$$\bar{C}_{CDIT} \approx \log_2(\rho) + \mathcal{E}\left\{\log_2\left(|h|^2\right)\right\} = \log_2(\rho) - 0.83 = C_{AWGN} - 0.83$$

  - SIMO ($n_t = n = 1$, $n_r = N$):

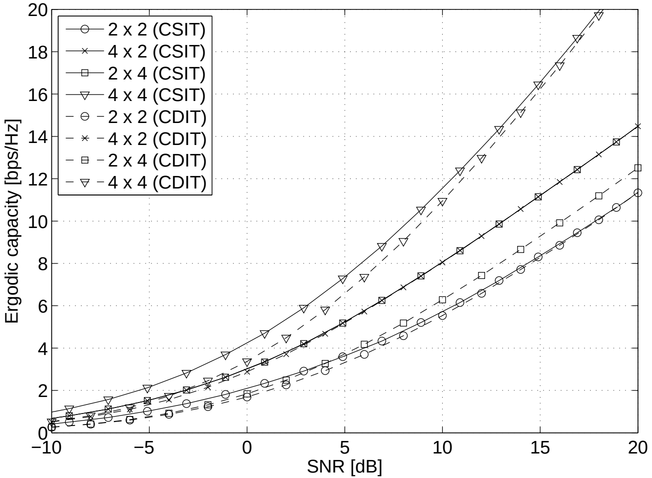
$$\bar{C}_{CDIT} \approx \log_2(n_r\rho)$$

  - MISO ($n_r = n = 1$, $n_t = N$):

$$\bar{C}_{CDIT} \approx \log_2(\rho) + \mathcal{E}\left\{\log_2\left(\|h\|^2/n_t\right)\right\} \overset{n_t\to\infty}{\approx} \log_2(\rho) = C_{AWGN}$$

I.I.D. Rayleigh Fast Fading Channels

- Ergodic capacity of various $n_r \times n_t$ i.i.d. Rayleigh channels with full (CSIT) and partial (CDIT) channel knowledge at the transmitter.



Correlated Rayleigh Fast Fading Channels: Uniform Power Allocation

- Assume the channel covariance matrix is unknown to the transmitter
- Mutual information with identity input covariance matrix

$$\bar{\mathcal{I}}_e = \mathcal{E}\left\{\log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right]\right\}.$$

- Low SNR

$$\bar{\mathcal{I}}_e \geq \mathcal{E}\left\{\log_2\left[1 + \frac{\rho}{n_t}\left\|\mathbf{H}\right\|_F^2\right]\right\}.$$

- High SNR in Kronecker Correlated Rayleigh Channels $\mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$ (with full rank correlation matrices) and $n_t = n_r$

$$\bar{\mathcal{I}}_e \approx \mathcal{E}\left\{\log_2 \det\left[\frac{\rho}{n_t}\mathbf{H}_w\mathbf{H}_w^H\right]\right\} + \log_2 \det(\mathbf{R}_r) + \log_2 \det(\mathbf{R}_t).$$

Observations:

- $\det(\mathbf{R}_r) \leq 1$ and $\det(\mathbf{R}_t) \leq 1$: receive and transmit correlations always degrade the mutual information (with power uniform allocation) with respect to the i.i.d. case.
- $\bar{\mathcal{I}}_e$ still scales linearly with $\min\{n_t, n_r\}$

Correlated Rayleigh Fast Fading Channels: Partial Transmit Channel Knowledge

- Assume the channel covariance matrix is known to the transmitter.

Proposition

In Kronecker correlated Rayleigh fast fading channels, the optimal input covariance matrix can again be expressed as

$$\mathbf{Q} = \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{Q}} \mathbf{U}_{\mathbf{R}_t}^H,$$

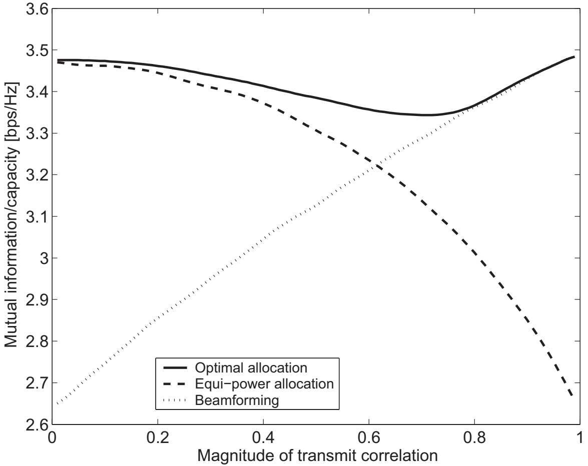
where $\mathbf{U}_{\mathbf{R}_t}$ is a unitary matrix formed by the eigenvectors of $\mathbf{R}_t$ (arranged in such order that they correspond to decreasing eigenvalues of $\mathbf{R}_t$), and $\mathbf{\Lambda}_{\mathbf{Q}}$ is a diagonal matrix whose elements are also arranged in decreasing order.

Power allocation has to be computed numerically. Approximation using Jensen's inequality is possible.

- Spatial correlation: beneficial or detrimental?
  - receive correlations degrade both the mutual information $\bar{\mathcal{I}}_e$ and the capacity with CDIT,
  - transmit correlations always decrease $\bar{\mathcal{I}}_e$ but may increase $\bar{C}_{CDIT}$ at low SNR (irrespective of n_t and n_r) or at higher SNR when $n_t > n_r$ (analogous to the full CSIT case).

Correlated Rayleigh Fast Fading Channels: Partial Transmit Channel Knowledge

- Mutual information of various strategies at 0 dB SNR as a function of the transmit correlation $|t|$ in TIMO. Beamforming refers here to the tranmsission of one stream along the dominant eigenvector of $\mathbf{R}_t$.



Outage Capacity and Probability in Slow Fading Channels

- In slow fading, the encoding still averages out the randomness of the noise but cannot fully average out the randomness of the channel.
- For a given channel realization $\mathbf{H}$ and a target rate R, reliable transmission if

$$\log_2 \det\left(\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right) > R$$

  If not met with any $\mathbf{Q}$, an outage occurs and the decoding error probability is strictly non-zero.
- Look at the tail probability of $\log_2 \det\left(\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right)$, not its average!

Definition

The outage probability $P_{out}(R)$ of a $n_r \times n_t$ MIMO channel with a target rate R is given by

$$P_{out}(R) = \min_{\mathbf{Q} \succeq 0 : \mathrm{Tr}\{\mathbf{Q}\} \leq 1} P\left(\log_2 \det\left(\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right) < R\right),$$

where $\mathbf{Q}$ is the input covariance matrix optimized as to minimize the outage probability.

- More meaningful in the absence of CSI knowledge at the transmitter: the transmitter cannot adjust its transmit strategy $\rightarrow$ hopes the channel is good enough

Diversity-Multiplexing Trade-Off in Slow Fading Channels

- Compound channel coding theorem: there exist "universal" codes with rate R bits/s/Hz that achieve reliable transmission over any slow fading channel realization which is not in outage.
  - CSIT is actually not necessary in slow fading channels if the aim is transmit reliably when the channel is not in outage.
- For a given R, how does P_{out} behave as a function of the SNR ρ?

Definition

A diversity gain $g_d^\star(g_s, \infty)$ is achieved at multiplexing gain g_s *at infinite SNR* if

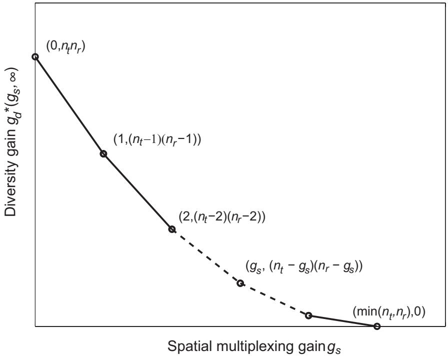
$$\begin{aligned} \lim_{\rho \longrightarrow \infty} \frac{R(\rho)}{\log_2(\rho)} &= g_s \\ \lim_{\rho \longrightarrow \infty} \frac{\log_2\left(P_{out}(R)\right)}{\log_2(\rho)} &= -g_d^\star(g_s, \infty) \end{aligned}$$

The curve $g_d^\star(g_s, \infty)$ as function of g_s is known as the asymptotic diversity-multiplexing trade-off of the channel.

- The multiplexing gain indicates how fast the transmission rate increases with the SNR.
- The diversity gain represents how fast the outage probability decays with the SNR.

Diversity-Multiplexing Trade-Off in I.I.D. Rayleigh Slow Fading Channels

- Point $(0, n_t n_r)$: for a spatial multiplexing gain of zero (i.e., R is fixed), the maximal diversity gain achievable is $n_t n_r$.
- Point $(\min\{n_t, n_r\}, 0)$: transmitting at diversity gain $g_d^\star =$ (i.e., P_{out} is kept fixed) allows the data rate to increase with SNR as $n = \min\{n_t, n_r\}$.
- Intermediate points: possible to transmit at non-zero diversity and multiplexing gains but that any increase of one of those quantities leads to a decrease of the other quantity.



Diversity-Multiplexing Trade-Off in I.I.D. Rayleigh Slow Fading Channels

- For fixed rates $R = 2, 4, ..., 40$ bits/s/Hz,
  - The asymptotic slope of each curve is four and matches the maximum diversity gain $g_d^\star(0, \infty)$.
  - The horizontal separation is 2 bits/s/Hz per 3 dB, which corresponds to the maximum multiplexing gain equal to $n(= 2)$.
- As the rate increases more rapidly with SNR (i.e., as the multiplexing gain g_s increases), the slope of the outage probability curve (given by the diversity gain g_d^*) vanishes.

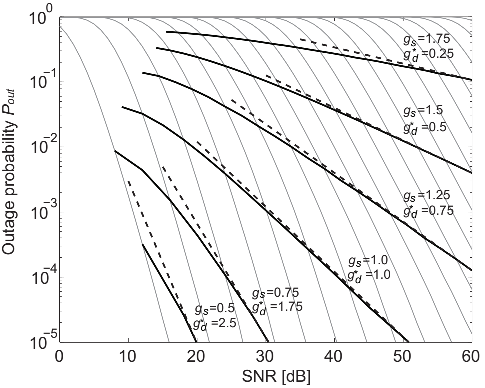


Figure: 2×2 MIMO i.i.d. Rayleigh fading channels

Diversity-Multiplexing Trade-Off of a Scalar Rayleigh Channel h

- Determine for a transmission rate R scaling with ρ as $g_s \log_2(\rho)$, the rate at which the outage probability decreases with ρ as ρ increases.
- Outage probability

$$
\begin{aligned}
P_{\text{out}}(R) &= P\left(\log_2\left[1+\rho\left|h\right|^2\right] < g_s \log_2(\rho)\right) \\
&= P\left(1+\rho\left|h\right|^2 < \rho^{g_s}\right)
\end{aligned}
$$

- At high SNR,

$$
P_{\text{out}}(R) \approx P\left(\left|h\right|^2 \leq \rho^{-(1-g_s)}\right)
$$

- Since $\left|h\right|^2$ is exponentially distributed, i.e., $P\left(\left|h\right|^2 \leq \epsilon\right) \approx \epsilon$ for small ϵ

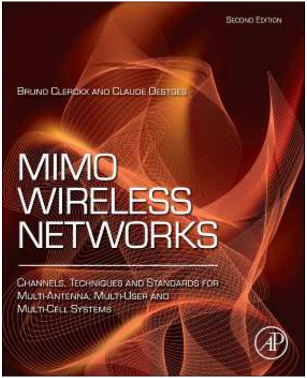
$$
P_{\text{out}}(R) \approx \rho^{-(1-g_s)}
$$

  An outage occurs at high SNR when $\left|h\right|^2 \leq \rho^{-(1-g_s)}$ with a probability $\rho^{-(1-g_s)}$.
- DMT for the scalar Rayleigh fading channel $g_d^\star(g_s, \infty) = 1 - g_s$ for $g_s \in [0,1]$.

Space-Time Coding over I.I.D. Rayleigh Flat Fading Channels

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Chapter 6
    - Section: 6.1, 6.2, 6.3 (except "Antenna Selection" in 6.3.2), 6.4.1, 6.4.2 (except the Proofs), 6.5.1, 6.5.2, 6.5.3, 6.5.4, 6.5.8, Figure 7.1

Introduction - Previous Lectures

- Previous lecture
  - Capacity of deterministic MIMO channels

$$C(\mathbf{H}) = \max_{\mathbf{Q} \geq 0 : \mathsf{Tr}\{\mathbf{Q}\}=1} \log_2 \det \left[\mathbf{I}_{n_r} + \rho \mathbf{H} \mathbf{Q} \mathbf{H}^H \right].$$

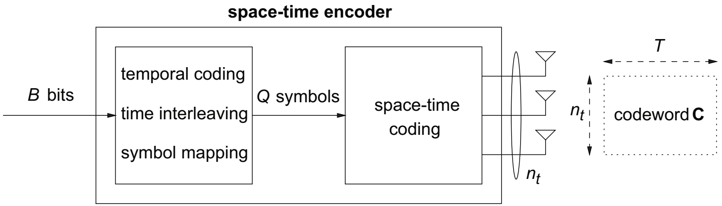
  - Ergodic capacity of fast fading channels
  - Outage capacity and probability of slow fading channels

- MIMO provides huge gains in terms of reliability and transmission rate
  - diversity gain, array gain, coding gain, spatial multiplexing gain, interference management

- What we further need
  - practical methodologies to achieve these gains?
  - how to code across space and time?
  - Some preliminary answers: multimode eigenmode transmission when channel knowledge available at the Tx, Alamouti scheme when no channel knowledge available at the Tx

Overview of a Space-Time Encoder

- Space-time encoder: sequence of two black boxes



- First black box: combat the randomness created by the noise at the receiver.
- Second black box: spatial interleaver which spreads symbols over several antennas in order to mitigate the spatial selective fading.
- The ratio B/T is the signaling rate of the transmission.
- The ratio Q/T is defined as the spatial multiplexing rate (representative of how many symbols are packed within a codeword per unit of time).

System Model

- MIMO system with n_t transmit and n_r receive antennas over a frequency flat-fading channel
- Transmit a codeword $\mathbf{C} = [\mathbf{c}_0 \dots \mathbf{c}_{T-1}]$ $[n_t \times T]$ contained in the codebook $\mathcal{C}$
- At the k^{th} time instant, the transmitted and received signals are related by

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}_k + \mathbf{n}_k$$

  where
  - $\mathbf{y}_k$ is the $n_r \times 1$ received signal vector,
  - $\mathbf{H}_k$ is the $n_r \times n_t$ channel matrix,
  - $\mathbf{n}_k$ is a $n_r \times 1$ zero mean complex AWGN vector with $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_l^H\} = \sigma_n^2\mathbf{I}_{n_r}\delta(k-l)$,
  - The parameter E_s is the energy normalization factor. SNR $\rho = E_s/\sigma_n^2$.
- No transmit channel knowledge but we know it is i.i.d. Rayleigh fading.
- Codeword average transmit power $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{C}\mathbf{C}^H\right\}\right\} = T$. Assume $\mathcal{E}\left\{\|\mathbf{H}\|_F^2\right\} = n_t n_r$.
- Channel time variation:
  - *slow fading*: $T_{coh} >> T$ and $\{\mathbf{H}_k = \mathbf{H}_w\}_{k=0}^{T-1}$, with $\mathbf{H}_w$ denoting an i.i.d. random fading matrix with unit variance circularly symmetric complex Gaussian entries.
  - *fast fading*: $T \geq T_{coh}$ and $\mathbf{H}_k = \mathbf{H}_{k,w}$, where $\left\{\mathbf{H}_{k,w}\right\}_{k=0}^{T-1}$ are uncorrelated matrices, each $\left\{\mathbf{H}_{k,w}\right\}$ being an i.i.d. random fading matrix with unit variance circularly symmetric complex Gaussian entries.

Error Probability Motivated Design Methodology

- With instantaneous channel realizations perfectly known at the receive side, the ML decoder computes an estimate of the transmitted codeword according to

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \sum_{k=0}^{T-1} \left\| \mathbf{y}_k - \sqrt{E_s}\mathbf{H}_k \mathbf{c}_k \right\|^2$$

  where the minimization is performed over all possible codeword vectors $\mathbf{C}$.
- Pairwise Error Probability (PEP): probability that the ML decoder decodes the codeword $\mathbf{E} = [\mathbf{e}_0 \ldots \mathbf{e}_{T-1}]$ instead of the transmitted codeword $\mathbf{C}$.
- When the PEP is conditioned on the channel realizations $\{\mathbf{H}_k\}_{k=0}^{T-1}$, it is defined as the conditional PEP,

$$P\left(\mathbf{C} \to \mathbf{E} | \{\mathbf{H}_k\}_{k=0}^{T-1}\right) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\sum_{k=0}^{T-1} \|\mathbf{H}_k\left(\mathbf{c}_k - \mathbf{e}_k\right)\|_F^2}\right)$$

  where $\mathcal{Q}(x)$ is the Gaussian $\mathcal{Q}$-function.
- The average PEP, $P(\mathbf{C} \to \mathbf{E})$, obtained by averaging the conditional PEP over the probability distribution of the channel gains.
- System performance dominated at high SNR by the couples of codewords that lead to the worst PEP.

- Assume a fixed rate transmission, i.e., spatial multiplexing gain $g_s = 0$.

Definition

The **diversity gain** $g_d^o(\rho)$ achieved by a pair of codewords $\{\mathbf{C}, \mathbf{E}\} \in \mathcal{C}$ is defined as the slope of $P(\mathbf{C} \to \mathbf{E})$ as a function of the SNR ρ on a log-log scale, usually evaluated at very high SNR, i.e.,

$$g_d^o(\infty) = \lim_{\rho \to \infty} g_d^o(\rho) = - \lim_{\rho \to \infty} \frac{\log_2 (P(\mathbf{C} \to \mathbf{E}))}{\log_2 \rho}.$$

PS: $g_d^o(\infty) \leftrightarrow P(\mathbf{C} \to \mathbf{E})$, $g_d^\star(0, \infty) \leftrightarrow P_{out}$.

Definition

The **coding gain** achieved by a pair of codewords $\{\mathbf{C}, \mathbf{E}\} \in \mathcal{C}$ is defined as the magnitude of the left shift of the $P(\mathbf{C} \to \mathbf{E})$ vs. ρ curve evaluated at very high SNR.

- If $P(\mathbf{C} \to \mathbf{E})$ is well approximated at high SNR by

$$P(\mathbf{C} \to \mathbf{E}) \approx c\,(g_c \rho)^{-g_d^o(\infty)}$$

with c being a constant, g_c is identified as the coding gain.

Fast Fading MIMO Channels

- In i.i.d. Rayleigh fast fading channels, average PEP reads as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{k=0}^{T-1} \left(1 + \eta \|\mathbf{c}_k - \mathbf{e}_k\|^2\right)^{-n_r} d\beta$$

  where $\eta = \rho/(4\sin^2\beta)$.
- Upper bound using the Chernoff bound

$$P(\mathbf{C} \to \mathbf{E}) \leq \prod_{k=0}^{T-1} \left(1 + \frac{\rho}{4} \|\mathbf{c}_k - \mathbf{e}_k\|^2\right)^{-n_r}.$$

- In the high SNR regime, the average PEP is further upper-bounded by

$$P(\mathbf{C} \to \mathbf{E}) \leq \left(\frac{\rho}{4}\right)^{-n_r l_{\mathbf{C},\mathbf{E}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r}$$

  with $l_{\mathbf{C},\mathbf{E}}$ the effective length of the pair of codewords $\{\mathbf{C}, \mathbf{E}\}$, i.e., $l_{\mathbf{C},\mathbf{E}} = \sharp\tau_{\mathbf{C},\mathbf{E}}$ with $\tau_{\mathbf{C},\mathbf{E}} = \{k \mid \mathbf{c}_k - \mathbf{e}_k \neq 0\}$.
- Diversity gain: $n_r l_{\mathbf{C},\mathbf{E}}$, coding gain: $\prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r}$.

The Distance-Product Criterion

- At high SNR, the error probability is naturally dominated by the worst-case PEP

Design Criterion

(**Distance-product criterion**) *Over i.i.d. Rayleigh fast fading channels,*

1. **distance criterion**: *maximize the minimum effective length L_{min} of the code over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$*

$$L_{min} = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C}\neq\mathbf{E}}} l_{\mathbf{C},\mathbf{E}}$$

2. **product criterion**: *maximize the minimum product distance d_p of the code over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$*

$$d_p = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C}\neq\mathbf{E} \\ l_{\mathbf{C},\mathbf{E}}=L_{min}}} \prod_{k\in\tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^2$$

- The presence of multiple antennas at the transmitter does not impact the achievable diversity gain $g_d^o(\infty) = n_r L_{min}$ but improves the coding gain $g_c = d_p$.
- The diversity gain is maximized first, and the coding gain is maximized only in a second step.

Slow Fading MIMO Channels

- In i.i.d. Rayleigh fast fading channels, average PEP reads as

$$P\left(\mathbf{C}\rightarrow\mathbf{E}\right)=\frac{1}{\pi}\int_{0}^{\pi/2}\left[\det\left(\mathbf{I}_{n_t}+\eta\tilde{\mathbf{E}}\right)\right]^{-n_r}d\beta \quad (1)$$

  where $\tilde{\mathbf{E}}\triangleq\left(\mathbf{C}-\mathbf{E}\right)\left(\mathbf{C}-\mathbf{E}\right)^{H}$.
- Upper bound using the Chernoff bound

$$\begin{aligned}P\left(\mathbf{C}\rightarrow\mathbf{E}\right)&\leq\left[\det\left(\mathbf{I}_{n_t}+\frac{\rho}{4}\tilde{\mathbf{E}}\right)\right]^{-n_r}\\&=\prod_{i=1}^{r(\tilde{\mathbf{E}})}\left(1+\frac{\rho}{4}\lambda_i(\tilde{\mathbf{E}})\right)^{-n_r}\end{aligned}$$

  with $r(\tilde{\mathbf{E}})$ denotes the rank of the error matrix $\tilde{\mathbf{E}}$ and $\{\lambda_i(\tilde{\mathbf{E}})\}$ for $i=1,\ldots,r(\tilde{\mathbf{E}})$ the set of its non-zero eigenvalues.
- At high SNR, $\frac{\rho}{4}\lambda_i(\tilde{\mathbf{E}})>>1$

$$P\left(\mathbf{C}\rightarrow\mathbf{E}\right)\leq\left(\frac{\rho}{4}\right)^{-n_r r(\tilde{\mathbf{E}})}\prod_{i=1}^{r(\tilde{\mathbf{E}})}\lambda_i^{-n_r}(\tilde{\mathbf{E}})$$

- diversity gain: $n_r\, r(\tilde{\mathbf{E}})$, coding gain: $\prod_{i=1}^{r(\tilde{\mathbf{E}})}\lambda_i(\tilde{\mathbf{E}})$.

The Rank-Determinant Criterion

Design Criterion

(**Rank-determinant criterion**) *Over i.i.d. Rayleigh slow fading channels,*

1. **rank criterion***: maximize the minimum rank* r_{min} *of* $\tilde{\mathbf{E}}$ *over all pairs of codewords* $\{\mathbf{C}, \mathbf{E}\}$ *with* $\mathbf{C} \neq \mathbf{E}$

$$r_{min} = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C}\neq\mathbf{E}}} r(\tilde{\mathbf{E}})$$

2. **determinant criterion***: over all pairs of codewords* $\{\mathbf{C}, \mathbf{E}\}$ *with* $\mathbf{C} \neq \mathbf{E}$*, maximize the minimum of the product* d_λ *of the non-zero eigenvalues of* $\tilde{\mathbf{E}}$*,*

$$d_\lambda = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C}\neq\mathbf{E}}} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i(\tilde{\mathbf{E}}).$$

If $r_{min} = n_t$*, the determinant criterion comes to maximize the minimum determinant of the error matrix over all pairs of codewords* $\{\mathbf{C}, \mathbf{E}\}$ *with* $\mathbf{C} \neq \mathbf{E}$

$$d_\lambda = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C}\neq\mathbf{E}}} \det(\tilde{\mathbf{E}}).$$

The Rank-Determinant Criterion

Definition

A full-rank (a.k.a. full-diversity) code is characterized by $r_{min} = n_t$. A rank-deficient code is characterized by $r_{min} < n_t$.

Example

Rank-deficient and full-rank codes for $n_t = 2$

- Rank-deficient code

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \mathbf{E} = \frac{1}{\sqrt{2}} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

- Full-rank code

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}, \mathbf{E} = \frac{1}{\sqrt{2}} \begin{bmatrix} e_1 & -e_2^* \\ e_2 & e_1^* \end{bmatrix}$$

$$(\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H = \frac{1}{2} \begin{bmatrix} |c_1 - e_1|^2 + |c_2 - e_2|^2 & 0 \\ 0 & |c_1 - e_1|^2 + |c_2 - e_2|^2 \end{bmatrix}$$

The Rank-Determinant Criterion

Example

Question: Relying on the rank-determinant criterion, show that delay diversity achieves full diversity. Assume for simplicity two transmit antennas.
Answer: The codeword for delay diversity can be written as

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & c_2 & \dots & c_{T-1} & 0 \\ 0 & c_1 & c_2 & \dots & c_{T-1} \end{bmatrix}.$$

Taking another codeword $\mathbf{E}$, different from $\mathbf{C}$,

$$\mathbf{E} = \frac{1}{\sqrt{2}} \begin{bmatrix} e_1 & e_2 & \dots & e_{T-1} & 0 \\ 0 & e_1 & e_2 & \dots & e_{T-1} \end{bmatrix}.$$

The diversity gain is given by the minimum rank of the error matrix over all possible pairs of (different) codewords, i.e.

$$r_{min} = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} r\big(\tilde{\mathbf{E}}\big) = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} r\big(\mathbf{C} - \mathbf{E}\big).$$

The Rank-Determinant Criterion

Example

With delay diversity, we have

$$\mathbf{C}-\mathbf{E}=\frac{1}{\sqrt{2}}\left[\begin{array}{ccccc} c_1-e_1 & c_2-e_2 & \dots & c_{T-1}-e_{T-1} & 0 \\ 0 & c_1-e_1 & c_2-e_2 & \dots & c_{T-1}-e_{T-1} \end{array}\right].$$

Obviously, $r\big(\mathbf{C}-\mathbf{E}\big)\leq 2$. Actually, $r\big(\mathbf{C}-\mathbf{E}\big)=2$ as long as $\mathbf{C}\neq\mathbf{E}$. Indeed even in the case where all $c_k-e_k=0$ except for one index k (in order to keep $\mathbf{C}\neq\mathbf{E}$), e.g. $k=1$,

$$\mathbf{C}-\mathbf{E}=\frac{1}{\sqrt{2}}\left[\begin{array}{ccccc} c_1-e_1 & 0 & \dots & 0 & 0 \\ 0 & c_1-e_1 & 0 & \dots & 0 \end{array}\right],$$

the rank is equal to 2. Hence diversity gain of $2n_r$.

The Rank-Determinant Criterion

Example

Question: Assume that c_1, c_2, c_3 and c_4 are constellation symbols taken from a unit average energy QAM constellation. Consider the Linear Space-Time Block Code, characterized by codewords

$$\mathbf{C} = \frac{1}{2}\left[\begin{array}{cc} c_1 + c_3 & c_2 + c_4 \\ c_2 - c_4 & c_1 - c_3 \end{array}\right].$$

What is the diversity gain achieved by this code over slow Rayleigh fading channels?

Answer: Check the rank of its error matrix

$$\mathbf{C} - \mathbf{E} = \frac{1}{2}\left[\begin{array}{cc} d_1 + d_3 & d_2 + d_4 \\ d_2 - d_4 & d_1 - d_3 \end{array}\right]$$

where $d_k = c_k - e_k$ for $k = 1, ..., 4$. This code is rank deficient. It is easily seen that by taking two codewords $\mathbf{C}$ and $\mathbf{E}$ such that $d_3 = d_4 = 0$ and $d_1 = d_2 = d$ (which is encountered for any constellations), $r\left(\mathbf{C} - \mathbf{E}\right) = 1$. Hence diversity gain of n_r.

Information Theory Motivated Design Methodology: Fast Fading MIMO Channels - Achieving The Ergodic Capacity

- Recall Lecture 5&6: ergodic capacity

$$\bar{C} = \max_{\mathbf{Q}:\text{Tr}\{\mathbf{Q}\}=1} \mathcal{E}\left\{\log_2 \det\left(\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right\}.$$

- Perfect Transmit Channel Knowledge
  - transmit independent streams in the directions of the eigenvectors of the channel matrix $\mathbf{H}$.
  - For a total transmission rate R, each stream k can then be encoded using a capacity-achieving Gaussian code with rate R_k such that $\sum_{k=1}^{n} R_k = R$, ascribed a power $\lambda_k(\mathbf{Q})$ and be decoded independently of the other streams.
  - The optimal power allocation $\{\lambda_k^\star\}$ based on the water-filling allocation strategy.
  - Capacity achievable using a variable-rate coding strategy ($T = T_{coh}$ is enough as long as the noise can be averaged out).
- Partial Transmit Channel Knowledge
  - When the channel is i.i.d. Rayleigh fading, $\mathbf{Q} = (1/n_t)\,\mathbf{I}_{n_t}$.
  - Transmission of independent information symbols may be performed in parallel over n virtual spatial channels.
  - The transmitter is very similar to the CSIT case except that all eigenmodes now receive the same amount of power.
  - Transmit with uniform power allocation over n_t independent streams, each stream using an AWGN capacity-achieving code and perform joint ML decoding (independent decoding of all streams is clearly suboptimal due to interference between streams).

Information Theory Motivated Design Methodology: Slow Fading MIMO Channels Achieving The DMT

- Impossible to code over a large number of independent channel realizations $\rightarrow$ separate coding leads to an outage as soon as one of the subchannels is in deep fade.
- Joint coding across all subchannels necessary in the absence of transmit channel knowledge!
- Rank-determinant criterion focuses on diversity maximization under fixed rate.
- What if we want to design codes achieving the diversity-multiplexing trade-off?

Definition

A scheme $\{\mathcal{C}(\rho)\}$, i.e., a family of codes indexed by the SNR ρ, is said to achieve a diversity gain $g_d(g_s, \infty)$ and a multiplexing gain g_s at high SNR if

$$\lim_{\rho \to \infty} \frac{R(\rho)}{\log_2(\rho)} = g_s$$

$$\lim_{\rho \to \infty} \frac{\log_2(P_e(\rho))}{\log_2(\rho)} = -g_d(g_s, \infty)$$

where $R(\rho)$ is the data rate and $P_e(\rho)$ the average error probability averaged over the additive noise, the i.i.d. channel statistics and the transmitted codewords. The curve $g_d(g_s, \infty)$ is the diversity-multiplexing trade-off achieved by the scheme in the high SNR regime.

Space-Time Block Coding (STBC)

- STBCs can be seen as a mapping of Q symbols (complex or real) onto a codeword $\mathbf{C}$ of size $n_t \times T$.

- Codewords are uncoded in the sense that no error correcting code is contained in the STBC.

- Linear STBCs are by far the most widely used
  - Spread information symbols in space and time in order to improve either the diversity gain, either the spatial multiplexing rate ($r_s = \frac{Q}{T}$) or both the diversity gain and the spatial multiplexing rate.
  - Pack more symbols into a given codeword, i.e., increase Q, to increase the data rate.

Example

Alamouti code: $n_t = 2$, $Q = 2$, $T = 2$, $r_s = 1$

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}.$$

A General Framework for Linear STBCs

- A linear STBC is expressed in its general form as

$$\mathbf{C} = \sum_{q=1}^{Q} \mathbf{\Phi}_q \Re[c_q] + \mathbf{\Phi}_{q+Q} \Im[c_q]$$

where
  - $\mathbf{\Phi}_q$ are complex basis matrices of size $n_t \times T$,
  - c_q stands for the complex information symbol (taken for example from PSK or QAM constellations),
  - Q is the number of complex symbols c_q transmitted over a codeword,
  - $\Re$ and $\Im$ stand for the real and imaginary parts.

Definition

Tall $(T \leq n_t)$ unitary basis matrices are such that $\mathbf{\Phi}_q^H \mathbf{\Phi}_q = \frac{1}{Q}\mathbf{I}_T$ $\forall q = 1, \ldots, 2Q$. *Wide* $(T \geq n_t)$ unitary basis matrices are such that $\mathbf{\Phi}_q \mathbf{\Phi}_q^H = \frac{T}{Qn_t}\mathbf{I}_{n_t}$ $\forall q = 1, \ldots, 2Q$.

Definition

The spatial multiplexing rate of a space-time block code is defined as $r_s = \frac{Q}{T}$. A full rate space-time block code is characterized by $r_s = n_t$.

A General Framework for Linear STBCs

- Apply the vec operator to $\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}_k + \mathbf{n}_k$ and make use of STBC structure

$$\mathcal{Y} = \sqrt{E_s}\mathcal{H}\mathcal{X}\mathcal{S} + \mathcal{N}$$

  where
  - $\mathcal{Y}[2n_rT \times 1]$ is the channel output vector

$$\mathcal{Y} = \text{vec}\left(\left[\begin{array}{c} \Re\left[\begin{array}{ccc} \mathbf{y}_0 & \cdots & \mathbf{y}_{T-1} \end{array}\right] \\ \Im\left[\begin{array}{ccc} \mathbf{y}_0 & \cdots & \mathbf{y}_{T-1} \end{array}\right] \end{array}\right]\right),$$

  - $\mathcal{H}[2n_rT \times 2n_tT]$ is the block diagonal channel

$$\mathcal{H} = \mathbf{I}_T \otimes \mathbf{H}', \quad \text{where} \quad \mathbf{H}' = \left[\begin{array}{cc} \Re[\mathbf{H}] & -\Im[\mathbf{H}] \\ \Im[\mathbf{H}] & \Re[\mathbf{H}] \end{array}\right],$$

  - $\mathcal{X}[2n_tT \times 2Q]$ is the linear code matrix

$$\mathcal{X} = \left[\begin{array}{ccc} \text{vec}\left(\left[\begin{array}{c} \Re[\mathbf{\Phi}_1] \\ \Im[\mathbf{\Phi}_1] \end{array}\right]\right) & \cdots & \text{vec}\left(\left[\begin{array}{c} \Re[\mathbf{\Phi}_{2Q}] \\ \Im[\mathbf{\Phi}_{2Q}] \end{array}\right]\right) \end{array}\right],$$

  - $\mathcal{S}[2Q \times 1]$ is a block of uncoded input symbols

$$\mathcal{S} = \left[\begin{array}{cccccc} \Re[c_1] & \cdots & \Re[c_Q] & \Im[c_1] & \cdots & \Im[c_Q] \end{array}\right]^T,$$

  - $\mathcal{N}[2n_rT \times 1]$ is the noise vector

$$\mathcal{N} = \text{vec}\left(\left[\begin{array}{c} \Re\left[\begin{array}{ccc} \mathbf{n}_0 & \cdots & \mathbf{n}_{T-1} \end{array}\right] \\ \Im\left[\begin{array}{ccc} \mathbf{n}_0 & \cdots & \mathbf{n}_{T-1} \end{array}\right] \end{array}\right]\right),$$

A General Framework for Linear STBCs

- Average Pairwise Error Probability of STBCs

Proposition

A PSK/QAM based linear STBC consisting of unitary basis matrices minimizes the worst-case PEP (1) averaged over i.i.d. Rayleigh slow fading channels if (sufficient condition) the unitary basis matrices $\{\boldsymbol{\Phi}_q\}_{q=1}^{2Q}$ *satisfy the conditions*

$$\boldsymbol{\Phi}_q \boldsymbol{\Phi}_p^H + \boldsymbol{\Phi}_p \boldsymbol{\Phi}_q^H = \mathbf{0}_{n_t}, \quad q \neq p \text{ for wide } \{\boldsymbol{\Phi}_q\}_{q=1}^{2Q},$$
$$\boldsymbol{\Phi}_q^H \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_p^H \boldsymbol{\Phi}_q = \mathbf{0}_T, \quad q \neq p \text{ for tall } \{\boldsymbol{\Phi}_q\}_{q=1}^{2Q}.$$

Proof: Using Hadamard's inequality and the unitarity of basis matrices,

$$\min_{q=1,\ldots,Q} \min_{d_q} \det\left(\mathbf{I}_{n_t} + \eta \tilde{\mathbf{E}}\right) \leq \det\left(\mathbf{I}_{n_t} + \eta \frac{T}{Q n_t} \mathbf{I}_{n_t} d_{min}^2\right).$$

Equality if unitary basis matrices are skew-hermitian. □

A General Framework for Linear STBCs

- Ergodic Capacity in i.i.d. Rayleigh Fading Channels

$$\bar{C} = \max_{\text{Tr}\left\{\mathcal{X}\mathcal{X}^T\right\} \leq 2T} \frac{1}{2T} \mathcal{E}_{\mathcal{H}} \left\{ \log \det \left(\mathbf{I}_{2n_r T} + \frac{\rho}{2} \mathcal{H}\mathcal{X}\mathcal{X}^T \mathcal{H}^T \right) \right\}$$

where it is assumed without loss of generality that $\mathcal{E}\left\{\mathcal{S}\mathcal{S}^H\right\} = \mathbf{I}_{2Q}$.

Proposition

In i.i.d. Rayleigh fading channels, linear STBCs with wide ($Q \geq n_t T$) matrices $\mathcal{X}$ that satisfy

$$\mathcal{X}\mathcal{X}^T = \frac{1}{n_t} \mathbf{I}_{2n_t T}$$

are capacity-efficient.

By capacity-efficient, we mean a code that maximizes the average mutual information in the sense that it preserves the capacity without inducing any loss of capacity if concatenated with capacity-achieving outer codes.

A General Framework for Linear STBCs

- Decoding

Proposition

Applying the space-time matched filter $\mathcal{X}^H\mathcal{H}^H$ to the output vector $\mathcal{Y}$ decouples the transmitted symbols

$$\mathcal{X}^H\mathcal{H}^H\mathcal{Y} = \frac{\sqrt{E_s}T}{Qn_t}\|\mathbf{H}\|_F^2\,\mathbf{I}_{2T}\mathcal{S} + \mathcal{X}^H\mathcal{H}^H\mathcal{N}$$

if and only if the basis matrices are wide unitary

$$\mathbf{\Phi}_q\mathbf{\Phi}_q^H = \frac{T}{Qn_t}\mathbf{I}_{n_t},\ \forall q = 1,\ldots,2Q$$

and pairwise skew-hermitian

$$\mathbf{\Phi}_q\mathbf{\Phi}_p^H + \mathbf{\Phi}_p\mathbf{\Phi}_q^H = \mathbf{0}_{n_t},\ \forall q \neq p.$$

The complexity of ML decoding of linear STBCs grows exponentially with n_t and Q. The decoupling property allows each symbol to be decoded independently of the presence of the other symbols through a simple space-time matched filter.

A General Framework for Linear STBCs

Example

A code such that $T = 1$, $n_t = 2$, $Q = 2$, $r_s = 2$ with the following **tall** basis matrices

$$\mathbf{\Phi}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \mathbf{\Phi}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ \mathbf{\Phi}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} j \\ 0 \end{bmatrix}, \ \mathbf{\Phi}_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ j \end{bmatrix},$$

or equivalently, with the following matrix $\mathcal{X}$

$$\mathcal{X} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This code is called Spatial Multiplexing. Optimal for worst-case PEP min. capacity-efficient, large decoding complexity.

A General Framework for Linear STBCs

Example

A code such that $T = 2$, $n_t = 2$, $Q = 2$, $r_s = 1$ with the following **wide** basis matrices

$$\begin{aligned} \mathbf{\Phi}_1 &= \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{\Phi}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \\ \mathbf{\Phi}_3 &= \tfrac{1}{\sqrt{2}} \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}, \quad \mathbf{\Phi}_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix}, \end{aligned}$$

or equivalently, with the following matrix $\mathcal{X}$

$$\mathcal{X} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

This code is called Alamouti code. Optimal for worst-case PEP min. not capacity-efficient, low decoding complexity.

Spatial Multiplexing/V-BLAST/D-BLAST

- Spatial Multiplexing (SM), also called V-BLAST, is a full rate code ($r_s = n_t$) that consists in transmitting independent data streams on each transmit antenna.
- In uncoded transmissions, we assume one symbol duration ($T = 1$) and codeword $\mathbf{C}$ is a symbol vector of size $n_t \times 1$.
- From previous results on error probability and capacity,

Proposition

Spatial Multiplexing with basis matrices characterized by square $\mathcal{X}$ such that

$$\mathcal{X}^T \mathcal{X} = \frac{1}{n_t} \mathbf{I}_{2n_t}$$

is capacity-efficient (Proposition 6) and optimal from an error rate minimization perspective (Proposition 5).

Example

$$\mathbf{C} = \frac{1}{\sqrt{n_t}} \begin{bmatrix} c_1 & \ldots & c_{n_t} \end{bmatrix}^T = \frac{1}{\sqrt{n_t}} \sum_{q=1}^{n_t} \mathbf{I}_{n_t}(:,q)\, \Re\left[c_q\right] + j \mathbf{I}_{n_t}(:,q)\, \Im\left[c_q\right].$$

Each element c_q is a symbol chosen from a given constellation.

ML decoding

- Error probability

$$
\begin{aligned}
P(\mathbf{C} \to \mathbf{E}) &= \frac{1}{\pi} \int_0^{\pi/2} \left[\det\left(\mathbf{I}_{n_t} + \eta \tilde{\mathbf{E}}\right)\right]^{-n_r} d\beta \\
&= \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \frac{\eta}{n_t} \sum_{q=1}^{n_t} |c_q - e_q|^2\right)^{-n_r} d\beta \\
&\leq \left(\frac{\rho}{4 n_t}\right)^{-n_r} \left(\sum_{q=1}^{n_t} |c_q - e_q|^2\right)^{-n_r}
\end{aligned}
$$

  The SNR exponent is equal to n_r. Due to the lack of coding across transmit antennas, no transmit diversity is achieved and only receive diversity is exploited.
- Over fast fading channels, we know that it is not necessary to code across antennas to achieve the ergodic capacity.

Proposition

Spatial Multiplexing with ML decoding and equal power allocation achieves the ergodic capacity of i.i.d. Rayleigh fast fading channels.

ML decoding

- Over slow fading channels, what is the multiplexing-diversity trade-off achieved by SM with ML decoding?

Proposition

For $n_r \geq n_t$, the diversity-multiplexing trade-off at high SNR achieved by Spatial Multiplexing with ML decoding and QAM constellation over i.i.d. Rayleigh fading channels is given by

$$g_d\left(g_s, \infty\right) = n_r \left(1 - \frac{g_s}{n_t}\right), \ \ g_s \in \left[0, n_t\right].$$

Zero-Forcing (ZF) Linear Receiver

- MIMO ZF receiver acts similarly to a ZF equalizer in frequency selective channels.
- ZF filtering effectively decouples the channel into n_t parallel channels
  - interference from other transmitted symbols is suppressed
  - scalar decoding may be performed on each of these channels
- The complexity of ZF decoding similar to SISO ML decoding, but the inversion step is responsible for the noise enhancement (especially at low SNR).
- Assuming that a symbol vector $\mathbf{C} = 1/\sqrt{n_t} \begin{bmatrix} c_1 & \dots & c_{n_t} \end{bmatrix}^T$ is transmitted, the output of the ZF filter $\mathbf{G}_{ZF}$ is given by

$$\mathbf{z} = \mathbf{G}_{ZF}\mathbf{y} = \begin{bmatrix} c_1 & \dots & c_{n_t} \end{bmatrix}^T + \mathbf{G}_{ZF}\mathbf{n}$$

where $\mathbf{G}_{ZF}$ inverts the channel,

$$\mathbf{G}_{ZF} = \sqrt{\frac{n_t}{E_s}}\mathbf{H}^\dagger$$

with $\mathbf{H}^\dagger = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ denoting the Moore-Penrose pseudo inverse.

Zero-Forcing (ZF) Linear Receiver

- Covariance matrix of the noise at the output of the ZF filter

$$\mathcal{E}\left\{\mathbf{G}_{ZF}\mathbf{n}\left(\mathbf{G}_{ZF}\mathbf{n}\right)^H\right\} = \frac{n_t}{\rho}\mathbf{H}^{\dagger}\left(\mathbf{H}^{\dagger}\right)^H = \frac{n_t}{\rho}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}.$$

- The output SNR on the q^{th} subchannel is thus given by

$$\rho_q = \frac{\rho}{n_t}\frac{1}{\left(\mathbf{H}^H\mathbf{H}\right)^{-1}(q,q)}, \qquad q = 1, \ldots, n_t.$$

- Inversion leads to noise enhancement. Severe degradation at low SNR.
- Assuming that the channel is i.i.d. Rayleigh distributed, ρ_q is a χ^2 random variable with $2\left(n_r - n_t + 1\right)$ degrees of freedom, denoted as $\chi^2_{2(n_r-n_t+1)}$. The average PEP on the q^{th} subchannel is thus upper-bounded by

$$P\left(c_q \to e_q\right) \le \left(\frac{\rho}{4n_t}\right)^{-(n_r-n_t+1)} \left|c_q - e_q\right|^{-2(n_r-n_t+1)}.$$

The lower complexity of the ZF receiver comes at the price of a diversity gain limited to $n_r - n_t + 1$. Clearly, the system is undetermined if $n_t > n_r$.

Zero-Forcing (ZF) Linear Receiver

- In fast fading channels, the average maximum achievable rate $\bar{C}_{ZF}$ is equal to the sum of the maximum rates achievable by all layers

$$\bar{C}_{ZF} = \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\left\{\log_2\left(1+\rho_q\right)\right\}$$

$$\stackrel{(\rho\nearrow)}{\approx} \min\left\{n_t,n_r\right\}\log_2\left(\frac{\rho}{n_t}\right) + \min\left\{n_t,n_r\right\}\mathcal{E}\left\{\log_2\left(\chi^2_{2(n_r-n_t+1)}\right)\right\}.$$

Note the difference with

$$\bar{C}_{CDIT} \approx n\text{log}_2\left(\frac{\rho}{n_t}\right) + \sum_{k=1}^{n}\mathcal{E}\left\{\text{log}_2(\chi^2_{2(N-n+k)})\right\}.$$

Spatial Multiplexing in combination with a ZF decoder allows for transmitting over $n = \min\{n_t,n_r\}$ independent data pipes.

Zero-Forcing (ZF) Linear Receiver

- In slow fading channels, what is the diversity-multiplexing trade-off achieved by SM with ZF?

Proposition

For $n_r \geq n_t$, the diversity-multiplexing trade-off achieved by Spatial Multiplexing with QAM constellation and ZF filtering in i.i.d. Rayleigh fading channels is given by

$$g_d\left(g_s, \infty\right) = \left(n_r - n_t + 1\right)\left(1 - \frac{g_s}{n_t}\right), \;\; g_s \in \left[0, n_t\right].$$

Zero-Forcing (ZF) Linear Receiver

- ZF receiver maximizes the SNR under the constraint that the interferences from all other layers are nulled out.
  - For a given layer q, the ZF combiner $\mathbf{g}_q$ is such that this layer is detected through a projection of the output vector $\mathbf{y}$ onto the direction closest to $\mathbf{H}(:,q)$ within the subspace orthogonal to the one spanned by the set of vectors $\mathbf{H}(:,p)$, $p \neq q$.
- Assume the following system model with $n_r \geq n_t$

$$\begin{aligned}\mathbf{y} &= \mathbf{H}\mathbf{c} + \mathbf{n}, \\ &= \mathbf{h}_q c_q + \sum_{p \neq q} \mathbf{h}_p c_p + \mathbf{n}\end{aligned}$$

where $\mathbf{h}_q$ is the q^{th} column of $\mathbf{H}$.
- Let us build the following $n_r \times (n_t - 1)$ matrix by collecting all h_p with $p \neq q$:

$$\begin{aligned}\mathbf{H}_{-q} &= \left[\begin{array}{ccc} \dots & \mathbf{h}_p & \dots \end{array}\right]_{p \neq q}, \\ &= \left[\begin{array}{cc} \mathbf{U}' & \tilde{\mathbf{U}} \end{array}\right] \mathbf{\Lambda} \mathbf{V}^H\end{aligned}$$

where $\tilde{\mathbf{U}}$ is the matrix containing the left singular vectors corresponding to the null singular values. Similarly we define

$$\mathbf{c}_{-q} = \left[\begin{array}{ccc} \dots & c_p & \dots \end{array}\right]_{p \neq q}^T .$$

Zero-Forcing (ZF) Linear Receiver

- By multiplying by $\tilde{\mathbf{U}}^H$, we project the output vector onto the subspace orthogonal to the one spanned by the columns of $\mathbf{H}'$

$$\begin{aligned}\tilde{\mathbf{U}}^H\mathbf{y} &= \tilde{\mathbf{U}}^H\mathbf{h}_q c_q + \tilde{\mathbf{U}}^H\mathbf{H}_{-q}\mathbf{c}_{-q} + \tilde{\mathbf{U}}^H\mathbf{n}\\ &= \tilde{\mathbf{U}}^H\mathbf{h}_q c_q + \tilde{\mathbf{U}}^H\mathbf{n}.\end{aligned}$$

- To maximize the SNR, noting the noise is still white, we match to the effective channel $\tilde{\mathbf{U}}^H\mathbf{h}_q$ such that

$$z = \left(\tilde{\mathbf{U}}^H\mathbf{h}_q\right)^H\tilde{\mathbf{U}}^H\mathbf{h}_q c_q + \left(\tilde{\mathbf{U}}^H\mathbf{h}_q\right)^H\tilde{\mathbf{U}}^H\mathbf{n}$$

and the ZF combiner is $g_q = \left(\tilde{\mathbf{U}}^H\mathbf{h}_q\right)^H\tilde{\mathbf{U}}^H = \mathbf{h}_q^H\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H$.

Minimum Mean Squared Error (MMSE) Linear Receiver

- Filter maximizing the SINR. Minimize the total resulting noise: find $\mathbf{G}$ such that $\mathcal{E}\{\left\|\mathbf{G}\mathbf{y} - \begin{bmatrix} c_1 & \ldots & c_{n_t} \end{bmatrix}^T\right\|^2\}$ is minimum.
- The combined noise plus interference signal $\mathbf{n}_{i,q}$ when estimating symbol c_q writes as

$$\mathbf{n}_{i,q} = \sum_{p \neq q} \sqrt{\frac{E_s}{n_t}} \mathbf{h}_p c_p + \mathbf{n}.$$

The covariance matrix of $\mathbf{n}_{i,q}$ reads as

$$\mathbf{R}_{\mathbf{n}_{i,q}} = \mathcal{E}\left\{\mathbf{n}_{i,q}\mathbf{n}_{i,q}^H\right\} = \sigma_n^2 \mathbf{I}_{n_r} + \sum_{p \neq q} \frac{E_s}{n_t} \mathbf{h}_p \mathbf{h}_p^H$$

and the MMSE combiner for stream q is given by

$$\mathbf{g}_{MMSE,q} = \sqrt{\frac{E_s}{n_t}} \mathbf{h}_q^H \left(\sigma_n^2 \mathbf{I}_{n_r} + \sum_{p \neq q} \frac{E_s}{n_t} \mathbf{h}_p \mathbf{h}_p^H\right)^{-1}.$$

- An alternative and popular representation of the MMSE filter can also be written as

$$\mathbf{G}_{MMSE} = \sqrt{\frac{n_t}{E_s}} \left(\mathbf{H}^H \mathbf{H} + \frac{n_t}{\rho} \mathbf{I}_{n_t}\right)^{-1} \mathbf{H}^H = \sqrt{\frac{n_t}{E_s}} \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H + \frac{n_t}{\rho} \mathbf{I}_{n_r}\right)^{-1}$$

- Bridge between matched filtering at low SNR and ZF at high SNR.

Minimum Mean Squared Error (MMSE) Linear Receiver

- The output SINR on the q^{th} subchannel (stream) is given by

$$\rho_q = \frac{E_s}{n_t}\mathbf{h}_q^H \left(\sigma_n^2 \mathbf{I}_{n_r} + \sum_{p \neq q} \frac{E_s}{n_t}\mathbf{h}_p\mathbf{h}_p^H\right)^{-1} \mathbf{h}_q.$$

- At high SNR, the MMSE filter is practically equivalent to ZF and the diversity achievable is thus limited to $n_r - n_t + 1$.

Successive Interference Canceler

- Successively decode one symbol (or more generally one layer/stream) and cancel the effect of this symbol from the received signal.
- Decoding order based on the SINR of each symbol/layer: the symbol/layer with the highest SINR is decoded first at each iteration.
- SM with (ordered) SIC is generally known as V-BLAST, and ZF and MMSE V-BLAST refer to SM with respectively ZF-SIC and MMSE-SIC receivers.

- The diversity order experienced by the decoded layer is increased by one at each iteration. Therefore, the symbol/layer detected at iteration i will achieve a diversity of $n_r - n_t + i$.
- Major issue: error propagation
  - The error performance is mostly dominated by the weakest stream.
  - Non-ordered SIC: diversity order approximately $n_r - n_t + 1$.
  - Ordered SIC: performance improved by reducing the error propagation caused by the first decoded stream. The diversity order remains lower than n_r.

Successive Interference Canceler

1. *Initialization:* $i \longleftarrow 1$, $\mathbf{y}^{(1)} = \mathbf{y}$,$\mathbf{G}^{(1)} = \mathbf{G}_{ZF}(\mathbf{H})$,$q_1 \overset{(*)}{=} \arg\min_j \left\|\mathbf{G}^{(1)}(j,:)\right\|^2$ where $\mathbf{G}_{ZF}(\mathbf{H})$ is defined as the ZF filter of the matrix $\mathbf{H}$.
2. *Recursion:*
   1. *step 1:* extract the q_i^{th} transmitted symbol from the received signal $\mathbf{y}^{(i)}$

      $$\tilde{c}_{q_i} = \mathbf{G}^{(i)}(q_i,:)\,\mathbf{y}^{(i)}$$

      where $\mathbf{G}^{(i)}(q_i,:)$ is the q_i^{th} row of $\mathbf{G}^{(i)}$;
   2. *step 2:* slice $\tilde{c}_{q_i}$ to obtain the estimated transmitted symbol $\hat{c}_{q_i}$;
   3. *step 3:* assume that $\hat{c}_{q_i} = c_{q_i}$ and construct the received signal

      $$\begin{aligned} \mathbf{y}^{(i+1)} &= \mathbf{y}^{(i)} - \sqrt{\frac{E_s}{n_t}}\mathbf{H}(:,q_i)\,\hat{c}_{q_i} \\ \mathbf{G}^{(i+1)} &= \mathbf{G}_{ZF}\left(\mathbf{H}_{\overline{q_i}}\right) \\ i &\longleftarrow i+1 \\ q_{i+1} &\overset{(*)}{=} \arg\min_{j \notin \{q_1,\ldots,q_i\}} \left\|\mathbf{G}^{(i+1)}(j,:)\right\|^2 \end{aligned}$$

      where $\mathbf{H}_{\overline{q_i}}$ is the matrix obtained by zeroing columns $q_1,\ldots,q_i$ of $\mathbf{H}$. Here $\mathbf{G}_{ZF}\left(\mathbf{H}_{\overline{q_i}}\right)$ denotes the ZF filter applied to $\mathbf{H}_{\overline{q_i}}$.

Successive Interference Canceler

- In fast fading channels, the maximum rate achievable with ZF-SIC

$$\bar{C}_{ZF-SIC} = \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\left\{\log_2\left(1+\rho_q\right)\right\}$$

$$\overset{(\rho\nearrow)}{\approx} \min\left\{n_t,n_r\right\}\log_2\left(\frac{\rho}{n_t}\right) + \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\left\{\log_2\left(\chi^2_{2(n_r-n_t+q)}\right)\right\} = \bar{C}_{CDIT}$$

The loss that was observed with ZF filtering is now compensated because the successive interference cancellation improves the SNR of each decoded layer.

Proposition

Spatial Multiplexing with ZF-SIC (ZF V-BLAST) and equal power allocation achieves the ergodic capacity of i.i.d. Rayleigh fast fading MIMO channels at asymptotically high SNR.

This only holds true when error propagation is neglected.

Successive Interference Canceler

- MMSE-SIC does better for any SNR

$$\bar{C}_{MMSE-SIC} = \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\left\{\log_2\left(1+\rho_q\right)\right\} = \mathcal{E}\left\{\log_2 \det\left(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right)\right\} = \bar{\mathcal{I}}_e,$$

Proposition

Spatial Multiplexing with MMSE-SIC (MMSE V-BLAST) and equal power allocation achieves the ergodic capacity for all SNR in i.i.d. Rayleigh fast fading MIMO channels.

Result also valid for a deterministic channel.

Successive Interference Canceler

- In slow fading channels, what is the diversity-multiplexing trade-off achieved by unordered ZF-SIC?

Proposition

For $n_r \geq n_t$, the diversity-multiplexing trade-off achieved by Spatial Multiplexing with QAM constellation and unordered ZF-SIC receiver over i.i.d. Rayleigh fading channels is given by

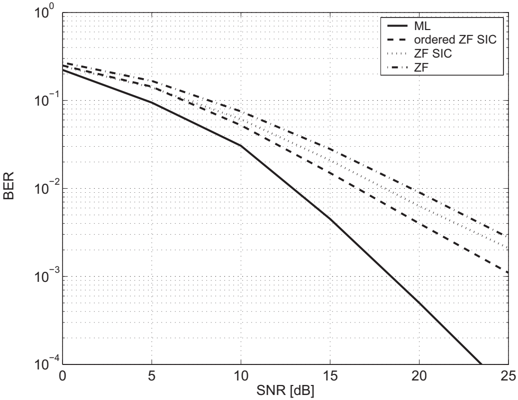
$$g_d(g_s, \infty) = (n_r - n_t + 1)\left(1 - \frac{g_s}{n_t}\right), \;\; g_s \in [0, n_t].$$

The achieved trade-off is similar to the trade-off achieved by a simple ZF receiver. This comes from the fact that the first layer dominates the error probability since its error exponent is the smallest.

- By increasing the number of receive antennas by 1,
  - with ZF or unordered ZF-SIC, we can *either* accommodate one extra stream with the same diversity order *or* increase the diversity order of every stream by 1,
  - with ML, we can accommodate one extra stream and *simultaneously* increase the diversity order of every stream by 1.

Impact of Decoding Strategy on Error Probability

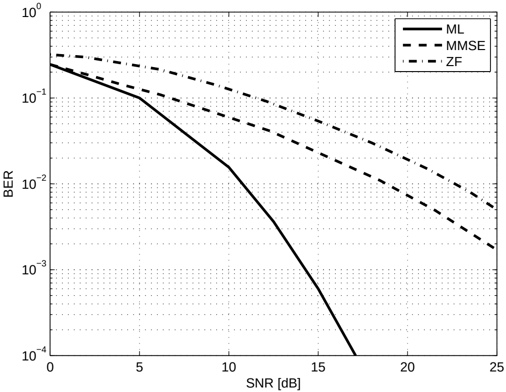
- SM with ML, ordered and non ordered ZF SIC and simple ZF decoding in 2×2 i.i.d. Rayleigh fading channels for 4 bits/s/Hz.



The slope of the ML curve approaches 2. ZF only achieves a diversity order of $n_r - n_t + 1 = 1$.

Impact of Decoding Strategy on Error Probability

- SM with ML, ZF and MMSE in i.i.d. Rayleigh slow fading channels with $n_t = n_r = 4$ and QPSK.



D-BLAST

- So far, transmission of independent data streams exploits a diversity order of at most n_r out of $n_t n_r$.
- Lack of coding across antennas: V-BLAST is in outage each time the SINR of a layer cannot support the rate allocated to that layer.
- Need for a spatial interleaving so that each layer encounters all antennas.
- D-BLAST: V-BLAST transmitter + a stream rotation following the encoding of all layers. Receiver is similar to V-BLAST.

Example

Consider two layers $\mathbf{a}$ and $\mathbf{b}$ and $n_t = 2$. Assume that layer $\mathbf{a}$ is made of two streams $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ and layer $\mathbf{b}$ of two streams as well $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$. Each stream can be seen as a block of symbols. The transmitted codeword $\mathbf{C}$ is now written as

$$\mathbf{C} = \begin{bmatrix} \mathbf{a}^{(1)} & \mathbf{b}^{(1)} & \\ & \mathbf{a}^{(2)} & \mathbf{b}^{(2)} \end{bmatrix}.$$

Orthogonal Space-Time Block Codes

- O-STBC vs. SM
  - Remarkable properties which make them extremely easy to decode: MIMO ML decoding decouples into several SIMO ML decoding
  - Achieve a full-diversity of $n_t n_r$.
  - Much smaller spatial multiplexing rate than SM.
- Linear STBC characterized by the two following properties
  1. the basis matrices are wide unitary

$$\mathbf{\Phi}_q \mathbf{\Phi}_q^H = \frac{T}{Q n_t} \mathbf{I}_{n_t} \;\; \forall q = 1 \ldots 2Q$$

  2. the basis matrices are pairwise skew-hermitian

$$\mathbf{\Phi}_q \mathbf{\Phi}_p^H + \mathbf{\Phi}_p \mathbf{\Phi}_q^H = 0, \;\; q \neq p$$

  or equivalently by this unique property

$$\mathbf{C}\mathbf{C}^H = \frac{T}{Q n_t} \left[\sum_{q=1}^{Q} |c_q|^2 \right] \mathbf{I}_{n_t}.$$

- Complex O-STBCs with $r_s = 1$ only exist for $n_t = 2$. For larger n_t, codes exist with $r_s \leq 1/2$. For some particular values of $n_t > 2$, complex O-STBCs with $1/2 < r_s < 1$ have been developed. This is the case for $n_t = 3$ and $n_t = 4$ with $r_s = 3/4$.

Orthogonal Space-Time Block Codes

Example

Alamouti code: complex O-STBC for $n_t = 2$ with a spatial multiplexing rate $r_s = 1$

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}.$$

- basis matrices are unitary and skew-hermitian (discussed before).
- $\mathbf{C}\mathbf{C}^H = \frac{1}{2}\left[\,|c_1|^2 + |c_2|^2\,\right]\mathbf{I}_2$.
- $r_s = 1$ since two symbols are transmitted over two symbol durations.

Example

For $n_t = 3$, a complex O-STBC expanding on four symbol durations ($T = 4$) and transmitting three symbols on each block ($Q = 3$)

$$\mathbf{C} = \frac{2}{3} \begin{bmatrix} c_1 & -c_2^* & c_3^* & 0 \\ c_2 & c_1^* & 0 & c_3^* \\ c_3 & 0 & -c_1^* & -c_2^* \end{bmatrix}.$$

The spatial multiplexing rate r_s is equal to $3/4$.

Orthogonal Space-Time Block Codes

Proposition

O-STBCs enjoy the decoupling property.

Example

Assume a MISO transmission based on the Alamouti code

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix} = \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_1 & h_2 \end{bmatrix} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix} + \begin{bmatrix} n_1 & n_2 \end{bmatrix}$$

or equivalently

$$\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{\frac{E_s}{2}} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\mathbf{H}_{eff}} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}.$$

Orthogonal Space-Time Block Codes

Example

Applying the space-time matched filter $\mathbf{H}_{eff}^H$ to the received vector decouples the transmitted symbols

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{H}_{eff}^H \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{\frac{E_s}{2}} \left[|h_1|^2 + |h_2|^2\right] \mathbf{I}_2 \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \mathbf{H}_{eff}^H \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}.$$

Expanding the original ML metric

$$\left| y_1 - \sqrt{\frac{E_s}{2}} \left(h_1 c_1 + h_2 c_2\right) \right|^2 + \left| y_2 - \sqrt{\frac{E_s}{2}} \left(-h_1 c_2^* + h_2 c_1^*\right) \right|^2$$

and making use of z_1 and z_2, the decision metric for c_1 is

$$\text{choose } c_i \text{ iff } \left| z_1 - \sqrt{\frac{E_s}{2}} \left(|h_1|^2 + |h_2|^2\right) c_i \right|^2 \leq \left| z_1 - \sqrt{\frac{E_s}{2}} \left(|h_1|^2 + |h_2|^2\right) c_k \right|^2 \; \forall i \neq k$$

and similarly for c_2. Independent decoding of symbols c_1 and c_2 is so performed.

Orthogonal Space-Time Block Codes

- Error Probability

$$
\begin{aligned}
P(\mathbf{C} \to \mathbf{E}) &= \frac{1}{\pi} \int_0^{\pi/2} \left[\det\left(\mathbf{I}_{n_t} + \eta \tilde{\mathbf{E}}\right)\right]^{-n_r} d\beta \\
&= \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta \frac{T}{Q n_t} \sum_{q=1}^{Q} |c_q - e_q|^2\right)^{-n_r n_t} d\beta \\
&\overset{(\rho \nearrow)}{\leq} \left(\frac{\rho}{4} \frac{T}{Q n_t}\right)^{-n_r n_t} \left(\sum_{q=1}^{Q} |c_q - e_q|^2\right)^{-n_r n_t} .
\end{aligned}
$$

Full diversity gain of $n_t n_r$.

Orthogonal Space-Time Block Codes

- O-STBCs are not capacity efficient $\mathcal{I}_{O-STBC}\left(\mathbf{H}\right) \leq \mathcal{I}_e\left(\mathbf{H}\right)$
  - mutual information of MIMO channel

$$\mathcal{I}_e\left(\mathbf{H}\right) = \log_2\left(1 + \frac{\rho}{n_t}\left\|\mathbf{H}\right\|_F^2 + \ldots + \left(\frac{\rho}{n_t}\right)^{r(\mathbf{H})}\prod_{k=1}^{r(\mathbf{H})}\lambda_k\left(\mathbf{H}\mathbf{H}^H\right)\right).$$

  - mutual information of MIMO channel transformed by the O-STBC

$$\mathcal{I}_{O-STBC}\left(\mathbf{H}\right) = \frac{Q}{T}\log_2\left(1 + \frac{\rho T}{Q n_t}\left\|\mathbf{H}\right\|_F^2\right).$$

Proposition

For a given channel realization $\mathbf{H}$, the mutual information achieved by any O-STBC is always upper-bounded by the channel mutual information with equal power allocation $\mathcal{I}_e$. Equality occurs if and only if both the rank of the channel and the spatial multiplexing rate of the code are equal to one.

Corollary

The Alamouti scheme is optimal with respect to the mutual information when used with only one receive antenna.

Orthogonal Space-Time Block Codes

- Diversity-Multiplexing Trade-off Achieved by O-STBCs

Proposition

The diversity-multiplexing trade-off at high SNR achieved by O-STBCs using QAM constellations in i.i.d. Rayleigh fading channels is given by

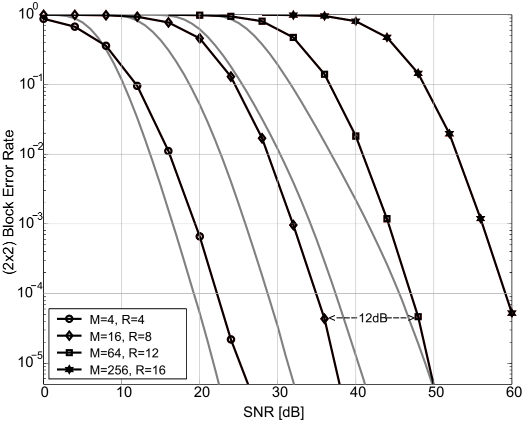
$$g_d\left(g_s, \infty\right) = n_r n_t \left(1 - \frac{g_s}{r_s}\right), \ \ g_s \in \left[0, r_s\right].$$

Proposition

The Alamouti code with any QAM constellation achieves the optimal diversity-multiplexing trade-off for two transmit and one receive antennas in i.i.d. Rayleigh fading channels.

Orthogonal Space-Time Block Codes

- Block error rate for 4 different rates $R = 4, 8, 12, 16$ bits/s/Hz in 2×2 i.i.d. slow Rayleigh fading channels.



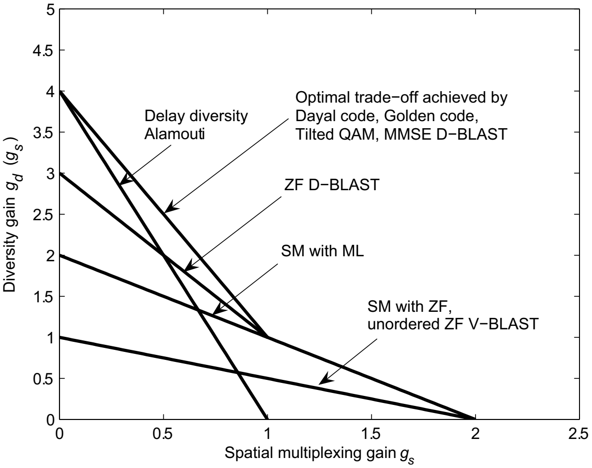
  - full diversity exploited: $g_d\left(g_s = 0, \infty\right) = g_d^o\left(\infty\right) = 4$.
  - the growth of the multiplexing gain is slow: 12 dB separate the curves, corresponding to a multiplexing gain $g_s = 1$, i.e., 1 bit/s/Hz increase per 3 dB SNR increase.

Other Code Constructions

- Quasi-Orthogonal Space-Time Block Codes
  - increase the spatial multiplexing rate while still partially enjoying the decoupling properties of O-STBCs
  - use O-STBCs of reduced dimensions as the building blocks of a higher dimensional code
- Linear Dispersion Codes
  - if a larger receiver complexity is authorized, it is possible to relax the skew-hermitian conditions and increase the data rates while still providing transmit diversity.
- Algebraic Space-Time Codes
  - structured codes using algebraic tools
  - many of them are designed to achieve the optimal diversity-multiplexing tradeoff

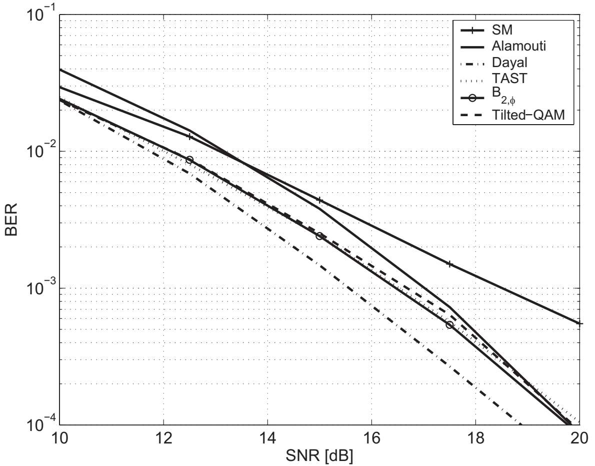
Global Performance Comparison

- Asymptotic diversity-multiplexing trade-off $g_d(g_s, \infty)$ achieved by several space-time codes in a 2×2 i.i.d. Rayleigh fading MIMO channel.



Global Performance Comparison

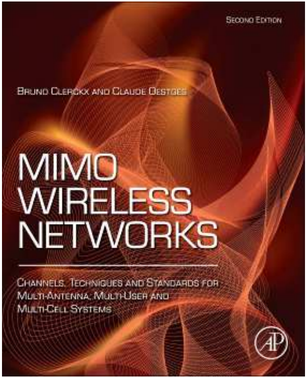
- Bit error rate (BER) of several space-time block codes in i.i.d. slow Rayleigh fading channels with $n_t = 2$ and $n_r = 2$ in a 4-bit/s/Hz transmission. ML decoding is used.



Space-Time Coding in Real-World MIMO Channels

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 8
  - Section: 8.1, 8.2, 8.3.1 ("Rayleigh Slow Fading Channels", "How Realistic is the High SNR Regime Approximation?"), 8.3.4

Introduction - Previous Lectures

- Space-time coding in i.i.d. Rayleigh fading
  - Distance-product criterion in fast fading
  - Rank-determinant criterion in slow fading

- Real-world channels span a large variety of propagation conditions.

- Some environments may highly deviate from the i.i.d. Rayleigh fading scenario.

- Objectives:
  - how codes developed under the *i.i.d. Rayleigh* assumption behave in more realistic propagation conditions, i.e., how these codes are affected by non ideal propagation conditions,
  - how a more adapted design criterion might significantly improve their performance.

System Model

- MIMO system with n_t transmit and n_r receive antennas communicating through a frequency flat-fading channel
- A codeword $\mathbf{C} = [\mathbf{c}_0 \ldots \mathbf{c}_{T-1}]$ of size $n_t \times T$ contained in the codebook $\mathcal{C}$ is transmitted over T symbol durations via n_t transmit antennas.
- At the k^{th} time instant, the transmitted and received signals are related by

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}_k + \mathbf{n}_k$$

  where
  - $\mathbf{y}_k$ is the $n_r \times 1$ received signal vector,
  - $\mathbf{H}_k$ is the $n_r \times n_t$ channel matrix,
  - $\mathbf{n}_k$ is a $n_r \times 1$ zero mean complex additive white Gaussian noise (AWGN) vector with $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_l^H\} = \sigma_n^2\mathbf{I}_{n_r}\delta(k-l)$,
  - The parameter E_s is the energy normalization factor, so that the ratio E_s/σ_n^2 represents the SNR denoted as ρ.
- We normalize the codeword average transmit power such that $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{C}\mathbf{C}^H\right\}\right\} = T$ and assume for simplicity that $\mathcal{E}\left\{\|\mathbf{H}\|_F^2\right\} = n_t n_r$.
- $\mathbf{H}_k$ is not i.i.d. anymore!

Radiation Patterns

- Decompose the channel $\mathbf{H}_k = \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)} = \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)}(:,1)\,\mathbf{a}_t^T\left(\theta_{t,k}^{(l)}\right)$, where $\mathbf{a}_t\left(\theta_{t,k}^{(l)}\right)$ is the transmit array response in the direction of departure $\theta_{t,k}^{(l)}$.
- PEP argument writes as

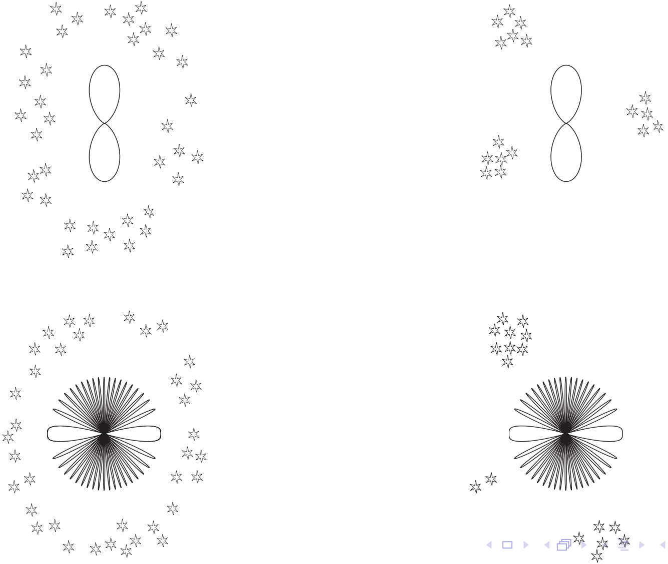
$$\sum_{k=0}^{T-1} \left\|\mathbf{H}_k\left(\mathbf{c}_k - \mathbf{e}_k\right)\right\|_F^2 = \sum_{k=0}^{T-1} \left\| \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)}(:,1)\mathbf{a}_t^T\left(\theta_{t,k}^{(l)}\right)\left(\mathbf{c}_k - \mathbf{e}_k\right) \right\|^2$$

- The original MIMO transmission can be considered as the SIMO transmission of an equivalent codeword, given at the k^{th} time instant by

$$\mathbf{a}_t^T \mathbf{c}_k$$

- It may be thought of as an array factor function of the transmitted codewords. At every symbol period,
  - the energy radiated in any direction varies as a function of the transmitted codewords.
  - for a given codeword and omnidirectional antennas, the radiated energy is not uniformly distributed in all directions, but may present maxima and minima in certain directions.

Radiation Patterns



Radiation Patterns

- What if the transmit angle spread decreases?

Definition

A MIMO channel is said to be *degenerate in the direction of departure* θ_t if all scatterers surrounding the transmitter are located along the same direction θ_t.

In the presence of small angle spread at the transmit side, the MIMO channel degenerates into a SIMO channel where the $1 \times T$ transmitted codeword is given by $\mathbf{a}_t^T(\theta_t)\,\mathbf{C}$

$$\sum_{k=0}^{T-1} \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2 = \sum_{k=0}^{T-1} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_t)\right|^2 \left[\sum_{n=1}^{n_r} |\mathbf{H}_k(n,1)|^2\right].$$

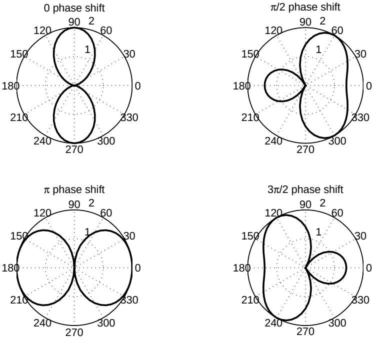
Since a space-time code designed for i.i.d. channels is only concerned with $\mathbf{C}$ and $\mathbf{E}$, its interaction with $\mathbf{a}_t(\theta_t)$ is not taken into account.

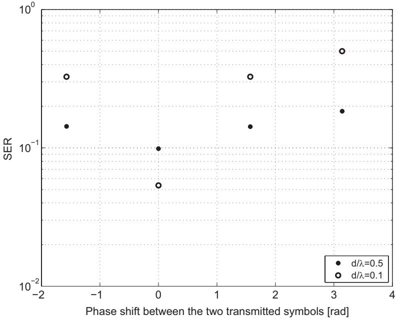
Radiation Patterns

Example

The Spatial Multiplexing example for $n_t = 2$: $\mathbf{c}_k = [\; c_1[k] \quad c_2[k] \;]^T$

$$\mathbf{c}_k^T \mathbf{a}_t(\theta_t) = c_1[k] \underbrace{\left[1 + \frac{c_2[k]}{c_1[k]} e^{-2\pi j \frac{d_t}{\lambda} \cos\theta_t} \right]}_{G_t(\theta_t | \mathbf{c}_k)}$$





Derivation of the Average PEP

- Conditional PEP

$$P\left(\mathbf{C} \to \mathbf{E} | \left\{\mathbf{H}_k\right\}_{k=0}^{T-1}\right) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\sum_{k=0}^{T-1}\left\|\mathbf{H}_k\left(\mathbf{c}_k - \mathbf{e}_k\right)\right\|_F^2}\right)$$

where $\mathcal{Q}(x)$ is the Gaussian $\mathcal{Q}$-function defined as

$$\mathcal{Q}(x) \triangleq P(y \geq x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} \exp\left(-\frac{y^2}{2}\right) dy.$$

- Average PEP

$$P(\mathbf{C} \to \mathbf{E}) = \mathcal{E}_{\mathbf{H}_k}\left\{P\left(\mathbf{C} \to \mathbf{E} | \left\{\mathbf{H}_k\right\}_{k=0}^{T-1}\right)\right\}.$$

- This integration is sometimes difficult to calculate. Therefore, alternatives forms of the Gaussian $\mathcal{Q}$-function are used.
  - Craig's formula

$$\mathcal{Q}(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2(\beta)}\right) d\beta.$$

  - Chernoff bound

$$\mathcal{Q}(x) \leq \exp\left(-\frac{x^2}{2}\right).$$

Derivation of the Average PEP

- We can derive the average PEP as follows

$$\begin{aligned} P\left(\mathbf{C}\to\mathbf{E}\right) &= \mathcal{E}_{\mathbf{H}_k}\left\{P\left(\mathbf{C}\to\mathbf{E}\middle|\left\{\mathbf{H}_k\right\}_{k=0}^{T-1}\right)\right\} \\ &= \frac{1}{\pi}\int_0^{\pi/2} M_\Gamma\left(-\frac{1}{2\sin^2\left(\beta\right)}\right)d\beta \\ &\leq M_\Gamma\left(-\frac{1}{2}\right) \end{aligned}$$

with $M_\Gamma\left(\gamma\right)$ moment generating function (MGF) of $\Gamma = \frac{\rho}{2}\sum_{k=0}^{T-1}\left\|\mathbf{H}_k\left(\mathbf{c}_k-\mathbf{e}_k\right)\right\|_F^2$

$$M_\Gamma\left(\gamma\right) \triangleq \int_0^\infty \exp\left(\gamma\Gamma\right)p_\Gamma\left(\Gamma\right)d\Gamma$$

Theorem

The moment generating function of a Hermitian quadratic form in complex Gaussian random variable $y=\mathbf{z}\mathbf{F}\mathbf{z}^H$*, where* $\mathbf{z}$ *is a circularly symmetric complex Gaussian vector with mean* $\bar{\mathbf{z}}$ *and a covariance matrix* $\mathbf{R}_\mathbf{z}$ *and* $\mathbf{F}$ *a Hermitian matrix, is given by*

$$M_y\left(s\right) \triangleq \int_0^\infty \exp\left(sy\right)p_y\left(y\right)dy = \frac{\exp\left(s\bar{\mathbf{z}}\mathbf{F}\left(\mathbf{I}-s\mathbf{R}_\mathbf{z}\mathbf{F}\right)^{-1}\bar{\mathbf{z}}^H\right)}{\det\left(\mathbf{I}-s\mathbf{R}_\mathbf{z}\mathbf{F}\right)}$$

Derivation of the Average PEP

- Apply to a joint Space-Time Correlated Ricean Fading Channels. Defining

$$\mathbf{H} = \sqrt{\frac{K}{1+K}}\left(\mathbf{1}_{1\times T}\otimes\bar{\mathbf{H}}\right) + \sqrt{\frac{1}{1+K}}\begin{bmatrix}\tilde{\mathbf{H}}_1 & \tilde{\mathbf{H}}_2 & \cdots & \tilde{\mathbf{H}}_T\end{bmatrix}$$

$$\mathbf{D} = \mathrm{diag}\left\{\mathbf{c}_1-\mathbf{e}_1, \mathbf{c}_2-\mathbf{e}_2, \ldots, \mathbf{c}_T-\mathbf{e}_T\right\},$$

we may write

$$\sum_{k=1}^{T}\left\|\mathbf{H}_k\left(\mathbf{c}_k-\mathbf{e}_k\right)\right\|_F^2 = \left\|\mathbf{HD}\right\|_F^2 = \mathrm{Tr}\left\{\mathbf{HDD}^H\mathbf{H}^H\right\} = \mathrm{vec}\left(\mathbf{H}^H\right)^H \boldsymbol{\Delta}\,\mathrm{vec}\left(\mathbf{H}^H\right)$$

where $\boldsymbol{\Delta} = \mathbf{I}_{n_r}\otimes\mathbf{DD}^H$. This is a hermitian quadratic form of complex Gaussian random variables where

- Define

$$\tilde{\mathcal{H}} = \mathrm{vec}\left(\begin{bmatrix}\tilde{\mathbf{H}}_0 & \tilde{\mathbf{H}}_1 & \cdots & \tilde{\mathbf{H}}_{T-1}\end{bmatrix}^H\right),$$

$$\bar{\mathcal{H}} = \mathrm{vec}\left(\left(\mathbf{1}_{1\times T}\otimes\bar{\mathbf{H}}\right)^H\right),$$

and

$$\boldsymbol{\Xi} = \mathcal{E}\{\tilde{\mathcal{H}}\tilde{\mathcal{H}}^H\}$$

as the spatio-temporal correlation matrix.

Derivation of the Average PEP

- PEP averaged over the space-time correlated Ricean fading channel

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\eta K \bar{\mathcal{H}}^H \boldsymbol{\Delta} \left(\mathbf{I}_{Tn_r n_t} + \eta \boldsymbol{\Xi} \boldsymbol{\Delta}\right)^{-1} \bar{\mathcal{H}}\right) \left(\det\left(\mathbf{I}_{Tn_r n_t} + \eta \boldsymbol{\Xi} \boldsymbol{\Delta}\right)\right)^{-1} d\beta$$

where the effective SNR η is defined as $\eta = \rho/(4\sin^2(\beta)(1+K))$.

- PEP averaged over the space-time correlated Rayleigh fading channel

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{Tn_r n_t} + \eta \boldsymbol{\Xi} \boldsymbol{\Delta}\right)\right)^{-1} d\beta$$

where $\eta = \rho/(4\sin^2(\beta))$.

Average PEP in Rayleigh Slow Fading Channels

Example

Question: Assume a spatially correlated slow Rayleigh fading MIMO channel. Derive the Average PEP for ML receiver.
Answer: The conditional PEP writes as

$$P(\mathbf{C} \to \mathbf{E} \mid \mathbf{H}) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\left\|\mathbf{H}(\mathbf{C}-\mathbf{E})\right\|_F^2}\right).$$

The average PEP over Rayleigh slow fading channels is

$$P(\mathbf{C} \to \mathbf{E}) = \mathcal{E}_{\mathbf{H}}\left\{P(\mathbf{C} \to \mathbf{E}|\mathbf{H})\right\} = \frac{1}{\pi}\int_0^{\pi/2} M_\Gamma\left(-\frac{1}{2\sin^2(\beta)}\right) d\beta$$

where $M_\Gamma(\gamma)$ moment generating function (MGF) of $\Gamma = \frac{\rho}{2}\left\|\mathbf{H}(\mathbf{C}-\mathbf{E})\right\|_F^2$.
Note that

$$\left\|\mathbf{H}(\mathbf{C}-\mathbf{E})\right\|_F^2 = \mathrm{Tr}\left\{\mathbf{H}\tilde{\mathbf{E}}\mathbf{H}^H\right\} = \mathrm{vec}\left(\mathbf{H}^H\right)^H\left(\mathbf{I}_{n_r}\otimes\tilde{\mathbf{E}}\right)\mathrm{vec}\left(\mathbf{H}^H\right)$$

where $\tilde{\mathbf{E}} = (\mathbf{C}-\mathbf{E})(\mathbf{C}-\mathbf{E})^H$.

Average PEP in Rayleigh Slow Fading Channels

Example

This is a hermitian quadratic form of complex gaussian random variables of the form $\mathbf{z}\mathbf{F}\mathbf{z}^H$ (with $\mathbf{z} = \text{vec}\left(\mathbf{H}^H\right)^H$ and $\mathbf{F} = \mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}$) and we can use Theorem where the mean $\bar{\mathbf{z}} = \mathbf{0}$ is the zero vector and the covariance matrix is $\mathbf{R}_{\mathbf{z}} = \mathbf{R} = \mathcal{E}\left\{\text{vec}(\mathbf{H}^H)\text{vec}(\mathbf{H}^H)^H\right\}$.
We then write (with $\eta = \rho/(4\sin^2(\beta))$)

$$P\left(\mathbf{C} \rightarrow \mathbf{E}\right) = \frac{1}{\pi}\int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \mathbf{C}_{\mathbf{R}}\right)\right)^{-1} d\beta.$$

where $\mathbf{C}_{\mathbf{R}} = \mathbf{R}\mathbf{F}$. With the Kronecker model, $\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t$, and

$$\begin{aligned} P\left(\mathbf{C} \rightarrow \mathbf{E}\right) &= \frac{1}{\pi}\int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta\left(\mathbf{R}_r \otimes \mathbf{R}_t\right)\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\right)\right)^{-1} d\beta \\ &= \frac{1}{\pi}\int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \mathbf{R}_r \otimes \mathbf{R}_t \tilde{\mathbf{E}}\right)\right)^{-1} d\beta. \end{aligned}$$

□

Average PEP in Rayleigh Slow Fading Channels

- Average PEP writes as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \mathbf{C}_{\mathbf{R}}\right)\right)^{-1} d\beta$$
$$= \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{C}_{\mathbf{R}})} \left(1 + \eta \lambda_i\left(\mathbf{C}_{\mathbf{R}}\right)\right)^{-1} d\beta.$$

- What happens at infinite SNR? $P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\mathbf{C}_{\mathbf{R}})} \prod_{i=1}^{r(\mathbf{C}_{\mathbf{R}})} \lambda_i^{-1}\left(\mathbf{C}_{\mathbf{R}}\right) d\beta$
  - Full rank code, i.e., $r(\tilde{\mathbf{E}}) = n_t$

$$\prod_{i=1}^{n_t n_r} \lambda_i\left(\mathbf{C}_{\mathbf{R}}\right) = \prod_{i=1}^{n_t n_r} \lambda_i\left(\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\right) = \det\left(\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\right) = \left(\det\left(\tilde{\mathbf{E}}\right)\right)^{n_r} \det\left(\mathbf{R}\right).$$

    No interactions between the channel and the code at very high SNR!
  - Non-full rank code, i.e., $r(\tilde{\mathbf{E}}) < n_t$

$$\prod_{i=1}^{r(\mathbf{C}_{\mathbf{R}})} \lambda_i\left(\mathbf{C}_{\mathbf{R}}\right) = \det\left(\mathbf{Q}'\right) \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{n_r}\left(\tilde{\mathbf{E}}\right)$$

    where $\mathbf{Q}'$ is a $n_r r(\tilde{\mathbf{E}}) \times n_r r(\tilde{\mathbf{E}})$ principal submatrix of $\mathbf{V}_{\mathbf{C}}^H \mathbf{R} \mathbf{V}_{\mathbf{C}}$ with $\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H = \mathbf{U}_{\mathbf{C}} \mathbf{\Lambda}_{\mathbf{C}} \mathbf{V}_{\mathbf{C}}^H$.

Average PEP in Rayleigh Slow Fading Channels

- Rank deficient codes sensitive to spatial correlation!

Example

Question: In a 2×2 spatially correlated Rayleigh fading MIMO channel, derive the Average PEP for Spatial Multiplexing with ML receiver and discuss the effect of transmit and receive correlation on the performance.
Answer: For SM, $\mathbf{C} - \mathbf{E}$ is a $n_t \times 1$ vector and

$$
\begin{aligned}
&P\left(\mathbf{C} \rightarrow \mathbf{E}\right)\\
&= \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \left(\mathbf{R}_r \otimes \mathbf{R}_t\right) \left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\right)\right)^{-1} d\beta\\
&= \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \left(\mathbf{R}_r \otimes \mathbf{R}_t\right) \left(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})\right) \left(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H\right)\right)\right)^{-1} d\beta\\
&= \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r} + \eta (\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H) \left(\mathbf{R}_r \otimes \mathbf{R}_t\right) \left(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})\right)\right)\right)^{-1} d\beta\\
&= \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r} + \eta \left(\mathbf{R}_r \otimes (\mathbf{C} - \mathbf{E})^H \mathbf{R}_t \left(\mathbf{C} - \mathbf{E}\right)\right)\right)\right)^{-1} d\beta.
\end{aligned}
$$

Average PEP in Rayleigh Slow Fading Channels

Example

For SM over a 2×2 MIMO channel,

$$\mathbf{C}-\mathbf{E}=\left[\begin{array}{c} c_0-e_0 \\ c_1-e_1 \end{array}\right], \quad \mathbf{R}_t=\left[\begin{array}{cc} 1 & t^* \\ t & 1 \end{array}\right], \quad \mathbf{R}_r=\left[\begin{array}{cc} 1 & r^* \\ r & 1 \end{array}\right],$$

$$a=(\mathbf{C}-\mathbf{E})^H \mathbf{R}_t (\mathbf{C}-\mathbf{E}) = |c_0-e_0|^2+|c_1-e_1|^2+2\Re\left\{t\left(c_0-e_0\right)\left(c_1-e_1\right)^*\right\}.$$

Hence

$$\begin{aligned} P(\mathbf{C}\to\mathbf{E}) &= \frac{1}{\pi}\int_0^{\pi/2}\left(\det\left(\mathbf{I}_2+\eta a\mathbf{R}_r\right)\right)^{-1} d\beta. \\ &= \frac{1}{\pi}\int_0^{\pi/2}\left(1+\eta a(1+|r|)\right)^{-1}\left(1+\eta a(1-|r|)\right)^{-1} d\beta. \end{aligned}$$

At high SNR, assuming $a>0$ and $|r|<1$, we get

$$P(\mathbf{C}\to\mathbf{E}) \approx \frac{1}{\pi}\int_0^{\pi/2}\left(\det\left(\mathbf{R}_r\right)\right)^{-1}\eta^{-2}a^{-2}d\beta = \frac{1}{\pi}\int_0^{\pi/2}\left(1-|r|^2\right)^{-1}\eta^{-2}a^{-2}d\beta.$$

As $|r|$ increases, the PEP increases. For large value of $|t|$, a can be very small for some error vectors leading to detrimental performance of SM.

Average PEP in Rayleigh Slow Fading Channels

Example

Observations:

- performance of SM in correlated channels depends on the projection of $\mathbf{C} - \mathbf{E}$ onto the space spanned by the eigenvectors of $\mathbf{R}_t$:
- worse performance when $\mathbf{C} - \mathbf{E}$ is parallel to the eigenvector of $\mathbf{R}_t$ corresponding to the lowest eigenvalue.
  Intuition: transmitting all the information contained in the unique non-zero eigenvalue of the error matrix $\tilde{\mathbf{E}}$ in the direction of the space offering the lowest scatterer density.
- receive correlation induces a coding gain loss independent of the error matrix. $\square$

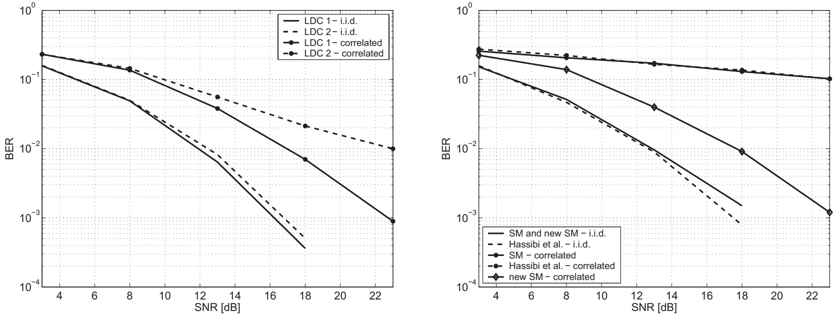
Average PEP in Rayleigh Slow Fading Channels

- Analysis can be extended to finite SNR and space-time correlated Rician channels (see more in Chapter 8 if interested).

- Main observations:
  - Rank deficient codes are very sensitive to spatial correlation.
  - Designing codes using the rank-determinant criterion is not sufficient to guarantee a good performance in spatially correlated Rayleigh slow fading channels when the code is rank-deficient.
  - Designing codes based on the distance-product criterion is not sufficient to guarantee a good performance in spatially correlated Rayleigh fast fading channels, irrespective of the rank of the code.
  - The maximization of the coding gain in i.i.d. Rayleigh channels is not a sufficient condition to guarantee the good performance of a code in correlated channels at finite SNR, even for full-rank codes.
  - Robust code design exist (see Chapter 9 if interested)

Average PEP in Rayleigh Slow Fading Channels

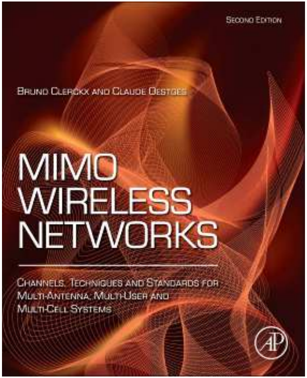
- Performance of full-rank and rank-deficient STBCs in i.i.d. and spatially correlated channels with $n_t = 2$ and $n_r = 2$.



MIMO with Partial Channel State Information at the Transmitter

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



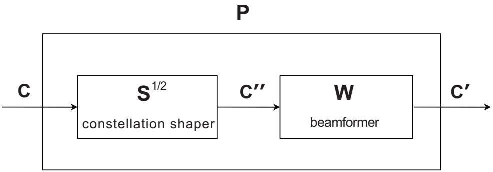
  - Chapter 10
    - Section: 10.1, 10.2.1, 10.5, 10.6.1, 10.6.2, 10.6.3, 10.6.4, 10.9

Introduction

- full CSIT
  - array and diversity gain
  - lower system complexity (parallel virtual transmissions)
  - hardly achievable (especially when the channel varies rapidly), costly in terms of feedback
- Exploiting Channel Statistics at the Transmitter
  - low rate feedback link
  - statistical properties of the channel (correlations, K-factor) vary at a much slower rate than the fading channel itself
  - The receiver estimates the channel stochastic properties and sends them back to the transmitter "once in a while" (if channel reciprocity cannot be exploited)
  - stationary channel: statistics do not change over time
- Exploiting a Limited Amount of Feedback at the Transmitter
  - codebook of precoding matrices, i.e., a finite set of precoders, designed off-line and known to both the transmitter and receiver.
  - The receiver estimates the best precoder as a function of the current channel and feeds back only the index of this best precoder in the codebook.

System Model

- MIMO system with n_t transmit and n_r receive antennas communicating through a frequency flat slow fading channel.
- The encoder outputs a codeword $\mathbf{C} = [\mathbf{c}_0 \dots \mathbf{c}_{T-1}]$ of size $n_e \times T$ contained in the codebook $\mathcal{C}$ over T symbol durations.
- Precoder $\mathbf{P}$ $[n_t \times n_e]$ processes the codeword $\mathbf{C}$ and the codeword $\mathbf{C}' = \mathbf{PC}$ $[n_t \times T]$ is transmitted over n_t antennas.



- Linear precoder $\mathbf{P} = \mathbf{W}\mathbf{S}^{1/2}$
  - multi-mode beamformer $\mathbf{W}$ whose columns have a unit-norm
  - constellation shaper $\mathbf{S}^{1/2}$ (if $\mathbf{S}$ real-valued and diagonal, it can be thought of as the power allocation scheme across the modes)
- nomalization: $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{C}'\mathbf{C}'^H\right\}\right\} = T$, $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{C}\mathbf{C}^H\right\}\right\} = T$ and $\text{Tr}\left\{\mathbf{P}\mathbf{P}^H\right\} = n_e$, $\mathcal{E}\left\{ \|\mathbf{H}\|_F^2 \right\} = n_t n_r$.

Channel Statistics based Precoding

- Information Theory motivated strategy.

Proposition

In Kronecker correlated Rayleigh fast fading channels, the optimal input covariance matrix can again be expressed as

$$\mathbf{Q} = \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{Q}} \mathbf{U}_{\mathbf{R}_t}^H,$$

where $\mathbf{U}_{\mathbf{R}_t}$ is a unitary matrix formed by the eigenvectors of $\mathbf{R}_t$ (arranged in such order that they correspond to decreasing eigenvalues of $\mathbf{R}_t$), and $\mathbf{\Lambda}_{\mathbf{Q}}$ is a diagonal matrix whose elements are also arranged in decreasing order.

Transmit a single stream along the dominant eigenvector of $\mathbf{R}_t$ if very large transmit correlation. Transmit multiple streams with uniform power allocation if very low transmit correltion.

- Error Probability motivated strategy

$$\mathbf{P}^{\star} = \arg\min_{\mathbf{P}} \max_{\tilde{\mathbf{E}} \neq \mathbf{0}} P\left(\mathbf{C} \rightarrow \mathbf{E}\right)$$

  - challenging problem for arbitrary codes
  - focus on O-STBC

Channel Statistics based Precoding

- O-STBCs in Kronecker Rayleigh fading channels

$$
\begin{aligned}
\mathbf{P}^{\star} &= \arg\max_{\mathbf{P}} \max_{\tilde{\mathbf{E}}\neq\mathbf{0}} \det\left(\mathbf{I}_{n_r n_t} + \zeta\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \mathbf{P}\mathbf{P}^H\right)\right) \\
&= \arg\max_{\mathbf{P}} \max_{\tilde{\mathbf{E}}\neq\mathbf{0}} \det\left(\mathbf{I}_{n_r n_t} + \zeta\left(\mathbf{R}_r \otimes \mathbf{R}_t\mathbf{P}\mathbf{P}^H\right)\right)
\end{aligned}
$$

where $\zeta = \eta T\delta^2/(Qn_e)$ and $\delta = d_{min}$.

Proposition

In Kronecker Rayleigh fading channels, the optimal precoder minimizing the average PEP/SER is given by $\mathbf{P} = \mathbf{W}\mathbf{S}^{1/2}$ *where*

- $\mathbf{W} = \mathbf{U}'_{\mathbf{R}_t}$ *with* $\mathbf{U}'_{\mathbf{R}_t}$ *the* $n_t \times n_e$ *submatrix of* $\mathbf{U}_{\mathbf{R}_t}$ *containing the* n_e *dominant eigenvector of* $\mathbf{R}_t$*, i.e.,* $\mathbf{R}_t = \mathbf{U}_{\mathbf{R}_t}\mathbf{\Lambda}_{\mathbf{R}_t}\mathbf{U}_{\mathbf{R}_t}^H$*,*
- $\mathbf{S}^{1/2} = \mathbf{D}$*,* $\mathbf{D}$ *being a real-valued diagonal matrix accounting for the power allocation.*

- Power allocation strategy follows the water-filling solution.

Channel Statistics based Precoding

Example

Let us consider the Alamouti O-STBC with two transmit antennas $(n_e = n_t = 2)$. Denoting $\mathbf{S} = \mathrm{diag}\{s_1, s_2\}$, the transmitted codewords are proportional, at the first time instant, to

$$\frac{1}{\sqrt{2}}\mathbf{U}_{\mathbf{R}_t}\mathbf{S}^{1/2}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{\sqrt{2}}\mathbf{U}_{\mathbf{R}_t}(:,1)\sqrt{s_1}c_1 + \frac{1}{\sqrt{2}}\mathbf{U}_{\mathbf{R}_t}(:,2)\sqrt{s_2}c_2$$

and, at the second time instant, to

$$\frac{1}{\sqrt{2}}\mathbf{U}_{\mathbf{R}_t}\mathbf{S}^{1/2}\begin{bmatrix} -c_2^* \\ c_1^* \end{bmatrix} = -\frac{1}{\sqrt{2}}\mathbf{U}_{\mathbf{R}_t}(:,1)\sqrt{s_1}c_2^* + \frac{1}{\sqrt{2}}\mathbf{U}_{\mathbf{R}_t}(:,2)\sqrt{s_2}c_1^*.$$

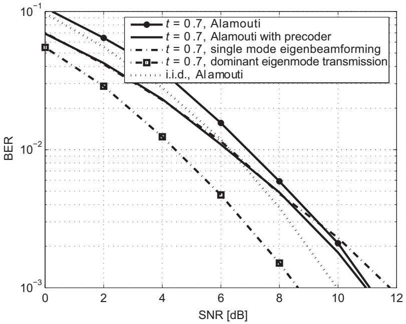
Extreme cases:

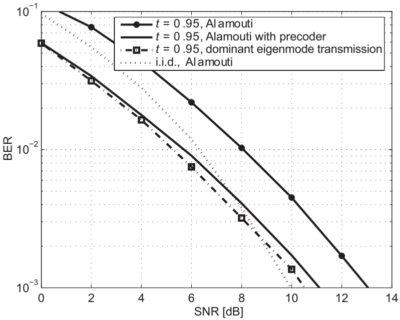
- $s_1 = s_2 = 1$: Alamouti scheme
- $s_1 = 2$, $s_2 = 0$: beamforming in the dominant eigenbeam

The precoder allocates more power to angular directions corresponding to the peaks of the transmit direction power spectrum.

Channel Statistics based Precoding

- Performance of a transmit correlation based precoded Alamouti scheme in 2×2 transmit correlated ($t = 0.7$ and $t = 0.95$) Rayleigh channels





Quantized Precoding: dominant eigenmode transmission

- Assume dominant eigenmode transmission (i.e. beamforming)

$$\begin{aligned} \mathbf{y} &= \sqrt{E_s}\mathbf{H}\mathbf{w}c + \mathbf{n}, \\ z &= \mathbf{g}^H\mathbf{y}, \\ &= \sqrt{E_s}\mathbf{g}^H\mathbf{H}\mathbf{w}c + \mathbf{g}^H\mathbf{n} \end{aligned}$$

  where $\mathbf{g}$ and $\mathbf{w}$ are respectively $n_r \times 1$ and $n_t \times 1$ vectors.
- Assuming MRC, the optimal beamforming vector $\mathbf{w}$ that maximizes the SNR is given by

$$\mathbf{w}^\star = \arg\max_{\mathbf{w}\in\mathcal{C}_w} \|\mathbf{H}\mathbf{w}\|^2$$

  with $\mathcal{C}_w$ set of unit-norm vectors. The best precoder is the dominant right singular vector of $\mathbf{H}$.
- Reduce the number of feedback bits: limit the space $\mathcal{C}_w$ over which $\mathbf{w}$ can be chosen to a codebook called $\mathcal{W}$. The receiver evaluates the best precoder $\mathbf{w}^\star$ among all unit-norm precoders $\mathbf{w}_i \in \mathcal{W}$ (with $i = 1, \ldots, n_p$) such that

$$\mathbf{w}^\star = \arg\max_{\substack{1\leq i\leq n_p \\ \mathbf{w}_i\in\mathcal{W}}} \|\mathbf{H}\mathbf{w}_i\|^2 .$$

Quantized Precoding: distortion function

- How to design the codebook? Need for a distortion function, i.e. measure of the average (over all channel realizations) array gain loss induced by the quantization process

$$d_f = \mathcal{E}_{\mathbf{H}} \left\{ \lambda_{max} - \left\| \mathbf{H}\mathbf{w}^{\star} \right\|^2 \right\}$$

- Upper-bound

$$\begin{aligned} d_f &\leq \mathcal{E}_{\mathbf{H}} \left\{ \lambda_{max} - \lambda_{max} \left| \mathbf{v}_{max}^H \mathbf{w}^{\star} \right|^2 \right\}, \\ &\stackrel{(a)}{=} \underbrace{\mathcal{E}_{\mathbf{H}} \left\{ \lambda_{max} \right\}}_{\text{quality of the channel}} \underbrace{\mathcal{E}_{\mathbf{H}} \left\{ 1 - \left| \mathbf{v}_{max}^H \mathbf{w}^{\star} \right|^2 \right\}}_{\text{quality of the codebook}} \end{aligned}$$

where $\mathbf{v}_{max}$ is the dominant right singular vector of $\mathbf{H}$. Equality (a) is only valid for i.i.d. Rayleigh fading channels.

Quantized Precoding: Lloyd

- Generalized Lloyd Algorithm:

Algorithm

For the given codebook, find the optimal quantization cells using the nearest neighbor rule. For the so-obtained quantization cells, determine that optimal quantized precoders using the centroid condition. Iterate till convergence.

- Essential conditions:
  - Assume MISO channel.
  - *centroid condition*: the optimal quantized precoder $\mathbf{w}_k$ of any quantization cell $\mathbf{R}_k$ is to be chosen as the dominant eigenvector of $\mathbf{R}_k = \mathcal{E}\left\{\mathbf{h}^H\mathbf{h} \mid \mathbf{h} \in \mathcal{R}_k\right\}$.
  - *nearest neighbor rule*: all channel vectors that are closer to the quantized precoder $\mathbf{w}_k$ are assigned to quantization cell $\mathcal{R}_k$, i.e. $\mathbf{h} \in \mathcal{R}_k$ if $\|\mathbf{h}\|^2 - |\mathbf{h}\mathbf{w}_k|^2 \leq \|\mathbf{h}\|^2 - |\mathbf{h}\mathbf{w}_j|^2$,
- Optimal codebook design for arbitrary fading channels

Quantized Precoding: Grassmannian

- Grassmannian Line Packing

Design Criterion

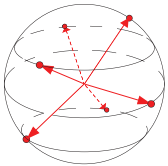
Choose a codebook $\mathcal{W}$ *made of* n_p *unit-norm vectors* $\mathbf{w}_i$ $(i = 1, \ldots, n_p)$ *such that the minimum distance*

$$\delta_{line}\left(\mathcal{W}\right) = \min_{1 \leq k < l \leq n_p} \sqrt{1 - \left|\mathbf{w}_k^H \mathbf{w}_l\right|^2},$$

is maximized.

- Problem of packing n_p lines in ^{n_t} in such a way that the minimum distance between any pair of lines is maximized.
- Close to optimal only for i.i.d. Rayleigh Fading Channels.

Grassmannian codebook



Quantized Precoding: How many bits?

- How many feedback bits $B = \log_2(n_p)$ are required? In i.i.d. channels

$$\bar{C}_{quant} \approx \mathcal{E}_{\mathbf{h}} \left\{ \log_2 \left(1 + \rho \left\| \mathbf{h} \right\|^2 \left(1 - 2^{-\frac{B}{n_t - 1}} \right) \right) \right\},$$

leading to an SNR degradation of $10 \log_{10} \left(1 - 2^{-\frac{B}{n_t - 1}} \right)$ dB relative to perfect CSIT.

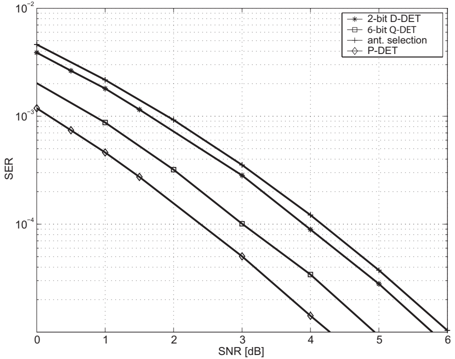
Proposition

In order to maintain a constant SNR or capacity gap between perfect CSIT and quantized feedback, it is not necessary to scale the number of feedback bits as a function of the SNR. The multiplexing gain g_s is not affected by the quality of CSIT.

- Achievable diversity gain?
  - Antenna selection (AS) is a particular case of a quantized precoding whose codebook is chosen as the columns of the identity matrix $\mathbf{I}_{n_t}$.
  - AS achieves a diversity gain of n_t.
  - Sufficient to take a full rank codebook matrix with $n_p \geq n_t$ to extract the full diversity

Quantized Precoding: Evaluations

- SER of a 3×3 MIMO system using 2-bit and 6-bit quantized BPSK-based DET.



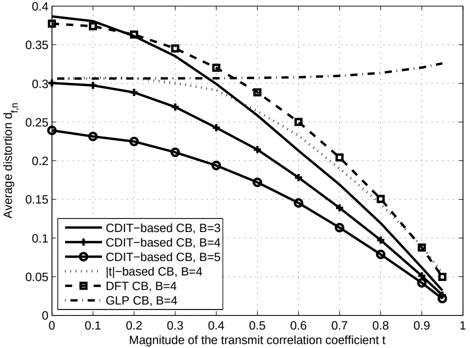
Example

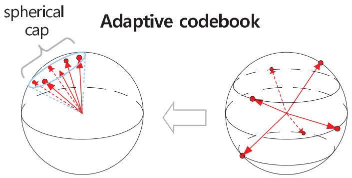
Codebook for quantized DET for $n_t = 3$ and $n_p = 4$.

$$\mathcal{W} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{j}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \\ \frac{-j}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{-j}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \\ \frac{j}{\sqrt{3}} \end{bmatrix} \right\}$$

Quantized Precoding: spatially correlated channels

- Grassmiannian only appropriate for i.i.d. Rayleigh fading.
- Spatial correlation decreases the quantization space compared to i.i.d. channels.
  - e.g. Lloyd, adaptive/CDIT-based codebook, DFT (for uniform linear arrays)
- Normalized average distortion (SNR loss) $d_{f,n} = d_f / \mathcal{E}_{\mathbf{H}} \{\lambda_{max}\}$ as a function of the codebook size $n_p = 2^B$ and the transmit correlation coefficient t with $n_t = 4$.





$$\mathcal{W}_c = \left\{ \frac{\mathbf{R}_t^{1/2}\mathbf{w}_1}{\left\|\mathbf{R}_t^{1/2}\mathbf{w}_1\right\|}, \ldots, \frac{\mathbf{R}_t^{1/2}\mathbf{w}_{n_p}}{\left\|\mathbf{R}_t^{1/2}\mathbf{w}_{n_p}\right\|} \right\}$$

Quantized Precoding: some extensions

- Extension to other kinds of channel models (e.g. spatial/time correlation, polarization), transmission strategies (e.g. O-STBCs, SM), reception strategies (e.g. MRC, ZF, MMSE, ML), criteria (e.g. error rate or transmission rate)
- Quantized precoding for SM with rank adaptation and rate maximization

$$\mathbf{W}^{\star} = \arg\max_{n_e} \max_{\mathbf{W}_i^{(n_e)} \in \mathcal{W}_{n_e}} R.$$

The codebooks $\mathcal{W}_{n_e}$ are defined for ranks $n_e = 1, \ldots, \min\{n_t, n_r\}$. Rate is computed on the equivalent precoded channel $\mathbf{H}\mathbf{W}_i^{(n_e)}$.

  - Uniform power allocation and joint ML decoding

$$R = \log_2 \det\left[\mathbf{I}_{n_e} + \frac{\rho}{n_e}\left(\mathbf{W}_i^{(n_e)}\right)^H \mathbf{H}^H \mathbf{H}\mathbf{W}_i^{(n_e)}\right].$$

  - With other types of receivers/combiner

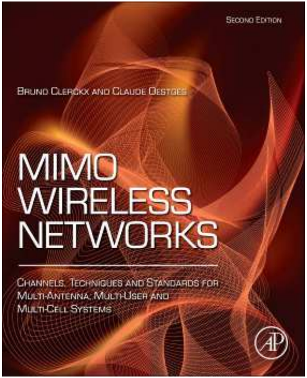
$$R = \sum_{q=1}^{n_e} \log_2\left(1 + \rho_q\left(\mathbf{H}\mathbf{W}_i^{(n_e)}\right)\right).$$

where ρ_q is the SINR of stream q on at the output of the combiner for the equivalent channel $\mathbf{H}\mathbf{W}_i^{(n_e)}$.

Frequency-Selective MIMO Channels - MIMO-OFDM

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Chapter 11
    - Section: 11.1, 11.4.1, 11.5.2

Introduction

- Two major different approaches to transmit information over frequency selective MIMO channels:
  - modulating a single carrier over the full bandwidth B.
  - converts the frequency selective channel into a set of multiple parallel flat fading channels in the frequency domain by means of Orthogonal Frequency Division Multiplexing (OFDM) modulation.

Single-Carrier Transmissions

- Transmission of a codeword $\mathbf{C} = [\mathbf{c}_0 \ldots \mathbf{c}_{T-1}]$ (of size $n_t \times T$)
- Presence of L resolvable taps, which are responsible for inter-symbol interference (ISI)

$$\mathbf{H}[\tau] = \sum_{l=0}^{L-1} \mathbf{H}[l]\delta\left(\tau - \tau_l\right)$$

  - L replicas of the same codeword $\rightarrow$ L^{th} order diversity!
- How to design codewords such that a ML decoder is able to efficiently exploit the frequency and spatial diversity without suffering from the ISI?
- System model

$$\mathbf{y}_k = \sqrt{E_s} \sum_{l=0}^{L-1} \mathbf{H}\left[l\right] \mathbf{c}_{k-l} + \mathbf{n}_k$$

  where $\mathbf{y}_k$, $\mathbf{n}_k$, E_s are defined analogous to frequency flat fading channels.

Virtual Transmit Antenna Array

- Equivalent system model

$$\mathbf{y}_k = \sqrt{E_s}\underline{\mathbf{H}} \begin{bmatrix} \mathbf{c}_k^T & \dots & \mathbf{c}_{k-L+1}^T \end{bmatrix}^T + \mathbf{n}_k$$
$$\underline{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{H}[0] & \dots & \mathbf{H}[L-1] \end{bmatrix}$$

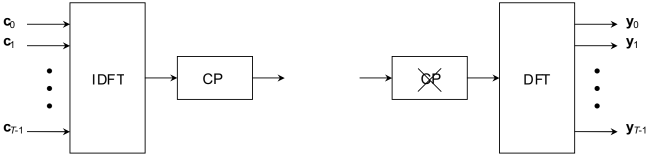
  - L taps may be thought of as virtual transmit antennas
  - Virtual transmit array of $n_t L$ antennas where the $n_r \times n_t L$ virtual channel matrix is $\underline{\mathbf{H}}$ and equivalent codeword

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_0 & \mathbf{c}_1 & \dots & \mathbf{c}_{T-1} \\ \mathbf{c}_{-1} & \mathbf{c}_0 & \dots & \mathbf{c}_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_{1-L} & \mathbf{c}_{2-L} & \dots & \mathbf{c}_{T-L} \end{bmatrix}$$

  - Maximum diversity gain of $n_r n_t L$

Multi-Carrier Transmissions: MIMO-OFDM

- Frequency domain approach to exploit the frequency diversity
- **Basic idea of OFDM:** Turn the channel matrix into a circulant matrix via the addition of a cyclic prefix to the transmitted sequence
  - A circulant matrix has the property that its left and right singular vector matrices are respectively DFT and IDFT matrices.
  - The multiplication by an IDFT matrix at the transmitter and by a DFT matrix at the receiver transforms the frequency selective channel into a diagonal matrix, whose elements are the singular values of the circulant matrix.
  - The original frequency-selective channel in the time domain becomes a set of parallel flat fading channels in the frequency domain



- This construction allows for a considerable reduction of complexity in terms of equalization and demodulation.

- **Fundamental Steps 1 to 6**

1. Apply an IDFT to the codeword $\mathbf{C}$. We obtain as output at the n^{th} time interval $(n = 0, \ldots, T-1)$,

$$\mathbf{x}_n = \frac{1}{\sqrt{T}} \sum_{k=0}^{T-1} \mathbf{c}_k e^{j \frac{2\pi}{T} kn},$$

or equivalently in a matrix form,

$$\begin{aligned} \left[\begin{array}{ccc} \mathbf{x}_0 & \ldots & \mathbf{x}_{T-1} \end{array} \right]^T &= \mathcal{D}^H \left[\begin{array}{ccc} \mathbf{c}_0 & \ldots & \mathbf{c}_{T-1} \end{array} \right]^T, \\ \left[\begin{array}{ccc} \mathbf{x}_0^T & \ldots & \mathbf{x}_{T-1}^T \end{array} \right]^T &= \left(\mathcal{D}^H \otimes \mathbf{I}_{n_t} \right) \left[\begin{array}{ccc} \mathbf{c}_0^T & \ldots & \mathbf{c}_{T-1}^T \end{array} \right]^T. \end{aligned}$$

The $T \times T$ matrix $\mathcal{D}^H$ realizes the IDFT operation. Hence, $\mathcal{D}$ is a DFT matrix reading as

$$\mathcal{D} = \frac{1}{\sqrt{T}} \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & e^{-j \frac{2\pi}{T}} & e^{-j \frac{2\pi}{T} 2} & \ldots & e^{-j \frac{2\pi}{T} (T-1)} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & e^{-j \frac{2\pi}{T} (T-2)} & e^{-j \frac{2\pi}{T} (T-2) 2} & \ldots & e^{-j \frac{2\pi}{T} (T-2)(T-1)} \\ 1 & e^{-j \frac{2\pi}{T} (T-1)} & e^{-j \frac{2\pi}{T} (T-1) 2} & \ldots & e^{-j \frac{2\pi}{T} (T-1)(T-1)} \end{bmatrix}.$$

Orthogonal Frequency Division Multiplexing (OFDM)

2. Add the guard interval vector $\mathbf{X}_g = \left[\, \mathbf{x}_{-(L-1)} \; \ldots \; \mathbf{x}_{-1} \,\right]$ of length $L-1$ in front of the codeword $\mathbf{X} = [\, \mathbf{x}_0 \; \ldots \; \mathbf{x}_{T-1} \,]$ to avoid inter-symbol interference.
   Choose the guard interval vector $\mathbf{X}_g$ in such way that $\mathbf{x}_{-n} = \mathbf{x}_{T-n}$, for $n = 1, \ldots, L-1$. Hence, the guard interval vector becomes $\mathbf{X}_g = \left[\, \mathbf{x}_{T-(L-1)} \; \ldots \; \mathbf{x}_{T-1} \,\right]$ and is commonly known as the cyclic prefix.
3. Transmit the OFDM symbol $\mathbf{X}' = \left[\, \mathbf{X}_g \quad \mathbf{X} \,\right]$ of size $n_t \times (T+L-1)$.
4. Remove the guard interval (CP) at the receiver and gather T output samples as

$$\left[\, \mathbf{r}_0^T \quad \ldots \quad \mathbf{r}_{T-1}^T \,\right]^T = \mathbf{H}_g \left[\, \mathbf{x}_{-(L-1)}^T \quad \ldots \quad \mathbf{x}_{T-1}^T \,\right]^T + \left[\, \mathbf{n}_0^T \quad \ldots \quad \mathbf{n}_{T-1}^T \,\right]^T$$

where

$$\mathbf{H}_g = \begin{bmatrix} \mathbf{H}[L-1] & \ldots & \mathbf{H}[1] & \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} & \ldots & \mathbf{0}_{n_r \times n_t} \\ \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] & \ddots & \mathbf{H}[1] & \mathbf{H}[0] & \ldots & \mathbf{0}_{n_r \times n_t} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n_r \times n_t} & \ldots & \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] & \mathbf{H}[L-2] & \ldots & \mathbf{H}[0] \end{bmatrix}$$

is a $n_r T \times n_t(T+L-1)$ matrix representing the channel seen by the OFDM symbol.

Orthogonal Frequency Division Multiplexing (OFDM)

5. Observe that the choice of the CP creates a blockwise circulant matrix $\mathbf{H}_{cp}$ of size $n_r T \times n_t T$

$$\begin{bmatrix} \mathbf{r}_0^T & \dots & \mathbf{r}_{T-1}^T \end{bmatrix}^T = \mathbf{H}_{cp} \begin{bmatrix} \mathbf{x}_0^T & \dots & \mathbf{x}_{T-1}^T \end{bmatrix}^T + \begin{bmatrix} \mathbf{n}_0^T & \dots & \mathbf{n}_{T-1}^T \end{bmatrix}^T$$

with

$$\mathbf{H}_{cp} = \begin{bmatrix} \mathbf{H}[0] & \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{0}_{n_r\times n_t} & \mathbf{H}[L-1] & \dots & \mathbf{H}[1] \\ \mathbf{H}[1] & \mathbf{H}[0] & \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{H}[2] \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{H}[L-2] & \dots & \mathbf{H}[0] & \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{0}_{n_r\times n_t} & \mathbf{H}[L-1] \\ \mathbf{H}[L-1] & \dots & \mathbf{H}[1] & \mathbf{H}[0] & \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{0}_{n_r\times n_t} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{H}[L-1] & \mathbf{H}[L-2] & \dots & \mathbf{H}[0] & \mathbf{0}_{n_r\times n_t} \\ \mathbf{0}_{n_r\times n_t} & \dots & \mathbf{0}_{n_r\times n_t} & \mathbf{H}[L-1] & \dots & \mathbf{H}[1] & \mathbf{H}[0] \end{bmatrix}.$$

SVD decomposition $\mathbf{H}_{cp} = \left(\mathcal{D}^H \otimes \mathbf{I}_{n_r}\right) \mathbf{\Lambda}_{cp} \left(\mathcal{D} \otimes \mathbf{I}_{n_t}\right)$ is such that $\mathbf{\Lambda}_{cp}$ is a block diagonal matrix whose blocks are obtained by a blockwise DFT of $\begin{bmatrix} \mathbf{H}[0] & \mathbf{H}[1] & \dots & \mathbf{H}[L-1] \end{bmatrix}$, i.e., for the $(k,k)^{\text{th}}$ block

$$\mathbf{\Lambda}_{cp}^{(kk)} = \sum_{l=0}^{L-1} \mathbf{H}[l]\ e^{-j\frac{2\pi}{T}kl}, \qquad k = 0, \dots, T-1,$$

irrespective of the channel matrix.

Orthogonal Frequency Division Multiplexing (OFDM)

6. The use of IDFT matrices at the transmitter is now clear as

$$\begin{bmatrix} \mathbf{r}_0^T & \dots & \mathbf{r}_{T-1}^T \end{bmatrix}^T = \left(\mathcal{D}^H \otimes \mathbf{I}_{n_r}\right) \boldsymbol{\Lambda}_{cp} \begin{bmatrix} \mathbf{c}_0^T & \dots & \mathbf{c}_{T-1}^T \end{bmatrix}^T + \begin{bmatrix} \mathbf{n}_0^T & \dots & \mathbf{n}_{T-1}^T \end{bmatrix}^T$$

Applying a DFT operation to the received vector, we finally obtain

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_0^T & \dots & \mathbf{y}_{T-1}^T \end{bmatrix}^T &= \left(\mathcal{D} \otimes \mathbf{I}_{n_r}\right) \begin{bmatrix} \mathbf{r}_0^T & \dots & \mathbf{r}_{T-1}^T \end{bmatrix}^T \\ &= \boldsymbol{\Lambda}_{cp} \begin{bmatrix} \mathbf{c}_0^T & \dots & \mathbf{c}_{T-1}^T \end{bmatrix}^T + \left(\mathcal{D} \otimes \mathbf{I}_{n_r}\right) \begin{bmatrix} \mathbf{n}_0^T & \dots & \mathbf{n}_{T-1}^T \end{bmatrix}^T . \end{aligned}$$

The original frequency selective channel has been converted into a set of T parallel flat fading channels in the frequency domain, the channel gains being given by the diagonal blocks of $\boldsymbol{\Lambda}_{cp}$.

Orthogonal Frequency Division Multiplexing (OFDM)

- **MIMO-OFDM System Model:** The input-output relationship on each parallel channel $k = 0, \ldots, T-1$ may be expressed without loss of generality as

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_{(k)}\mathbf{c}_k + \mathbf{n}_k$$

with

$$\mathbf{H}_{(k)} = \mathbf{\Lambda}_{cp}^{(kk)} = \sum_{l=0}^{L-1} \mathbf{H}[l] \; e^{-j\frac{2\pi}{T}kl}.$$

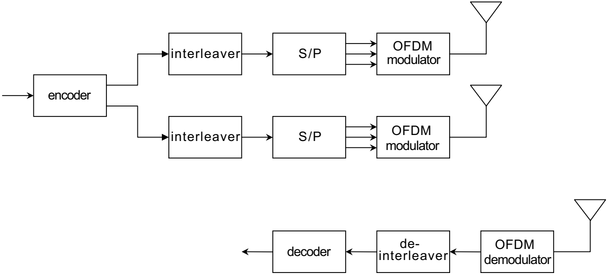
The $n_r \times 1$ vector $\mathbf{y}_k$ is the received signal to be decoded, and $\mathbf{n}_k$ is a $n_r \times 1$ zero mean complex AWGN vector with $\mathcal{E}\{\mathbf{n}_k \mathbf{n}_{k'}^H\} = \sigma_n^2 \mathbf{I}_{n_r} \delta[k-k']$.

- If ML decoding is applied, the decoder computes an estimate of the transmitted codeword according to

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \sum_{k=0}^{T-1} \left\| \mathbf{y}_k - \sqrt{E_s}\mathbf{H}_{(k)}\mathbf{c}_k \right\|^2.$$

Orthogonal Frequency Division Multiplexing (OFDM)

- Block diagram of a MIMO-OFDM system



- Strong analogy with the input-output relationship over a flat fading MIMO channel:
  - Temporal dimension replaced by frequency dimension
  - $\mathbf{H}_k \leftrightarrow \mathbf{H}_{(k)}$
  - Commonly known as Space-Frequency Coded MIMO-OFDM.
  - If the coherence bandwidth of the channel is small, the channel gains $\mathbf{H}_{(k)}$ vary significantly from tone to tone. The channel in the frequency domain can then be considered as a fast fading channel in the frequency domain.

Virtual Transmit Antenna Array

- Virtual transmit array

$$\mathbf{y}_k = \sqrt{E_s}\underline{\mathbf{H}}\begin{bmatrix} \mathbf{c}_k^T & \dots & e^{-j\frac{2\pi}{T}kl}\mathbf{c}_k^T & \dots & e^{-j\frac{2\pi}{T}k(L-1)}\mathbf{c}_k^T \end{bmatrix}^T + \mathbf{n}_k$$

where $\underline{\mathbf{H}}$ is the $n_r \times n_t L$ virtual channel matrix and the equivalent codeword is

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_0 & \dots & \mathbf{c}_k & \dots & \mathbf{c}_{T-1} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ & & e^{-j\frac{2\pi}{T}kl}\mathbf{c}_k & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_0 & \dots & e^{-j\frac{2\pi}{T}k(L-1)}\mathbf{c}_k & \dots & e^{-j\frac{2\pi}{T}(T-1)(L-1)}\mathbf{c}_{T-1} \end{bmatrix}$$

- Equivalent codewords differ from the single-carrier case.
- Maximum diversity gain of $n_r n_t L$.

Unified Representation for Single and Multi-Carrier Transmissions

- Unique virtual $n_r \times Ln_t$ MIMO channel

$$\mathbf{Y} = \left[\begin{array}{ccc} \mathbf{y}_0 & \cdots & \mathbf{y}_{T-1} \end{array}\right] = \sqrt{E_s}\underline{\mathbf{H}}\ \underline{\mathbf{C}} + \left[\begin{array}{ccc} \mathbf{n}_0 & \cdots & \mathbf{n}_{T-1} \end{array}\right]$$

- Equivalent transmitted codewords in the virtual $n_r \times Ln_t$ MIMO representation

$$\underline{\mathbf{C}} = \left[\begin{array}{c} \mathbf{C}_{(0)} \\ \vdots \\ \mathbf{C}_{(L-1)} \end{array}\right]$$

  - Space-frequency coded MIMO-OFDM

$$\left[\begin{array}{c} \mathbf{C}_{(0)} \\ \vdots \\ \mathbf{C}_{(L-1)} \end{array}\right] = \left[\begin{array}{c} \mathbf{C}\mathbf{D}_{(0)} \\ \vdots \\ \mathbf{C}\mathbf{D}_{(L-1)} \end{array}\right] = \left[\mathbf{I}_L \otimes \mathbf{C}\right] \underbrace{\left[\begin{array}{c} \mathbf{D}_{(0)} \\ \vdots \\ \mathbf{D}_{(L-1)} \end{array}\right]}_{\underline{\mathbf{D}}}$$

  with $\mathbf{D}_{(l)} = \text{diag}\left\{1, \ldots, e^{-j\frac{2\pi}{T}kl}, \ldots, e^{-j\frac{2\pi}{T}(T-1)l}\right\}$.

  - Single carrier transmissions,

$$\mathbf{C}_{(l)}(m,k) = \mathbf{c}_{k-l}(m,1)\,, \quad k = 0, \ldots, T-1,\ m = 1, \ldots, n_t.$$

- Maximum-likelihood (ML) decoding: $\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \left\|\mathbf{Y} - \sqrt{E_s}\underline{\mathbf{H}}\ \underline{\mathbf{C}}\right\|^2$

Capacity of Frequency Selective MIMO Channels

- MIMO frequency selective channels
  - Mutual information obtained by an integration over the frequency band of interest B

$$\mathcal{I}_{FS}\left(\{\mathbf{H}(f)\}_f, \{\mathbf{Q}(f)\}_f\right) = \frac{1}{B}\int_B \log_2 \det\left[\mathbf{I}_{n_r} + \rho(f)\mathbf{H}(f)\mathbf{Q}(f)\mathbf{H}(f)^H\right] df$$

  subject to $\int_B \mathrm{Tr}\{\mathbf{Q}(f)\} = P_B$.
  - Capacity

$$C_{CSIT,FS} = \max_{\int_B \mathrm{Tr}\{\mathbf{Q}(f)\}=P_B} \mathcal{I}_{FS}\left(\{\mathbf{H}(f)\}_f, \{\mathbf{Q}(f)\}_f\right).$$

- MIMO-OFDM (neglecting the loss in spectral efficiency due to the cyclic prefix):
  - Mutual information

$$\mathcal{I}_{FS}\left(\{\mathbf{H}_{(k)}\}_k, \{\mathbf{Q}_{(k)}\}_k\right) = \frac{1}{T}\sum_{k=0}^{T-1}\mathcal{I}_{(k)} = \frac{1}{T}\sum_{k=0}^{T-1}\log_2\det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}_{(k)}\mathbf{Q}_{(k)}\mathbf{H}_{(k)}^H\right]$$

  subject to $\sum_{k=0}^{T-1}\mathrm{Tr}\left\{\mathbf{Q}_{(k)}\right\} = T$.
  - Capacity

$$C_{CSIT,FS} = \frac{1}{T}\max_{\sum_{k=0}^{T-1}\mathrm{Tr}\{\mathbf{Q}_{(k)}\}=T}\sum_{k=0}^{T-1}\log_2\det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}_{(k)}\mathbf{Q}_{(k)}\mathbf{H}_{(k)}^H\right]$$

$$= \frac{1}{T}\max_{\sum_{k=0}^{T-1}\sum_{l=1}^{n} s_{(k),l}=T}\sum_{k=0}^{T-1}\sum_{l=1}^{n}\log_2\left[1 + \rho s_{(k),l}\lambda_{(k),l}\right].$$

  Solved using a space-frequency water-filling.

Average Pairwise Error Probability

- Conditional pairwise error probability (PEP)

$$P\left(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}} | \underline{\mathbf{H}}\right) = \mathcal{Q}\left(\sqrt{\frac{E_s}{2\sigma_n^2}\left\|\underline{\mathbf{H}}\left(\underline{\mathbf{C}} - \underline{\mathbf{E}}\right)\right\|_F^2}\right)$$

- Average PEP in Rayleigh slow fading channels

$$P\left(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}}\right) = \frac{1}{\pi}\int_0^{\pi/2}\left(\det\left(\mathbf{I}_{Tn_r} + \eta \underline{\mathbf{C}}_{\underline{\mathbf{R}}}\right)\right)^{-1} d\beta$$

where

$$\underline{\mathbf{C}}_{\underline{\mathbf{R}}} = \left(\mathbf{I}_{n_r} \otimes \left(\underline{\mathbf{C}} - \underline{\mathbf{E}}\right)^H\right)\underline{\mathbf{R}}\left(\mathbf{I}_{n_r} \otimes \left(\underline{\mathbf{C}} - \underline{\mathbf{E}}\right)\right)$$

$$\underline{\mathbf{R}} = \mathcal{E}\left\{\text{vec}\left(\underline{\mathbf{H}}^H\right)\text{vec}\left(\underline{\mathbf{H}}^H\right)^H\right\}$$

Assuming full rank $\underline{\mathbf{R}}$, full diversity at high SNR if $r\left(\underline{\mathbf{C}}_{\underline{\mathbf{R}}}\right) = n_r n_t L$.

- If each tap is spatially i.i.d. Rayleigh distributed with an average power β_l and if there is no correlation between taps $\underline{\mathbf{R}} = \mathbf{I}_{n_r} \otimes \text{diag}\left\{\beta_0, \ldots, \beta_{L-1}\right\} \otimes \mathbf{I}_{n_t}$,

$$P\left(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}}\right) = \frac{1}{\pi}\int_0^{\pi/2}\left[\det\left(\mathbf{I}_{Ln_t} + \eta\left[\text{diag}\left\{\beta_0, \ldots, \beta_{L-1}\right\} \otimes \mathbf{I}_{n_t}\right]\underline{\tilde{\mathbf{E}}}\right)\right]^{-n_r} d\beta$$

where $\underline{\tilde{\mathbf{E}}} \triangleq \left(\underline{\mathbf{C}} - \underline{\mathbf{E}}\right)\left(\underline{\mathbf{C}} - \underline{\mathbf{E}}\right)^H$. Full diversity at high SNR if $r\left(\underline{\tilde{\mathbf{E}}}\right) = n_t L$.

Code Design for Single-Carrier Transmissions

Example

Send $c_0, \ldots, c_{T-1}$ over two antennas in a 2-tap Rayleigh fading channel.

- classical delay-diversity scheme

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & \ldots & c_{T-1} & 0 & 0 \\ 0 & c_0 & c_1 & \ldots & c_{T-1} & 0 \end{bmatrix},$$

$$\underline{\mathbf{C}} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & \ldots & c_{T-1} & 0 & 0 \\ 0 & c_0 & c_1 & \ldots & c_{T-1} & 0 \\ 0 & c_0 & c_1 & \ldots & c_{T-1} & 0 \\ 0 & 0 & c_0 & c_1 & \ldots & c_{T-1} \end{bmatrix}.$$

Diversity of $3n_r$ only!

- Generalized delay-diversity scheme

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & \ldots & c_{T-1} & 0 & 0 & 0 \\ 0 & 0 & c_0 & c_1 & \ldots & c_{T-1} & 0 \end{bmatrix},$$

$$\underline{\mathbf{C}} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & \ldots & c_{T-1} & 0 & 0 & 0 \\ 0 & 0 & c_0 & c_1 & \ldots & c_{T-1} & 0 \\ 0 & c_0 & c_1 & \ldots & c_{T-1} & 0 & 0 \\ 0 & 0 & 0 & c_0 & c_1 & \ldots & c_{T-1} \end{bmatrix}.$$

Diversity of $4n_r$!

Code Design for Space-Frequency Coded MIMO-OFDM

- Diversity gain
  - Define $l_{\mathbf{C},\mathbf{E}}(l) = \sharp\tau_{\mathbf{C},\mathbf{E}}(l)$ $(l = 1, \ldots, n_t)$ with $\tau_{\mathbf{C},\mathbf{E}}(l) = \{k \mid \mathbf{c}_k(l) - \mathbf{e}_k(l) \neq 0\}$.
  - Define $l_{\mathbf{C},\mathbf{E}} = \sharp\tau_{\mathbf{C},\mathbf{E}}$ with $\tau_{\mathbf{C},\mathbf{E}} = \{k \mid \mathbf{c}_k - \mathbf{e}_k \neq 0\}$.

Proposition

For full rank space-tap correlation matrix $\underline{\mathbf{R}}$, a pair of space-frequency codewords $\{\mathbf{C}, \mathbf{E}\}$ with an effective length $l_{\mathbf{C},\mathbf{E}}$, effective lengths $\{l_{\mathbf{C},\mathbf{E}}(l)\}_{l=1}^{n_t}$ and a rank $r(\tilde{\mathbf{E}})$ achieves the full diversity $n_t n_r L$ if

$$\begin{aligned} r(\tilde{\mathbf{E}}) &= n_t, \\ l_{\mathbf{C},\mathbf{E}}(l) &\geq L, \ \forall l = 1, \ldots, n_t, \\ l_{\mathbf{C},\mathbf{E}} &\geq n_t L. \end{aligned}$$

Code Design for Space-Frequency Coded MIMO-OFDM

- Coding gain
  - Assume each tap l is i.i.d. Rayleigh distributed with an average power β_l and no correlation between taps

$$\underline{\mathbf{C}_{\mathbf{R}}} = \mathbf{R}_f \odot \left[\mathbf{1}_{n_r \times n_r} \otimes \left((\mathbf{C} - \mathbf{E})^H (\mathbf{C} - \mathbf{E}) \right) \right],$$

$$\mathbf{R}_f = \mathbf{I}_{n_r} \otimes \mathbf{R}_F, \qquad \text{(space-frequency correlation matrix)}$$

$$\mathbf{R}_F = \sum_{l=0}^{L-1} \beta_l \mathbf{d}_{(l)}^H \mathbf{d}_{(l)} \quad \text{with} \quad \mathbf{d}_{(l)} = \begin{bmatrix} 1 & \ldots & e^{-j2\pi kl/T} & \ldots & e^{-j2\pi(T-1)l/T} \end{bmatrix}$$

Example

SISO channel $h_{(k)} = \sum_{l=0}^{L-1} h[l]\, e^{-j\frac{2\pi}{T}kl}$. Frequency correlation between channel on subcarrier k and $k+K$

$$\mathcal{E}\left\{ h_{(k)} h_{(k+K)}^* \right\} = \sum_{l=0}^{L-1} \underbrace{\mathcal{E}\left\{ |h[l]|^2 \right\}}_{\beta_l} e^{j\frac{2\pi}{T}lK}$$

Code Design for Space-Frequency Coded MIMO-OFDM

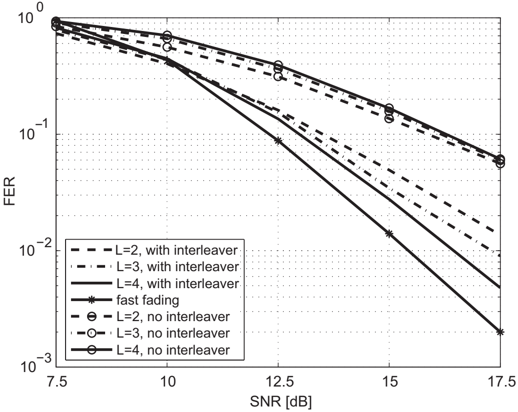
- Impact of Frequency Correlation on the Coding Gain

$$\begin{aligned}
\det\left(\mathbf{I}_{Tn_r} + \eta \underline{\mathbf{C}}_{\underline{\mathbf{R}}}\right) &= \det\left(\mathbf{I}_{Tn_r} + \eta \mathbf{R}_f \odot \left[\mathbf{1}_{n_r \times n_r} \otimes \left((\mathbf{C}-\mathbf{E})^H (\mathbf{C}-\mathbf{E})\right)\right]\right) \\
&= \left(\det\left(\mathbf{I}_T + \eta \left[\mathbf{R}_F \odot \left((\mathbf{C}-\mathbf{E})^H (\mathbf{C}-\mathbf{E})\right)\right]\right)\right)^{n_r} \\
&\leq \prod_{k=0}^{T-1} \left(1 + \eta \left[\sum_{l=0}^{L-1} \beta_l\right] \|\mathbf{c}_k - \mathbf{e}_k\|^2\right)^{n_r}
\end{aligned}$$

- MIMO-OFDM $\approx$ narrowband MIMO transmissions over i.i.d. fast fading Rayleigh channels if $\mathbf{R}_f$ is diagonal ($L >> 0$).
- Frequency correlation reduces the achievable coding gain. Reduce the frequency correlation between adjacent tones by means of an interleaver.

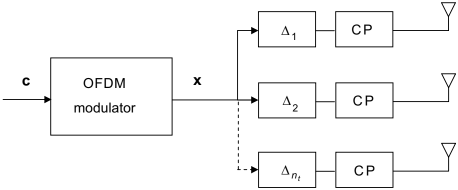
Code Design for Space-Frequency Coded MIMO-OFDM

- With interleaver, use codes with large effective length and product distance for OFDM transmissions as in fast fading channels.
- FER of the 16-state space-time trellis code for $L = 2$, 3 and 4 in uniformly distributed i.i.d. Rayleigh channels with and without interleaver.



Code Design for Space-Frequency Coded MIMO-OFDM

- Space-Frequency Linear Block Coding
  - Orthogonal codes: O-STBC $\rightarrow$ O-SFBC (time replaced by frequency)
  - Make sure that O-SFBC is operated on adjacent subcarriers. Recall that the channel has to be constant within a O-STBC/O-SFBC block!
  - In practice, O-SFBC often preferred over O-STBC.
- Cyclic Delay Diversity
  - Adaptation of the generalized delay-diversity (GDD) scheme to OFDM systems.
  - Send on each antenna a circularly shifted version of the same OFDM symbol in the time domain. Hence, the temporal delay introduced on each antenna in the GDD scheme is transformed into a cyclic delay in the CDD scheme.



1. A sequence $\mathbf{c}$ of symbols c_k with $k = 0, \ldots, T-1$ is OFDM modulated.
2. The output sequence $\mathbf{x}$ is transmitted on each antenna with a cyclic delay Δ_m, $m = 1, \ldots, n_t$ so that the output symbol on antenna m $(m = 1, \ldots, n_t)$ at time n $(n = 0, \ldots, T-1)$ is given by $x_{(n-\Delta_m) \bmod T}$.
3. Finally, CP is added on each antenna, analogous to conventional OFDM transmissions.
4. At the receiver, the cyclic prefix is removed, and OFDM demodulation and decoding are performed.

Code Design for Space-Frequency Coded MIMO-OFDM

- Analogous to GDD, CC converts a MIMO channel into a SIMO channel with enhanced frequency selectivity. The subsequent frequency diversity is extracted by appropriate outer codes.
- A cyclic shift in the time domain corresponds to the multiplication by a phase shift in the frequency domain. Therefore, the received signal in the frequency domain reads as

$$\mathbf{y}_k = \sqrt{\frac{E_s}{n_t}}\mathbf{h}_{eq,(k)} c_k + \mathbf{n}_k$$

  where the equivalent SIMO channel matrix on the k^{th} tone, denoted as $\mathbf{h}_{eq,(k)}$, is given by

$$\mathbf{h}_{eq,(k)} = \sum_{m=1}^{n_t} \mathbf{H}_{(k)}(:,m)\; e^{-j\frac{2\pi}{T}k\Delta_m}$$

  and $\mathbf{H}_{(k)}$ is the DFT of the impulse response evaluated on the k^{th} subcarrier $\mathbf{H}_{(k)} = \mathbf{\Lambda}_{cp}^{(kk)} = \sum_{l=0}^{L-1} \mathbf{H}\left[l\right]\; e^{-j\frac{2\pi}{T}kl}$.

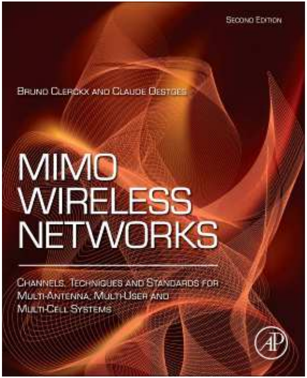
Code Design for Space-Frequency Coded MIMO-OFDM

- CDD vs. GDD vs. O-SFBC
  - over GDD: reduced guard interval.
  - over O-SFBC/O-STBC: increased flexibility and scalability to any n_t, no rate loss if $n_t > 2$, no requirement on constant channel over several tones, unlike in O-SFBC and O-STBC.
  - CDD receiver is essentially the same as a classical SIMO receiver.
  - the number of states of the outer code necessary to exploit the full diversity is much larger with CDD than with O-SFBC.
- Precoder cycling
  - Codevector on subcarrier k writes as $\mathbf{c}_k = \mathbf{w}_k c_k$ where $\mathbf{w}_k$ is the precoding vector and c_k is a complex symbol.
  - The precoder changes every M contiguous physical subcarriers.
  - Appropriate design (using Grassmanian Line Packing) of the precoders converts the MIMO channel into a frequency selective SIMO channel.

Multi-User MIMO - Multiple Access Channels (Uplink) & Broadcast Channels (Downlink)

Reference Book

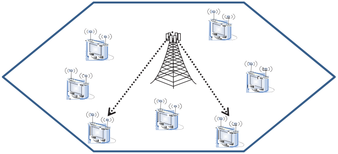
- Bruno Clerckx and Claude Oestges, “MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems,” Academic Press (Elsevier), Oxford, UK, Jan 2013.



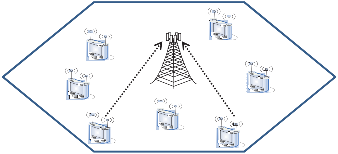
– Chapter 12
  - Section: 12.1, 12.2, 12.3, 12.4

Introduction

- So far, we looked at a single link/user. Most systems are multi-user!
- How to deal with multiple users? What is the benefit of MIMO in a multi-user setting?
- MIMO Broadcast Channel (BC) and Multiple Access Channel(MAC)



(a) Broadcast Channel - Downlink



(b) Multiple Access Channel - Uplink

Differences between BC and MAC:

- there are multiple independent receivers (and therefore multiple independent additive noises) in BC while there is a single receiver (and therefore a single noise term) in MAC.
- there is a single transmitter (and therefore a single transmit power constraint) in BC while there are multiple transmitters (and therefore multiple transmit power constraints) in MAC.
- the desired signal and the interference (originating from the co-scheduled signals) propagate through the same channel in the BC while they propagate through different channels in the MAC.

MIMO MAC System Model

- Uplink multi-user MIMO (MU-MIMO) transmission
  - total number of K users ($\mathcal{K} = \{1, \ldots, K\}$) distributed in a cell,
  - $n_{t,q}$ transmit antennas at mobile terminal q (we simply drop the index q and write n_t if $n_{t,q} = n_t \ \forall q$)
  - n_r receive antenna at the base station
- Received signal (we drop the time dimension)

$$\mathbf{y}_{ul} = \sum_{q=1}^{K} \Lambda_q^{-1/2} \mathbf{H}_{ul,q} \mathbf{c}'_{ul,q} + \mathbf{n}_{ul}$$

where

  - $\mathbf{y}_{ul} \in {}^{n_r}$
  - $\mathbf{H}_{ul,q} \in {}^{n_r \times n_{t,q}}$ models the small scale time-varying fading process and Λ_q^{-1} refers to the large-scale fading accounting for path loss and shadowing
  - $\mathbf{n}_{ul}$ is a complex Gaussian noise $\mathcal{CN}\left(0, \sigma_n^2 \mathbf{I}_{n_r}\right)$.
- User q's input covariance matrix is defined as the covariance matrix of the transmit signal of user q as $\mathbf{Q}_{ul,q} = \mathcal{E}\left\{\mathbf{c}'_{ul,q} {\mathbf{c}'_{ul,q}}^H\right\}$.
- Power constraint: Tr$\{\mathbf{Q}_{ul,q}\} \leq E_{s,q}$.

MIMO MAC System Model

- By stacking up the transmit signal vectors and the channel matrices of all K users,

$$\mathbf{c}'_{ul} = \left[\mathbf{c}'^{T}_{ul,1}, \ldots, \mathbf{c}'^{T}_{ul,K}\right]^T,$$

$$\mathbf{H}_{ul} = \left[\Lambda_1^{-1/2}\mathbf{H}_{ul,1}, \ldots, \Lambda_K^{-1/2}\mathbf{H}_{ul,K}\right],$$

  the system model also writes as

$$\mathbf{y}_{ul} = \mathbf{H}_{ul}\mathbf{c}'_{ul} + \mathbf{n}_{ul}.$$

  $\mathbf{H}_{ul}$ is assumed to be full-rank as it would be the case in a typical user deployment.
- Long term SNR of user q defined as $\eta_q = E_{s,q}\Lambda_q^{-1}/\sigma_n^2$.
- Note on the notations: the dependence on the path loss and shadowing is made explicit in order to stress that the co-scheduled users experience different path losses and shadowings and therefore receive power.
- We assume that the receiver (i.e. the BS in a UL scenario) has always perfect knowledge of the CSI, but we will consider strategies where the transmitters have perfect or partial knowledge of the CSI.

Capacity Region of Deterministic Channels

- In a multi-user setup, given that all users share the same spectrum, the rate achievable by a given user q, denoted as R_q, will depend on the rate of the other users R_p, $p \neq q \rightarrow$ Trade-off between rates achievable by different users!
- The capacity region $\mathcal{C}$ formulates this trade-off by expressing the set of all user rates $(R_1, \ldots, R_K)$ that are simultaneously achievable.

Definition

The capacity region $\mathcal{C}$ of a channel $\mathbf{H}_{ul}$ is the set of all rate vectors $(R_1, \ldots, R_K)$ such that simultaneously user 1 to user K can reliably communicate at rate R_1 to rate R_K, respectively.

Any rate vector not in the capacity region is not achievable (i.e. transmission at those rates will lead to errors).

Definition

The sum-rate capacity C of a capacity region $\mathcal{C}$ is the maximum achievable sum of rates

$$C = \max_{(R_1, \ldots, R_K) \in \mathcal{C}} \sum_{q=1}^{K} R_q.$$

Rate Region of MIMO MAC

- For given input covariance matrices $\mathbf{Q}_{ul,1}, \ldots, \mathbf{Q}_{ul,K}$, the achievable rate region is defined by
  1. The rate achievable by a given user q *with a given transmit strategy* $\mathbf{Q}_{ul,q}$ cannot be larger than its achievable rate in a single-user setup, i.e.

$$R_q \leq \log_2 \det \left[\mathbf{I}_{n_r} + \frac{\Lambda_q^{-1}}{\sigma_n^2} \mathbf{H}_{ul,q} \mathbf{Q}_{ul,q} \mathbf{H}_{ul,q}^H \right], \quad q = 1, \ldots, K$$

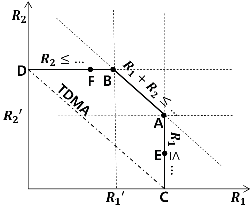
     where $\mathbf{Q}_{ul,q} = \mathcal{E}\{\mathbf{c}'_q \mathbf{c}'^H_q\}$ is subject to the power constraint $\mathrm{Tr}\{\mathbf{Q}_{ul,q}\} \leq E_{s,q}$.
  2. The sum of the rates achievable by a subset S of the users should be smaller than the total rate achievable when those users "cooperate" with each other to form a giant array with $n_{t,S} = \sum_{q \in S} n_{t,q}$ transmit antennas subject to their respective power constraints, i.e.

$$\begin{aligned} \sum_{q \in S} R_q &\leq \log_2 \det \left[\mathbf{I}_{n_r} + \frac{1}{\sigma_n^2} \mathbf{H}_{ul,S} \mathbf{Q}_{ul,S} \mathbf{H}_{ul,S}^H \right] \\ &= \log_2 \det \left[\mathbf{I}_{n_r} + \frac{1}{\sigma_n^2} \sum_{q \in S} \Lambda_q^{-1} \mathbf{H}_{ul,q} \mathbf{Q}_{ul,q} \mathbf{H}_{ul,q}^H \right], \end{aligned}$$

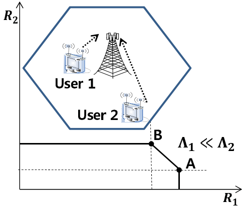
     with $\mathbf{H}_{ul,S} = \left[\Lambda_i^{-1/2} \mathbf{H}_{ul,i}, \ldots, \Lambda_j^{-1/2} \mathbf{H}_{ul,j} \right]_{i,j \in S}$,
     $\mathbf{Q}_{ul,S} = \mathrm{diag}\left\{ \mathbf{Q}_{ul,i}, \ldots, \mathbf{Q}_{ul,j} \right\}_{i,j \in S}$, subject to the constraints $\mathrm{Tr}\{\mathbf{Q}_{ul,q}\} \leq E_{s,q}$.
- The rate region looks like a K-dimensional polyhedron with $K!$ corner points on the boundary.

Rate Region of a Two-User MIMO MAC

- This rate region is a pentagon with two corner points A and B.



(a) Two-user MIMO MAC rate region for fixed $Q_{ul,1}$ and $Q_{ul,2}$



(b) Rate regions with various path losses

- Remarkably, at point A, user 1 can transmit at a rate equal to its single-link MIMO rate and user 2 can simultaneously transmit at a rate $R_2' > 0$ equal to

$$
\begin{aligned}
R_2' &= \log_2 \det \left[\mathbf{I}_{n_r} + \frac{\Lambda_1^{-1}}{\sigma_n^2} \mathbf{H}_{ul,1} \mathbf{Q}_{ul,1} \mathbf{H}_{ul,1}^H + \frac{\Lambda_2^{-1}}{\sigma_n^2} \mathbf{H}_{ul,2} \mathbf{Q}_{ul,2} \mathbf{H}_{ul,2}^H \right] \\
&\quad - \log_2 \det \left[\mathbf{I}_{n_r} + \frac{\Lambda_1^{-1}}{\sigma_n^2} \mathbf{H}_{ul,1} \mathbf{Q}_{ul,1} \mathbf{H}_{ul,1}^H \right] \\
&= \log_2 \det \left[\mathbf{I}_{n_r} + \frac{\Lambda_2^{-1}}{\sigma_n^2} \mathbf{H}_{ul,2} \mathbf{Q}_{ul,2} \mathbf{H}_{ul,2}^H \left(\mathbf{I}_{n_r} + \frac{\Lambda_1^{-1}}{\sigma_n^2} \mathbf{H}_{ul,1} \mathbf{Q}_{ul,1} \mathbf{H}_{ul,1}^H \right)^{-1} \right].
\end{aligned}
$$

Capacity Region of MIMO MAC

- We have assumed so far specific input covariance matrices.
  - A different choice of the beamforming matrix and the power allocation leads to a different transmit strategy $\mathbf{Q}_{ul,q}$ and generally a different shape of the pentagon (or more generally the K-dimensional polyhedron).
  - The trade-off between user rates is therefore affected by the choice of the input covariance matrices.
  - The optimal set of input covariance matrices that maximizes the sum-rate can be found using a generalization of the single-link water-filling solution (Detail in the book).
- The capacity region is equal to the union (over all transmit strategies satisfying the power constraints) of all the K-dimensional polyhedrons.

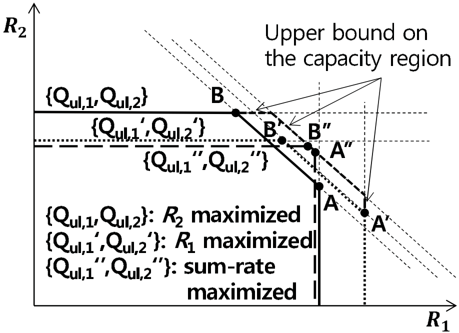
Proposition

The capacity region $\mathcal{C}_{MAC}$ of the Gaussian MIMO MAC for a determinsitic channel $\mathbf{H}_{ul}$ is the union of all achievable rate vectors $(R_1, \ldots, R_K)$ given by

$$\bigcup_{\substack{\mathrm{Tr}\left\{\mathbf{Q}_{ul,q}\right\} \leq E_{s,q} \\ \mathbf{Q}_{ul,q} \succeq 0, \forall q}} \left\{ \begin{array}{l} (R_1, \ldots, R_K) : \sum_{q \in S} R_q \leq \\ \log_2 \det \left[\mathbf{I}_{n_r} + \sum_{q \in S} \frac{\Lambda_q^{-1}}{\sigma_n^2} \mathbf{H}_{ul,q} \mathbf{Q}_{ul,q} \mathbf{H}_{ul,q}^H \right], \forall S \subseteq \mathcal{K} \end{array} \right\}.$$

Capacity Region of a Two-User MIMO MAC

- Due to the union of pentagons, the capacity region of the two-user MIMO MAC does not look like a pentagon in general.



(c) Bounds on the capacity region

- However, with a single antenna ($n_{t,q} = 1$), the capacity region remains a pentagon because a single data is transmitted per user at the full power, i.e. $E_{s,q}$.

Capacity Region of SISO MAC

Corollary

$$\mathcal{C}_{MAC} = \left\{ (R_1, \ldots, R_K) : \sum_{q \in S} R_q \leq \log_2 \left(1 + \sum_{q \in S} \eta_q \left|h_{ul,q}\right|^2 \right), \forall S \subseteq \mathcal{K} \right\}$$

where $\eta_q = \Lambda_q^{-1} E_{s,q} / \sigma_n^2$.

Example

Two-user SISO: $\mathcal{C}_{MAC}$ is the set of all rates pair (R_1, R_2) satisfying to

$$R_q \leq \log_2 \left(1 + \eta_q \left|h_{ul,q}\right|^2\right), q = 1, 2$$
$$R_1 + R_2 \leq \log_2 \left(1 + \eta_1 \left|h_{ul,1}\right|^2 + \eta_2 \left|h_{ul,2}\right|^2\right).$$

$$R_2' = \log_2 \left(1 + \eta_1 \left|h_{ul,1}\right|^2 + \eta_2 \left|h_{ul,2}\right|^2\right) - \log_2 \left(1 + \eta_1 \left|h_{ul,1}\right|^2\right)$$
$$= \log_2 \left(1 + \frac{\eta_2 \left|h_{ul,2}\right|^2}{1 + \eta_1 \left|h_{ul,1}\right|^2}\right) = \log_2 \left(1 + \frac{\Lambda_2^{-1} \left|h_{ul,2}\right|^2 E_{s,2}}{\sigma_n^2 + \Lambda_1^{-1} \left|h_{ul,1}\right|^2 E_{s,1}}\right).$$

Capacity Region of SIMO MAC

Corollary

$$\mathcal{C}_{MAC} = \left\{ \begin{array}{l} (R_1, \ldots, R_K) : \sum_{q \in S} R_q \leq \\ \log_2 \det \left[\mathbf{I}_{n_r} + \sum_{q \in S} \eta_q \mathbf{h}_{ul,q} \mathbf{h}_{ul,q}^H \right], \forall S \subseteq \mathcal{K} \end{array} \right\}$$

where $\eta_q = \Lambda_q^{-1} E_{s,q} / \sigma_n^2$.

Example

Two-user SIMO: $\mathcal{C}_{MAC}$ is the set of all rates pair (R_1, R_2) satisfying to

$$R_q \leq \log_2 \left(1 + \eta_q \left\| \mathbf{h}_{ul,q} \right\|^2\right) = \log_2 \det \left(\mathbf{I}_{n_r} + \eta_q \mathbf{h}_{ul,q} \mathbf{h}_{ul,q}^H \right), q = 1, 2$$

$$R_1 + R_2 \leq \log_2 \det \left(\mathbf{I}_{n_r} + \eta_1 \mathbf{h}_{ul,1} \mathbf{h}_{ul,1}^H + \eta_2 \mathbf{h}_{ul,2} \mathbf{h}_{ul,2}^H \right).$$

$$\begin{aligned} R_2' &= \log_2 \det \left(\mathbf{I}_{n_r} + \eta_1 \mathbf{h}_{ul,1} \mathbf{h}_{ul,1}^H + \eta_2 \mathbf{h}_{ul,2} \mathbf{h}_{ul,2}^H \right) - \log_2 \det \left(\mathbf{I}_{n_r} + \eta_1 \mathbf{h}_{ul,1} \mathbf{h}_{ul,1}^H \right) \\ &= \log_2 \det \left(\mathbf{I}_{n_r} + \eta_2 \mathbf{h}_{ul,2} \mathbf{h}_{ul,2}^H \left(\mathbf{I}_{n_r} + \eta_1 \mathbf{h}_{ul,1} \mathbf{h}_{ul,1}^H \right)^{-1} \right) \\ &= \log_2 \left(1 + \eta_2 \mathbf{h}_{ul,2}^H \left(\mathbf{I}_{n_r} + \eta_1 \mathbf{h}_{ul,1} \mathbf{h}_{ul,1}^H \right)^{-1} \mathbf{h}_{ul,2} \right) \end{aligned}$$

Achievability of the Capacity Region

- For $n_t = 1$, the SIMO MAC architecture is reminiscent of the Spatial Multiplexing architecture discussed for a single-link MIMO channel.
- We can therefore fully reuse the various receiver architectures derived for single-link MIMO.
- Recall the optimality of the MMSE V-BLAST (also called Spatial Multiplexing with MMSE-SIC receiver)

Proposition

MMSE-SIC is optimal for achieving the corner points of the MIMO MAC rate region.

- The exact corner point that is achieved on the rate region depends on the stream cancellation ordering:
  - Point A, user 2 is canceled first (i.e. all streams from user 2) such that user 1 is left with the Gaussian noise and can achieve a rate equal to the single-link bound.
  - Assuming $n_t = 1$, $R_2' = \log_2(1 + \rho_q)$ where ρ_q is the SINR of the MMSE receiver for user 2's stream treating user 1's stream as colored Gaussian interference.

Comparisons with TDMA

- TDMA allocates the time resources in an orthogonal manner such that users are never transmitting at the same time (line D-C in the rate region).
- SISO: both TDMA and SIC exploit a single degree of freedom but TDMA rate region is strictly smaller than the one achievable with SIC.
- SIMO: TDMA incurs a big loss compared to SIMO MAC (with MMSE-SIC) as it only exploits a single degree of freedom despite the presence of $\min\{n_r, K\}$ degrees of freedom achievable with SIMO MAC at high SNR.
- MIMO: As n_t increases, the gap between the TDMA and MIMO MAC rate regions decreases.

Ergodic Capacity Region of Fast Fading Channels: Perfect CSIT

- The ergodic capacity region is the set of achievable long-term average rates $R_1, \ldots, R_K$ where the averaging is taken w.r.t. all channel realizations.
- The rate region can therefore be extended to fast fading channels as

$$\sum_{q \in S} R_q \leq \mathcal{E}\left\{\log_2 \det\left[\mathbf{I}_{n_r} + \frac{1}{\sigma_n^2}\sum_{q \in S} \Lambda_q^{-1}\mathbf{H}_{ul,q}\mathbf{Q}_{ul,q}\mathbf{H}_{ul,q}^H\right]\right\}, \quad \forall S \subseteq \mathcal{K}$$

  where the input covariance matrices are subject to power constraints.
- short-term power constraint: $\text{Tr}\{\mathbf{Q}_{ul,q}\} = E_{s,q}$
  - similar to deterministic channels
- long-term power constraint: $\mathcal{E}\left\{\text{Tr}\{\mathbf{Q}_{ul,q}\}\right\} = E_{s,q}$ where the average power is computed over a duration $T_p >> T$
  - complicated scenario
  - $\mathbf{Q}_{ul,q}$ and its trace change according to the channel gain subject to the constraint that the average $\text{Tr}\{\mathbf{Q}_{ul,q}\}$ over a duration T_p should equal $E_{s,q}$.
  - change defined by a power control policy that maps a channel realization to $\text{Tr}\{\mathbf{Q}_{ul,q}\}$ $\forall q$
  - However, there may be multiple power control policies that meet the long-term power constraint. The ergodic capacity region is then given by the union of the capacity regions, each region corresponding to a given power control policy.

Fast Fading - Perfect CSIT

- SISO MAC with long-term power constraint
  - Sum-rate maximization strategy: allow a single user to transmit at a time! That user is the one with the largest weighted channel gain

$$q^\star = \arg\max_q \frac{\Lambda_q^{-1} \left|h_{ul,q}\right|^2}{\nu_q}$$

    where ν_q is a Lagrangian multiplier chosen to satisfy the power constraint. The other users remain quiet until their own weighted channel gain becomes the largest.
  - Reminiscent of the water-filling power allocation. A user is allocated more power when its channel is good and less power when its channel is bad.
  - Dynamic TDMA based on channel measurement and dynamic user selection and power control is optimal to maximize the sum-rate! $\rightarrow$ multi-user diversity!
- SIMO MAC with long-term power constraint
  - The power is allocated to more than one user at a time.
  - As n_r increases, irrespectively of the number of users K, the optimal power allocation relying on CSIT provides a marginal gain over the constant power allocation strategy that utilizes only the path loss and shadowing information (but no small scale fading information). Hence, perfect CSIT is of decreasing value as n_r increases.
  - The multi-user diversity gain indeed decreases as n_r increases due to channel hardening effect.

Fast Fading - Partial Transmit Channel Knowledge

- $\mathbf{H}_{ul,q}$ is not known to the transmitter $q \; \forall Q \rightarrow$ we cannot adapt $\mathbf{Q}_{ul,q}$ at all time instants
- Rate of information flow between a subset of users S and Rx at time instant k over channels $\mathbf{H}_{ul,k,q} \; \forall q \in S$

$$\log_2 \det \left[\mathbf{I}_{n_r} + \frac{1}{\sigma_n^2} \sum_{q \in S} \Lambda_q^{-1} \mathbf{H}_{ul,k,q} \mathbf{Q}_{ul,q} \mathbf{H}_{ul,k,q}^H \right].$$

Such a rate varies over time according to the channel fluctuations. The average rate of information flow over a time duration $T >> T_{coh}$ is

$$\frac{1}{T} \sum_{k=0}^{T-1} \log_2 \det \left[\mathbf{I}_{n_r} + \frac{1}{\sigma_n^2} \sum_{q \in S} \Lambda_q^{-1} \mathbf{H}_{ul,k,q} \mathbf{Q}_{ul,q} \mathbf{H}_{ul,q}^H \right].$$

- The rate region is a K-dimensional polyhedron in general and a pentagon in the two-user case. The corner points are still achieved by MMSE-SIC.

Fast Fading - Partial Transmit Channel Knowledge

- The ergodic capacity region is obtained as the union of all the K-dimensional polyhedrons whose corresponding input covariance matrices satisfy the power constraints.

Proposition

The ergodic capacity region $\bar{C}_{MAC}$ of the Gaussian fast fading MIMO MAC is the set of all achievable rate vectors $(R_1, \ldots, R_K)$ given by

$$\bigcup_{\substack{\mathrm{Tr}\left\{\mathbf{Q}_{ul,q}\right\} \leq E_{s,q} \\ \mathbf{Q}_{ul,q} \geq 0, \forall q}} \left\{ \begin{array}{l} (R_1, \ldots, R_K) : \sum_{q \in S} R_q \leq \\ \mathcal{E}\left\{\log_2 \det\left[\mathbf{I}_{n_r} + \sum_{q \in S} \frac{\Lambda_q^{-1}}{\sigma_n^2} \mathbf{H}_{ul,q} \mathbf{Q}_{ul,q} \mathbf{H}_{ul,q}^H\right]\right\}, \forall S \subseteq \mathcal{K} \end{array} \right\}.$$

- $T >> T_c$ to average out the noise and the channel fluctuations.
- Assuming i.i.d. Rayleigh fading for all the users, equal power allocation, i.e. $\mathbf{Q}_{ul,q} = \frac{E_{s,q}}{n_{t,q}} \mathbf{I}_{n_{t,q}}$, is optimal to achieve the entire ergodic capacity region of the MIMO MAC and the sum-rate capacity scales linearly with $\min(n_r, \sum_{q=1}^{K} n_{t,q})$.
- TDMA incurs a loss compared to MMSE SIC for SISO, SIMO and MIMO MAC.

Outage Capacity and Probability in Slow Fading Channels

- The transmitters have only partial transmit channel knowledge in the form of the channel distribution information.
- MAC outage event $\mathcal{O} = \bigcup_S \mathcal{O}_S$ where

$$\mathcal{O}_S = \left\{ \mathbf{H}_{ul} : \log_2 \det \left[\mathbf{I}_{n_r} + \frac{1}{\sigma_n^2} \mathbf{H}_{ul,S} \mathbf{Q}_{ul,S} \mathbf{H}_{ul,S}^H \right] < \sum_{q \in S} R_q \right\}.$$

- Outage probability of the MAC is defined as the probability that the target rate vector $(R_1, \ldots, R_K)$ lies outside the achievable rate region.

Definition

The outage probability $P_{out}(R_1, \ldots, R_K)$ of a MIMO MAC with target rate vector $(R_1, \ldots, R_K)$ is given by

$$P_{out}(R_1, \ldots, R_K) = \min_{\left\{\mathbf{Q}_{ul,q} \geq 0, \mathrm{Tr}\left(\mathbf{Q}_{ul,q}\right) \leq E_{s,q}\right\}_{\forall q}} P\left(\bigcup_S \mathcal{O}_S\right).$$

Diversity-Multiplexing trade-off of i.i.d. Rayleigh Slow Fading Channels

- Assume that all users have the same transmit power constraint $E_{s,q} = E_s \; \forall q$ and experience independent and identically distributed channels with $\Lambda_q = \Lambda$ (so that $\eta_q = \eta \; \forall q$) and $\mathbf{H}_q$ being i.i.d. Rayleigh fading.
- Asymptotic (i.e. large η) diversity-multiplexing trade-off of the K-user MIMO MAC

Definition

A diversity gain $g^{\star}_{d,MAC}(g_{s,1}, \ldots, g_{s,K}, \infty)$ is achieved for the set of K-tuple multiplexing gains $(g_{s,1}, \ldots, g_{s,K})$ if

$$\lim_{\eta \longrightarrow \infty} \frac{R_q(\eta)}{\log_2(\eta)} = g_{s,q}, \qquad \forall q$$

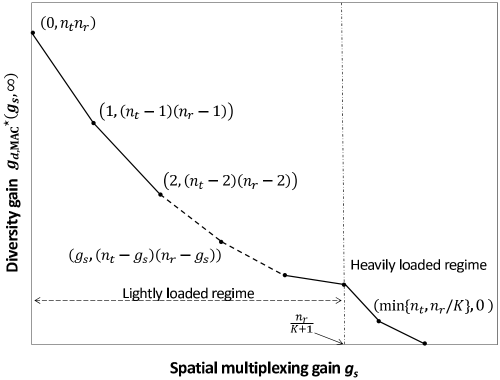
$$\lim_{\eta \longrightarrow \infty} \frac{\log_2\left(P_{out}(R_1, \ldots, R_K)\right)}{\log_2(\eta)} = -g^{\star}_{d,MAC}(g_{s,1}, \ldots, g_{s,K}, \infty)$$

The curve $g^{\star}_{d,MAC}(g_{s,1}, \ldots, g_{s,K}, \infty)$ as function of $(g_{s,1}, \ldots, g_{s,K})$ is known as the asymptotic diversity-multiplexing trade-off of the MIMO MAC.

- The DMT in the MAC differs from that of the single-link MIMO channel by the fact that coding can only be performed across antennas belonging to the same user and not jointly across all $\sum n_{t,q}$ antennas.

Diversity-Multiplexing trade-off of i.i.d. Rayleigh Slow Fading Channels

- Asymptotic symmetric ($n_{t,q} = n_t$ and $g_{s,q} = g_s\ \forall q$) DMT $g^{\star}_{d,MAC}(g_s, \infty)$ of MIMO MAC for $n_t > \frac{n_r}{K+1}$



  - *Lightly loaded*: multiple access is provided without compromising individual users' performance and admitting more users in the system does not degrade the users' performance.
  - *Heavily loaded*: tradeoff of a giant MIMO system made of Kn_t transmit antennas transmitting at a multiplexing rate Kg_s. User performance is affected by the presence of other users.

- Asymptotic symmetric DMT $g^{\star}_{d,MAC}(g_s, \infty)$ of MIMO MAC for $n_t \leq \frac{n_r}{K+1}$ is the same as single-link MIMO.

MIMO BC System Model

- Downlink multi-user MIMO (MU-MIMO) transmission
  - total number of K users ($\mathcal{K} = \{1, \dots, K\}$) distributed in a cell,
  - $n_{r,q}$ receive antennas at mobile terminal q (we simply drop the index q and write n_r if $n_{r,q} = n_r \ \forall q$)
  - n_t transmit antenna at the base station
- Received signal (we drop the time dimension)

$$\mathbf{y}_q = \Lambda_q^{-1/2} \mathbf{H}_q \mathbf{c}' + \mathbf{n}_q$$

where

  - $\mathbf{y}_q \in \quad ^{n_{r,q}}$
  - $\mathbf{H}_q \in \quad ^{n_{r,q} \times n_t}$ models the small scale time-varying fading process and Λ_q^{-1} refers to the large-scale fading accounting for path loss and shadowing
  - $\mathbf{n}_q$ is a complex Gaussian noise $\mathcal{CN}\left(0, \sigma_{n,q}^2 \mathbf{I}_{n_{r,q}}\right)$.
- The input covariance matrix is defined as the covariance matrix of the transmit signal as $\mathbf{Q} = \mathcal{E}\left\{\mathbf{c}'\mathbf{c}'^H\right\}$.
- Power constraint: Tr$\{\mathbf{Q}\} \leq E_s$.

MIMO BC System Model

- By stacking up the received signal vectors, the noise vectors and the channel matrices of all K users,

$$\mathbf{y} = \left[\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T\right]^T,$$

$$\mathbf{n} = \left[\mathbf{n}_1^T, \ldots, \mathbf{n}_K^T\right]^T,$$

$$\mathbf{H} = \left[\Lambda_1^{-1/2}\mathbf{H}_1^T, \ldots, \Lambda_K^{-1/2}\mathbf{H}_K^T\right]^T,$$

  the system model also writes as

$$\mathbf{y} = \mathbf{H}\mathbf{c}' + \mathbf{n}.$$

  $\mathbf{H}$ is assumed to be full-rank as it would be the case in a typical user deployment.
- SNR of user q defined as $\eta_q = E_s \Lambda_q^{-1}/\sigma_{n,q}^2$.
- Perfect instantaneous channel state information (CSI) at the Tx and all Rx.
- Generally speaking, $\mathbf{c}'$ is written as the superposition of statistically independent signals $\mathbf{c}'_q$

$$\mathbf{c}' = \sum_{q=1}^{K} \mathbf{c}'_q.$$

  The input covariance matrix of user q is defined as $\mathbf{Q}_q = \mathcal{E}\left\{\mathbf{c}'_q \mathbf{c}'^H_q\right\}$.

Capacity Region of two-user SISO Deterministic BC

- In *two-user* SISO MAC, point A was obtained by canceling user 2's signal first such that user 1 is left with Gaussian noise.
- Let us apply the same philosophy to the SISO BC:
  - transmit $c' = c_1' + c_2'$, with power of c_q' denoted as s_q
  - user 1 cancels user 2's signal c_2' so as to be left with its own Gaussian noise
  - user 2 decodes its signal by treating user 1's signal c_1' as Gaussian noise.
- Achievable rates of such strategy (with sum-power constraint $s_1 + s_2 = E_s$)

$$R_1 = \log_2 \left(1 + \frac{\Lambda_1^{-1} s_1}{\sigma_{n,1}^2} |h_1|^2\right)$$

$$R_2 = \log_2 \left(1 + \frac{\Lambda_2^{-1} |h_2|^2 s_2}{\sigma_{n,2}^2 + \Lambda_2^{-1} |h_2|^2 s_1}\right).$$

- *Careful!* For user 1 to be able to correctly cancel user 2's signal, user 1's channel has to be good enough to support R_2, i.e.

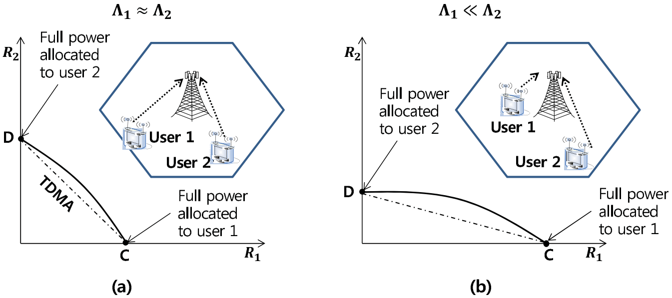
$$R_2 \leq \log_2 \left(1 + \frac{\Lambda_1^{-1} |h_1|^2 s_2}{\sigma_{n,1}^2 + \Lambda_1^{-1} |h_1|^2 s_1}\right).$$

- The channel gains normalized w.r.t. their respective noise power should be ordered

$$\frac{\Lambda_2^{-1} |h_2|^2}{\sigma_{n,2}^2} \leq \frac{\Lambda_1^{-1} |h_1|^2}{\sigma_{n,1}^2}.$$

Capacity Region of two-user SISO Deterministic BC

- If the ordering condition is satisfied, the above strategy achieves the boundary of the capacity region of the two-user SISO BC for any power allocation s_1 and s_2 satisfying $s_1 + s_2 = E_s$.
- The capacity region is given by the union of all rate pairs (R_1, R_2) over all power allocations s_1 and s_2 satisfying $s_1 + s_2 = E_s$.



Capacity Region of K-user SISO Deterministic BC

- Define $\mathsf{h}_q = \Lambda_q^{-1/2} h_q / \sigma_{n,q}$. Assume $|\mathsf{h}_1|^2 \geq |\mathsf{h}_2|^2 \geq \ldots \geq |\mathsf{h}_K|^2$.

Proposition

With the ordering $|\mathsf{h}_1|^2 \geq |\mathsf{h}_2|^2 \geq \ldots \geq |\mathsf{h}_K|^2$, the capacity region $\mathcal{C}_{BC}$ of the Gaussian SISO BC is the set of all achievable rate vectors $(R_1, \ldots, R_K)$ given by

$$\bigcup_{s_q : \sum_{q=1}^{K} s_q = E_s} \left\{ (R_1, \ldots, R_K) : R_q \leq \log_2 \left(1 + \frac{|\mathsf{h}_q|^2 s_q}{1 + |\mathsf{h}_q|^2 \left[\sum_{p=1}^{q-1} s_p \right]} \right), \forall q \right\}.$$

Proposition

The sum-rate capacity of the SISO BC is achieved by allocating the transmit power to the strongest user

$$C_{BC} = \log_2 \left(1 + E_s \max_{q=1,\ldots,K} |\mathsf{h}_q|^2 \right) = \log_2 \left(1 + \max_{q=1,\ldots,K} \eta_q |h_q|^2 \right).$$

Recall that the MAC sum-rate capacity is obtained with all users simultaneously transmitting at their respective full power.

Achievability of the SISO BC Capacity Region

- Receiver cancellation - *Superposition coding with SIC and appropriate ordering*:
  - User ordering: decode and cancel out weaker users signals before decoding their own signal.
  - The weakest user decodes only the coarsest constellation. The strongest user decodes and subtracts all constellation points in order to decode the finest constellation.
- Transmitter cancellation - *Dirty-Paper Coding* (DPC)
  - Assume a system model $y = hc' + i + n$ with i, n Gaussian interference and noise. Simply subtracting i for transmit signal is not a good idea!

Proposition

If Tx has full (non-causal) knowledge of the interference, the capacity of the dirty paper channel is equal to the capacity of the channel with the interference completely absent.

- By encoding users in the increasing order of their normalized channel gains, DPC achieves the capacity region of the SISO BC.

Example

Assume $|\mathrm{h}_1|^2 \geq |\mathrm{h}_2|^2$. By treating user 2's signal c_2' as known Gaussian interference at Tx and encoding user 1's signal c_1' using DPC, user 1 can achieve a rate as high as if user 2's signal was absent. User 2 treats user 1's signal as Gaussian noise.

Achievability of the SISO BC Capacity Region

Proposition

With the appropriate cancellation/encoding ordering, Superposition Coding with SIC and Dirty-Paper Coding are both optimal for achieving the SISO BC capacity region.

Proposition

The SISO BC sum-rate capacity is achievable with dynamic TDMA (to the strongest user), Superposition Coding with SIC (with the appropriate cancellation ordering) and Dirty-Paper Coding (with the appropriate encoding ordering).

Capacity Region of MIMO BC and its Achievability

- MAC with multiple Rx antennas provides a tremendous capacity increase compared to suboptimal TDMA. So does BC with multiple Tx antennas!
- MIMO BC difficult problem: users' channels cannot be ranked anymore.
- Assume an increasing encoding order from user 1 to K:
  1. Encode user 1's signal into $\mathbf{c}'_1$.
  2. With full knowledge of $\mathbf{c}'_1$, encode user 2's signal into $\mathbf{c}'_2$ using DPC: $\mathbf{c}'_1$ appears invisible to user 2 but $\mathbf{c}'_2$ appears like a Gaussian interference to user 1.
  3. With full knowledge of user 1 and user 2's signals, encode user 3's signal into $\mathbf{c}'_3$ using DPC.
  4. ... till K users are encoded.
- A given user q sees signals from users $p > q$ as a Gaussian interference but does not see any interference signals from users $p < q$:
  - Covariance of Noise plus Interference at user q: $\sigma_{n,q}^2 \mathbf{I}_{n_{r,q}} + \Lambda_q^{-1} \mathbf{H}_q \left[\sum_{p>q} \mathbf{Q}_p\right] \mathbf{H}_q^H$.
  - With a MMSE receiver that whitens the colored Gaussian interference (same as in MAC)

$$R_q = \log_2 \det \left[\mathbf{I}_{n_{r,q}} + \Lambda_q^{-1} \mathbf{H}_q \mathbf{Q}_q \mathbf{H}_q^H \left(\sigma_{n,q}^2 \mathbf{I}_{n_{r,q}} + \Lambda_q^{-1} \mathbf{H}_q \left[\sum_{p>q} \mathbf{Q}_p \right] \mathbf{H}_q^H \right)^{-1} \right]$$

- Capacity region: Repeat for all covariance matrices $\mathbf{Q}_1, \ldots, \mathbf{Q}_K$ satisfying the sum-power constraint $\sum_q \mathrm{Tr}\{\mathbf{Q}_q\} = E_s$ and all user ordering.
- Only DPC can achieve the MISO/MIMO BC sum-rate capacity.

BC-MAC Duality

- There are similarities between BC and MAC:
  - Both MAC and BC deal with received signal(s) expressed as a sum of K (Gaussian) codewords scaled by the wireless channel and perform SIC at the receiver(s).
  - The receiver in SISO MAC receives the sum of K (Gaussian) codewords (after propagating through the wireless channels) and decodes each of those signals using SIC.
  - In the degraded SISO BC, the transmitter sends a sum of K (Gaussian) codewords using superposition coding and each receiver also decodes its own codeword using SIC.
- Can those similarities be formally characterized? Yes, by the MAC-BC or UL-DL duality.
- Interestingly, the BC capacity region can be characterized in terms of the capacity region of a dual MAC and vice-versa.
  - By dual MAC, we here refer to the channel obtained by converting the transmitter in the BC into a receiver and by converting the receivers in the BC into the transmitters.
  - The BC and dual-MAC have the same channel gains and the noise variances at their respective receivers are equal.
  - The power constraint for the BC equals the sum of the individual power constraints of the dual MAC.

SISO BC-MAC Duality

- Assuming a SISO BC over a deterministic channel $\mathbf{h} = \left[\Lambda_1^{-1/2} h_1, \ldots, \Lambda_K^{-1/2} h_K\right]^T$ with receiver noise powers $\sigma_{n,1}^2, \ldots, \sigma_{n,K}^2$, we express the SISO BC in the equivalent system model with unit variance receiver noises and normalized channel gains $\mathsf{h}_q = \Lambda_q^{-1/2} h_q / \sigma_{n,q}$ such that $\mathsf{\mathbf{h}} = [\mathsf{h}_1, \ldots, \mathsf{h}_K]^T$.
- The system model $\mathbf{y} = \mathbf{h} c' + \mathbf{n}$ for SISO then writes equivalently as

$$\mathbf{y}_{dl} = \mathsf{\mathbf{h}} c'_{dl} + \tilde{\mathbf{n}}_{dl}$$

where $\tilde{\mathbf{n}}$ is a complex Gaussian noise $\mathcal{CN}\left(0, \mathbf{I}_K\right)$.

- We can also define the MAC where user q's uplink channel $h_{ul,q}$ is given by h_q and the receiver noise power is equal to one as

$$y_{ul} = \mathsf{\mathbf{h}}^T \mathbf{c}'_{ul} + \tilde{n}_{ul}$$

where $\tilde{n}$ is a complex Gaussian noise $\mathcal{CN}\left(0, 1\right)$.

- The MAC is the dual of the BC and vice-versa.

SISO BC-MAC Duality

- Usefulness: Given the difficulty to characterize the BC capacity region, the MAC-to-BC duality is very helpful to express the BC capacity region as a function of the capacity region of its dual MAC.

Proposition

The capacity region of a Gaussian SISO BC with power constraint E_s over a deterministic channel $\mathbf{h} = [\mathsf{h}_1, \ldots, \mathsf{h}_K]^T$ with unit variance receiver noise, denoted explicitly as $\mathcal{C}_{BC}(E_s, \mathbf{h})$, is equal to the union of the capacity regions of the dual MAC with individual power constraints $E_{s,q}$ $(q = 1, \ldots, K)$ such that $\sum_{q=1}^{K} E_{s,q} = E_s$

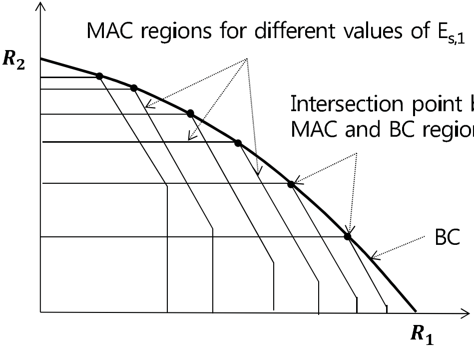
$$\mathcal{C}_{BC}(E_s, \mathbf{h}) = \bigcup_{\{E_{s,q}\}_{\forall q} : \sum_{q=1}^{K} E_{s,q} = E_s} \mathcal{C}_{MAC}(E_{s,1}, \ldots, E_{s,K}, \mathbf{h})$$

$$= \bigcup_{\{E_{s,q}\}_{\forall q} : \sum_{q=1}^{K} E_{s,q} = E_s} \left\{ \begin{array}{l} (R_1, \ldots, R_K) : \sum_{q \in S} R_q \leq \\ \log_2 \left(1 + \sum_{q \in S} |\mathsf{h}_q|^2 E_{s,q}\right), \forall S \subseteq \mathcal{K} \end{array} \right\}$$

where $\mathcal{C}_{MAC}(E_{s,1}, \ldots, E_{s,K}, \mathbf{h})$ is capacity region of the SISO MAC with the channel gains $\Lambda_q^{-1/2} h_{ul,q}$ replaced by the normalized channel gain h_q and the noise power $\sigma_n^2 = 1$.

SISO BC-MAC Duality

- Two-user SISO BC capacity region characterized in the terms of the capacity region of its dual MAC



MIMO BC-MAC Duality

- The system model $\mathbf{y} = \mathbf{H}\mathbf{c}' + \mathbf{n}$ can be written equivalently as

$$\mathbf{y}_{dl} = \mathbf{H}\mathbf{c}'_{dl} + \tilde{\mathbf{n}}_{dl}$$

where $\mathbf{H} = \left[\mathbf{H}_1^T, \ldots, \mathbf{H}_K^T\right]^T$ with $\mathbf{H}_q = \frac{\Lambda_q^{-1/2}\mathbf{H}_q}{\sigma_{n,q}}$ and $\tilde{\mathbf{n}}_{dl}$ is a complex Gaussian noise $\mathcal{CN}\left(0, \mathbf{I}_{\sum_q n_{r,q}}\right)$.

- The dual uplink channel has K users and n_t receive antennas

$$\mathbf{y}_{ul} = \mathbf{H}^H\mathbf{c}'_{ul} + \tilde{\mathbf{n}}_{ul}$$

where $\mathbf{c}'_{ul}$ is the vector of transmitted signals from the K users, $\mathbf{y}_{ul}$ is the received signal vector at the n_t receive antennas and $\tilde{\mathbf{n}}_{ul}$ is a complex Gaussian noise $\mathcal{CN}\left(0, \mathbf{I}_{n_t}\right)$.

MIMO BC-MAC Duality

Proposition

The capacity region of the MIMO BC with power constraint E_s over a deterministic channel $\mathbf{H}$ *is equal to the union of the capacity region of the dual MIMO MAC with individual power constraints $E_{s,q}$ such that $\sum_{q=1}^{K} E_{s,q} = E_s$*

$$\mathcal{C}_{BC}\left(E_s, \mathbf{H}\right) = \bigcup_{\left\{E_{s,q}\right\}_{\forall q} : \sum_{q=1}^{K} E_{s,q} = E_s} \mathcal{C}_{MAC}\left(E_{s,1}, \ldots, E_{s,K}, \mathbf{H}^H\right)$$

$$= \bigcup_{\substack{\left\{\mathbf{Q}_{ul,q} \geq 0\right\}_{\forall q}, \\ \sum_{q=1}^{K} \mathrm{Tr}\left\{\mathbf{Q}_{ul,q}\right\} \leq E_s}} \left\{ \begin{array}{l} (R_1, \ldots, R_K) : \sum_{q \in S} R_q \leq \\ \log_2 \det \left[\mathbf{I}_{n_t} + \sum_{q \in S} \mathbf{H}_q^H \mathbf{Q}_{ul,q} \mathbf{H}_q\right], \forall S \subseteq \mathcal{K} \end{array} \right\}$$

where $\mathcal{C}_{MAC}\left(E_{s,1}, \ldots, E_{s,K}, \mathbf{H}^H\right)$ is teh MIMO MAC capacity region with the channel matrix $\mathbf{H}_{ul}$ replaced by $\mathbf{H}^H$ and the noise power $\sigma_n^2 = 1$.

MIMO BC-MAC Duality

Proposition

The sum-rate capacity of the MIMO BC is equal to the sum-rate capacity of the sum power dual MIMO MAC

$$C_{BC}(\mathbf{H}, E_s) = C_{MAC}\left(\mathbf{H}^H, E_s\right)$$
$$= \max_{\substack{\left\{\mathbf{Q}_{ul,q} \geq 0\right\}_{\forall q}, \\ \sum_{q=1}^{K} \mathrm{Tr}\left\{\mathbf{Q}_{ul,q}\right\} \leq E_s}} \log_2 \det\left[\mathbf{I}_{n_t} + \sum_{q=1}^{K} \mathbf{H}_q^H \mathbf{Q}_{ul,q} \mathbf{H}_q\right].$$

Bounds on Sum-Rate Capacity of MIMO BC

- Define
  - $\tilde{n} = \min\{n_t, Kn_r\}$ ($n_{r,q} = n_r \ \forall q$ is assumed),
  - $\lambda_{max,q}$ as the dominant eigenvalue of $\mathbf{H}_q^H \mathbf{H}_q$,
- Achievable multiplexing gain of $\tilde{n}$

Proposition

The sum-rate capacity of MIMO BC for a deterministic channel $\mathbf{H}$, achievable with DPC, $C_{BC}(\mathbf{H})$, is lower-bounded as

$$C_{BC}(\mathbf{H}) \geq C_{BF}(\mathbf{H}) = \sum_{q=1}^{\tilde{n}} \log_2\left(1 + \alpha_q^2 \frac{\eta_q}{\tilde{n}}\right)$$

for some non-zero channel gains $\alpha_1^2, \ldots, \alpha_{\tilde{n}}^2$ and is upper-bounded as

$$C_{BC}(\mathbf{H}) \leq n_t \log_2\left(1 + \frac{1}{n_t} \max_{q=1,\ldots,K} \eta_q \lambda_{max,q}\right),$$

$$C_{BC}(\mathbf{H}) \leq C_{CSIT}(\mathbf{H}),$$

$$C_{BC}(\mathbf{H}) \leq \sum_{q=1}^{K} C_{CSIT}(\mathbf{H}_q).$$

Comparisons with TDMA

- SISO
  - Similarly to MAC, TDMA rate region is contained in the BC capacity region.
  - The gap between the BC capacity region and the TDMA rate region increases proportionally with the asymmetry between users normalized channel gains.
  - TDMA achieves the sum-rate capacity of SISO BC.
- MIMO
  - The maximum sum-rate $C_{TDMA}(\mathbf{H})$ is the largest single-link capacity among K users
$$C_{TDMA}(\mathbf{H}) = \max_{q=1,\ldots,K} C_{CSIT}(\mathbf{H}_q).$$
  - Define $\lambda_{max,q}$ and λ_{max} as the largest eigenvalue of $\mathbf{H}_q^H\mathbf{H}_q$ and $\mathbf{H}^H\mathbf{H}$, $\lambda_k\left(\mathbf{H}_q\mathbf{H}_q^H\right)$ as the non-zero eigenvalues of $\mathbf{H}_q\mathbf{H}_q^H$ and $n = \min\{n_t, n_r\}$. Assume $n_{r,q} = n_r\ \forall q$.

Proposition

The maximum TDMA sum-rate, $C_{TDMA}(\mathbf{H})$, is lower bounded as

$$C_{TDMA}(\mathbf{H}) \geq \log_2\left(1 + \max_{q=1,\ldots,K} \eta_q\lambda_{max,q}\right),$$

$$C_{TDMA}(\mathbf{H}) \geq C_{CSIT}(\mathbf{H}_q) \geq \sum_{k=1}^{n} \log_2\left(1 + \frac{\eta_q}{n}\lambda_k\left(\mathbf{H}_q\mathbf{H}_q^H\right)\right),$$

for $q = 1, \ldots, K$, and is upper bounded as

$$C_{TDMA}(\mathbf{H}) \leq n\log_2\left(1 + \frac{1}{n}\max_{q=1,\ldots,K} \eta_q\lambda_{max,q}\right).$$

Comparisons with TDMA

Proposition

For channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$, SNR η_q, number of receive antennas n_r, the gain of DPC over TDMA is upper-bounded by the minimum between the number of transmit antennas n_t and the number of users K

$$\frac{C_{BC}(\mathbf{H})}{C_{TDMA}(\mathbf{H})} \leq \min\{n_t, K\}.$$

Intuition:

- TDMA exploits at least one spatial dimension with the largest effective SNR among all users.
- DPC exploits up to n_t dimensions. Since the quality of each of those n_t dimensions cannot be larger than the single dimension used in the TDMA lower bound, DPC cannot achieve a rate larger than n_t times the TDMA capacity.

Proposition

For any n_t, n_r and K, at high SNR ($E_s \to \infty$, i.e. $\eta_q \to \infty \; \forall q$),

$$\frac{C_{BC}(\mathbf{H})}{C_{TDMA}(\mathbf{H})} \approx \frac{\min\{n_t, Kn_r\}}{\min\{n_t, n_r\}}.$$

Ergodic Capacity Region of SISO Fast Fading Channels

Perfect CSIT

- short-term power constraint $\sum_q s_q = E_s$: similar to deterministic channels
- long-term power constraint $\mathcal{E}_{\mathbf{H}}\left\{\sum_q s_q\left(\mathbf{H}\right)\right\} = E_s$:
  - Power control policy that maps a channel realization to a set of transmit power.
  - Ergodic capacity region is given by the union of all achievable rate regions over all power control policies that satisfy the long-term power constraint.
  - Sum-rate maximization: the sum-rate capacity can also be achieved by transmitting to the strongest user in each fading state

$$q^\star = \arg \max_{q=1,\ldots,K} \left|\mathsf{h}_q\right|^2 = \arg \max_{q=1,\ldots,K} \frac{\Lambda_q^{-1}\left|h_q\right|^2}{\sigma_{n,q}^2}$$

    and the power in each fading state can be optimized following the time domain water-filling solution.
  - Observe the similarity with SISO MAC. The user to be selected is slightly different.
- In MISO and MIMO BC, not sufficient to transmit to a single user at a time to achieve the sum-rate capacity (similarly to SIMO/MIMO MAC).

Partial CSIT

- No channel ordering and no way know the interference to other users.
- TDMA is an appropriate strategy. Huge loss compared to perfect CSIT!

Outage Capacity and Probability in Slow Fading Channels

- The notion of diversity-multiplexing trade-off is more meaningful in the absence of CSIT when the transmitter cannot adapt its transmit strategy as a function of the channel realization.
- However, SISO/MIMO BC both critically depend on CSIT. With only partial channel knowledge at the transmitter, the performance drops significantly.
- Assume perfect CSIT

Proposition

For a MISO BC with n_t transmit antennas and $K \leq n_t$ single-antenna users (whose concatenated channel matrix entries are Rayleigh i.i.d.) and given the fixed rates $R_1, \ldots, R_K$,

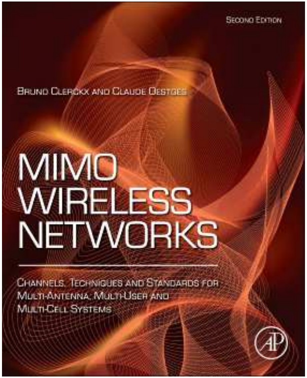
$$g_{d,BC}^{\star}(0, \ldots, 0, \infty) \leq n_t.$$

The diversity gain of MU-MIMO precoding in MISO BC is not larger than the diversity gain of transmit beamforming in a single link.

Multi-User MIMO - Scheduling and Precoding (Downlink)

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Chapter 12
    - Section: 12.1,12.5,12.6,12.8

Introduction

- BC: $K >> n_t$, MAC: $K >> n_r$ $\rightarrow$ All users cannot be scheduled at the same time.
  - Which users to schedule?
  - How to account for fairness?

- DPC is optimal in MIMO BC but is very complex to implement.
  - Can we derive suboptimal strategies? Yes, there are various linear and non-linear precoding techniques
  - How to design suboptimal linear precoders?
  - What is the performance of those precoders combined with scheduling?

- What if we do not have perfect channel knowledge at the transmitter to design the precoders in MIMO BC?

System Model

- Downlink multi-user MIMO (MU-MIMO) transmission
  - total number of K users ($\mathcal{K} = \{1, \ldots, K\}$) distributed in a cell,
  - $n_{r,q}$ receive antennas at mobile terminal q (we simply drop the index q and write n_r if $n_{r,q} = n_r \; \forall q$)
  - n_t transmit antenna at the base station
- Received signal (we drop the time dimension)

$$\mathbf{y}_q = \Lambda_q^{-1/2} \mathbf{H}_q \mathbf{c}' + \mathbf{n}_q$$

where

  - $\mathbf{y}_q \in {}^{n_{r,q}}$
  - $\mathbf{H}_q \in {}^{n_{r,q} \times n_t}$ models the small scale time-varying fading process and Λ_q^{-1} refers to the large-scale fading accounting for path loss and shadowing
  - $\mathbf{n}_q$ is a complex Gaussian noise $\mathcal{CN}\big(0, \sigma_{n,q}^2 \mathbf{I}_{n_{r,q}}\big)$.
- Long term SNR of user q defined as $\eta_q = E_s \Lambda_q^{-1} / \sigma_{n,q}^2$.
- Generally speaking, $\mathbf{c}'$ is written as the superposition of statistically independent signals $\mathbf{c}'_q$

$$\mathbf{c}' = \sum_{q=1}^{K} \mathbf{c}'_q.$$

- Power constraint: Tr$\{\mathbf{Q}\} \leq E_s$ with $\mathbf{Q} = \mathcal{E}\big\{\mathbf{c}'\mathbf{c}'^H\big\}$.

System Model - Linear Precoding

- *scheduled user set*, denoted as $\mathbf{K} \subset \mathcal{K}$, is the set of users who are actually scheduled (with a non-zero transmit power) by the transmitter at the time instant of interest.
- The transmitter serves users belonging to $\mathbf{K}$ with n_e data streams and user $q \in \mathbf{K}$ is served with $n_{u,q}$ data streams ($n_{u,q} \leq n_e$). Hence, $n_e = \sum_{q \in \mathbf{K}} n_{u,q}$.
- Linear Precoding

$$\mathbf{c}' = \mathbf{P}\mathbf{c} = \mathbf{W}\mathbf{S}^{1/2}\mathbf{c} = \sum_{q \in \mathbf{K}} \mathbf{P}_q \mathbf{c}_q = \sum_{q \in \mathbf{K}} \mathbf{W}_q \mathbf{S}_q^{1/2} \mathbf{c}_q$$

  where
  - $\mathbf{c}$ is the symbol vector made of n_e unit-energy independent symbols.
  - $\mathbf{P} \in {}^{n_t \times n_e}$ is the precoder subject to $\text{Tr}\{\mathbf{P}\mathbf{P}^H\} \leq E_s$, made of two matrices: a power control diagonal matrix denoted as $\mathbf{S} \in {}^{n_e \times n_e}$ and a transmit beamforming matrix $\mathbf{W} \in {}^{n_t \times n_e}$.
  - $\mathbf{P}_q \in {}^{n_t \times n_{u,q}}$, $\mathbf{W}_q \in {}^{n_t \times n_{u,q}}$, $\mathbf{S}_q \in {}^{n_{u,q} \times n_{u,q}}$, and $\mathbf{c}_q \in {}^{n_{u,q}}$ are user q's sub-matrices and sub-vector of $\mathbf{P}$, $\mathbf{W}$, $\mathbf{S}$, and $\mathbf{c}$, respectively.
- The received signal $\mathbf{y}_q \in {}^{n_{r,q}}$ is shaped by $\mathbf{G}_q \in {}^{n_{u,q} \times n_{r,q}}$ and the filtered received signal $\mathbf{z}_q \in {}^{n_{u,q}}$ at user q writes as

$$\begin{aligned} \mathbf{z}_q &= \mathbf{G}_q \mathbf{y}_q, \\ &= \Lambda_q^{-1/2} \mathbf{G}_q \mathbf{H}_q \mathbf{W}_q \mathbf{S}_q^{1/2} \mathbf{c}_q + \sum_{p \in \mathbf{K},\, p \neq q} \Lambda_q^{-1/2} \mathbf{G}_q \mathbf{H}_q \mathbf{W}_p \mathbf{S}_p^{1/2} \mathbf{c}_p + \mathbf{G}_q \mathbf{n}_q. \end{aligned}$$

Multi-User Diversity

- In single-link systems, channel fading is viewed as a source of unreliability mitigated through diversity techniques (e.g. space-time coding).
- In multi-user communications, fading is viewed as a source of randomization that can be exploited!
- Multi-User (MU) diversity is a form of selection diversity among users provided by independent time-varying channels across the different users.
- Provided that the BS is able to track the user channel fluctuations (based on feedback), it can schedule transmissions to the users with favorable channel fading conditions, i.e. near their peaks, to improve the total cell throughput.
- Recall that MU diversity was already identified as part of the sum-rate maximization in SISO BC.

Multi-User Diversity Gain in SISO

- Assume that the fading distribution of the K users are independent and identically ($\Lambda_q^{-1} = \Lambda^{-1}$ and channel gains h_q are drawn from the same) Rayleigh distributed and that users experience the same average SNR $\eta_q = \eta$ $(\sigma_{n,q}^2 = \sigma_n^2)$ $\forall q$:

$$\mathbf{y}_q = \Lambda^{-1/2} h_q c' + \mathbf{n}_q.$$

- Assume MU-SISO where one user is scheduled at a time in a TDMA manner: select the user with the largest channel gain.
- Mathematically same as antenna selection diversity.
- Average SNR gain
  - Average SNR after user selection $\bar{\rho}_{out}$

$$\bar{\rho}_{out} = \mathcal{E}\left\{\eta \max_{q=1,\ldots,K} |h_q|^2\right\} = \eta \sum_{q=1}^{K} \frac{1}{q}.$$

  - SNR gain provided by MU diversity g_m

$$g_m = \frac{\bar{\rho}_{out}}{\eta} = \sum_{q=1}^{K} \frac{1}{q} \overset{K\to\infty}{\cong} \log(K).$$

  g_m is of the order of $\log(K)$ and hence the gain of the strongest user grows as $\log(K)$!
- Heavily relies on CSIT (partial or imperfect feedback impacts the performance) and independent user fading distributions (correlated fading or LOS are not good for MU diversity)

Multi-User Diversity Gain in SISO

- Sum-rate capacity

$$\bar{C}_{TDMA} = \mathcal{E}\left\{C_{TDMA}\right\} = \mathcal{E}\left\{\log_2\left(1 + \eta \max_{q=1,\ldots,K} |h_q|^2\right)\right\}.$$

  - low SNR

$$\bar{C}_{TDMA} \approx \mathcal{E}\left\{\max_{q=1,\ldots,K} |h_q|^2\right\} \eta \log_2(e) \approx g_m C_{awgn}.$$

    Observations: capacity of the fading channel $\log(K)$ times larger than the AWGN capacity.
  - high SNR (Use Jensen's inequality: $\mathcal{E}_x\left\{\mathcal{F}(x)\right\} \leq \mathcal{F}\left(\mathcal{E}_x\left\{x\right\}\right)$ if $\mathcal{F}$ concave)

$$\begin{aligned}
\bar{C}_{TDMA} &\approx \log_2(\eta) + \mathcal{E}\left\{\log_2\left(\max_{q=1,\ldots,K} |h_q|^2\right)\right\}, \\
&\approx C_{awgn} + \mathcal{E}\left\{\log_2\left(\max_{q=1,\ldots,K} |h_q|^2\right)\right\}, \\
&\overset{(a)}{\leq} C_{awgn} + \log_2\left(\mathcal{E}\left\{\max_{q=1,\ldots,K} |h_q|^2\right\}\right), \\
&= C_{awgn} + \log_2(g_m).
\end{aligned}$$

    Observations: capacity of a fading channel is larger than the AWGN capacity by a factor roughly equal to $\log_2(g_m) \approx \log\log(K)$.
- Fading channels are significantly more useful in a multi-user setting than in a single-user setting

Multi-User Diversity

- In MU-MIMO, the performance is function of the channel magnitude but also of the spatial directions and properties of the channel matrices.
- MU diversity offers abundant spatial channel directions and allows to appropriately choose users with good channel matrix properties or spatial separations.
- Opportunistic Beamforming: precode multiple streams along the unitary precoding matrix $\mathbf{W}$ (orthogonal beams). For a large number of users, thanks to MU diversity, each beam matches one user channel with a high probability and orthogonality of beams prevents users from experiencing multi-user interference

$$y_q = \Lambda_q^{-1/2}\mathbf{h}_q\mathbf{W}\mathbf{S}^{1/2}\mathbf{c} + n_q \overset{K\to\infty}{=} \Lambda_q^{-1/2}\left\|\mathbf{h}_q\right\| s_q^{1/2} c_q + n_q.$$

  - The terminal only measures the effective channel, i.e. the channel precoded by each beam, and reports the SNR (or CQI) for one or multiple beam(s).
  - Works well only for very large K.

Multi-User Diversity

- Few fundamental differences with classical spatial/time/frequency diversity:
  - Diversity techniques, like space-time coding, mainly focus on improving reliability by decreasing the outage probability in slow fading channels. MU diversity on the other hand increases the data rate over time-varying channels.
  - Classical diversity techniques mitigate fading while MU diversity exploits fading.
  - MU diversity takes a system-level view while classical diversity approaches focus on a single-link. This system-level view becomes increasingly important as we shift from single-cell to multi-cell scenarios.

Resource Allocation, Fairness and Scheduling Criteria

- An appropriate scheduler should allocate resources (time, frequency, spatial, power) to the users in a fair manner while exploiting the MU diversity gain.
- Goal of the resource allocation strategy at the scheduler: maximize the utility metric $\mathcal{U}$.

$$\left\{\mathbf{c}'^{\star}, \mathbf{G}^{\star}, \mathbf{K}^{\star}\right\} = \arg \max_{\mathbf{c}', \mathbf{G}, \mathbf{K} \subset \mathcal{K}} \mathcal{U}$$

  where $\mathbf{c}'^{\star}$ is the optimum transmit vector, $\mathbf{G}^{\star}$ denotes the optimum set of receive beamformers, and $\mathbf{K}^{\star} \subset \mathcal{K}$ refers to the optimum subset of users to be scheduled.
- Two major kinds of resource allocation strategies:
  - *rate-maximization policy*: maximizes the sum-rate - no fairness among users
  - *fairness oriented policy*, commonly relying on a *proportional fair* (PF) metric: maximizes a weighted sum-rate and guarantees fairness among users.
- Those two strategies can be addressed by using two different utility metrics:

$$\left\{\mathbf{c}'^{\star}, \mathbf{G}^{\star}, \mathbf{K}^{\star}\right\} = \arg \max_{\mathbf{c}', \mathbf{G}, \mathbf{K} \subset \mathcal{K}} \sum_{q \in \mathbf{K}} w_q R_q$$

  where
  - rate-maximization approach: $w_q = 1$
  - proportional fair approach: $w_q = \frac{\gamma_q}{\bar{R}_q}$ ($\bar{R}_q$ is the long-term average rate of user q and γ_q is the Quality of Service (QoS) of each user).

Practical Proportional Fair Scheduling

- The long-term average rate $\bar{R}_q$ of user q is updated using an exponentially weighted low-pass filter such that the estimate of $\bar{R}_q$ at time $k+1$, denoted as $\bar{R}_q\left(k+1\right)$, is function of the long-term average rate $\bar{R}_q\left(k\right)$ and of the current rate $R_q(k)$ at current time instant k as outlined by

$$\bar{R}_q\left(k+1\right)=\begin{cases}\left(1-1/t_c\right)\bar{R}_q\left(k\right)+1/t_c R_q(k), & q\in\mathbf{K}^\star\\ \left(1-1/t_c\right)\bar{R}_q\left(k\right), & q\notin\mathbf{K}^\star\end{cases}$$

  where t_c is the scheduling time scale and $\mathbf{K}^\star$ refers to the scheduled user set at time k. The resources should thus be allocated at time instant k as

$$\left\{\mathbf{c}'^\star,\mathbf{G}^\star,\mathbf{K}^\star\right\}=\arg\max_{\mathbf{c}',\mathbf{G},\mathbf{K}\subset\mathcal{K}}\sum_{q\in\mathbf{K}}\gamma_q\frac{R_q(k)}{\bar{R}_q(k)}.$$

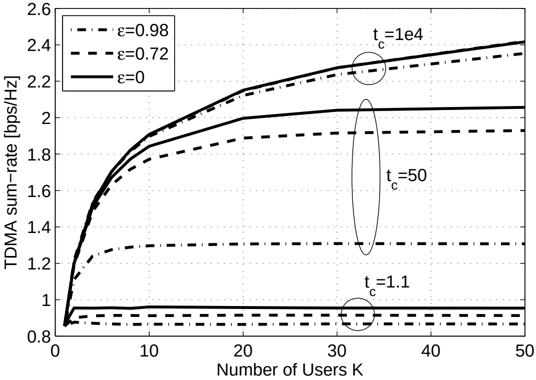
- The scheduling time scale t_c is a design parameter of the system that highly influences the user fairness and the performance
  - Very large t_c: assuming all users experience identical fading statistics and have the same QoS, the PF scheduler is equivalent to the rate-maximization scheduler, i.e. users contributing to the highest sum-rate are selected.
  - Small t_c: assuming all users have the same QoS, the scheduler divides the available resources equally among users (*Round-Robin* scheduling). No MU diversity is exploited.

Proportional Fair Scheduling

- Sum-rate of SISO TDMA with PF scheduling at SNR=0 dB as a function of the number of users K, the scheduling time scale t_c and the channel model

$$h_k = \epsilon h_{k-1} + \sqrt{1-\epsilon^2} n_k$$

with ϵ the channel time correlation coefficient.



User Grouping

- Given the presence of K users in the cell, the scheduler for MU-MIMO aims at finding the best scheduled user set among all possible candidates within $\mathcal{K}$.

- The *exhaustive search* is computationally intensive. Assuming a single stream transmission per user and $n_e \leq \min\{n_t, K\}$, a search like (with $R(\mathbf{K}) = \sum_{q \in \mathbf{K}} w_q R_q$)

$$\mathbf{K}^{\star} = \arg \max_{\substack{\mathbf{K} \subset \mathcal{K} \\ n_e \leq \min\{n_t, K\}}} R(\mathbf{K})$$

  requires to consider a large number of different sets and has a complexity that quickly becomes cumbersome as K increases.

User Grouping

- Lower complexity user grouping algorithms: Greedy User Selection, Semi-orthogonal User Selection.

- Greedy User Selection consists in successively adding a user to the tentative scheduled user set only if the weighted sum-rate is increased.
  - *Initialization step:* We fix $n = 1$, $\mathbf{K}^{(0)} = \emptyset$, $R\left(\mathbf{K}^{(0)}\right) = 0$ and *Done* $= 0$.
  - *Iteration-n:* While $(n \leq \min\{K, n_t\})$ and (not *Done*), select the user

$$q^{(n)} = \arg\max_{q \in \mathcal{K} \setminus \mathbf{K}} R\left(\mathbf{K}^{(n-1)} \cup q^{(n)}\right).$$

  If $R\left(\mathbf{K}^{(n-1)} \cup q^{(n)}\right) < R\left(\mathbf{K}^{(n-1)}\right)$, $\mathbf{K}^{(n)} = \mathbf{K}^{(n-1)}$ and *Done* $= 1$, otherwise $\mathbf{K}^{(n)} = \mathbf{K}^{(n-1)} \cup q^{(n)}$ and we move the next iteration $n + 1$. The final scheduled user set $\mathbf{K} = \mathbf{K}^{(n)}$.

User Correlation

- Scheduling orthogonal users provides a larger sum-rate by removing naturally the multi-user interference.

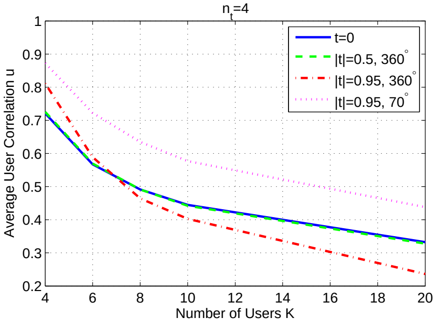
- The probability of finding orthogonal users is an important indicator of the performance of MU-MIMO.

- Among K active single-antenna users, the probability of finding n_t (semi-)orthogonal users is investigated.
  - Denote as $\mathcal{K}_{n_t}$ all sets of n_t users among the K active users, and as $\mathbf{K}$ one of those sets,
  - user correlation u

$$u = \mathcal{E}_{\{\mathbf{h}_{w,q}\}_{\forall q}} \left\{ \min_{\mathbf{K} \in \mathcal{K}_{n_t}} \max_{k,l \in \mathbf{K}} \left| \bar{\mathbf{h}}_k \bar{\mathbf{h}}_l^H \right| \right\},$$

User Correlation

- Average value of u as a function of $|t|$ and K, where the average value here refers to the averaging over different sets of transmit correlation matrices $\mathbf{R}_{t,q}$ ($|t_q| = |t|\ \forall q$, phases of the correlation coefficients t_q are randomly generated).



  - User correlation decreases with K at a higher rate in spatially correlated channels.
  - Transmit correlation decreases the probability of finding semi-orthogonal users for moderate K while it increases such probability for large K.
  - Spatially correlated fading at TX detrimental (resp. beneficial) to MU-MIMO for small (resp. large) K.
  - Cell sectorization increases user correlation.

Precoding with Perfect Transmit Channel Knowledge

- Single-link Spatial Multiplexing: Multiple Eigenmode Transmission relies on CSI knowledge at both the transmitter and the receiver and splits the spatial channel equalization between the transmitter and the receiver. As a result, the channel is decoupled into multiple parallel data pipes.
- Unfortunately, this approach cannot be applied to MU-MIMO as the receivers do not cooperate.
- MIMO BC, DPC optimal but extremely complex. Any suboptimal strategies?
- In MU-MIMO where CSI is available at the transmitter, precoding techniques reminiscent of the receiver architectures for SM

precoding	transmitter side	receiver side
Linear	Matched Beamforming (MBF)	MRC
Linear	**Zero-Forcing Beamforming (ZFBF)**	ZF
Linear	Regularized Zero-Forcing Beamforming (R-ZFBF)	MMSE
Non-linear	**Tomlinson-Harashima Precoding (THP)**	SIC
Non-linear	Vector Perturbation (VP)	sphere decoder

Achievable rate

- Maximum rate achievable by user q with linear precoding is

$$R_q = \sum_{l=1}^{n_{u,q}} \log_2 (1 + \rho_{q,l})$$

where $\rho_{q,l}$ denotes the SINR experienced by stream l of user q

$$\rho_{q,l} = \frac{\Lambda_q^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_q\mathbf{p}_{q,l}\right|^2}{I_l + I_c + \left\|\mathbf{g}_{q,l}\right\|^2 \sigma_{n,q}^2} = \frac{\Lambda_q^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_q\mathbf{w}_{q,l}\right|^2 s_{q,l}}{I_l + I_c + \left\|\mathbf{g}_{q,l}\right\|^2 \sigma_{n,q}^2}$$

with $\mathbf{p}_{q,l} = \mathbf{w}_{q,l} s_{q,l}$ (resp. $\mathbf{g}_{q,l}$) the precoder (resp. combiner) attached to stream l of user q, I_l the inter-stream interference and I_c the intra-cell interference (also called multi-user interference)

$$I_l = \sum_{m \neq l} \Lambda_q^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_q\mathbf{p}_{q,m}\right|^2 = \sum_{m \neq l} \Lambda_q^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_q\mathbf{w}_{q,m}\right|^2 s_{q,m},$$

$$I_c = \sum_{\substack{p \in \mathbf{K} \\ p \neq q}} \sum_{m=1}^{n_{u,p}} \Lambda_q^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_q\mathbf{p}_{p,m}\right|^2 = \sum_{\substack{p \in \mathbf{K} \\ p \neq q}} \sum_{m=1}^{n_{u,p}} \Lambda_q^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_q\mathbf{w}_{p,m}\right|^2 s_{p,m}.$$

- If $n_r = 1$, the SINR of user q simply reads as $\rho_q = \frac{\Lambda_q^{-1}\left|\mathbf{h}_q\mathbf{w}_q\right|^2 s_q}{\sum_{\substack{p \in \mathbf{K} \\ p \neq q}} \Lambda_q^{-1}\left|\mathbf{h}_q\mathbf{w}_p\right|^2 s_p + \sigma_{n,q}^2}$.

Zero-Forcing Beamforming (ZFBF)

- Most popular MU-MIMO precoder. Assume single receive antenna per user.
- Channel Direction Information (CDI) of user q: $\bar{\mathbf{h}}_q = \mathbf{h}_q / \|\mathbf{h}_q\|$.
- Idea is to force the intra-cell interference I_c to zero: the precoder of a user q, $\mathbf{w}_q$, is chosen such that $\mathbf{h}_p \mathbf{w}_q = 0 \ \forall p \in \mathbf{K} \setminus q$. Only possible if $n_e \leq n_t$!
- Define

$$\mathbf{H} = \left[\Lambda_i^{-1/2} \mathbf{h}_i^T, \ldots, \Lambda_j^{-1/2} \mathbf{h}_j^T\right]_{i,j \in \mathbf{K}}^T = \mathbf{D}\bar{\mathbf{H}}$$

with

$$\mathbf{D} = \text{diag}\left\{\Lambda_i^{-1/2} \|\mathbf{h}_i\|, \ldots, \Lambda_j^{-1/2} \|\mathbf{h}_j\|\right\}_{i,j \in \mathbf{K}},$$

$$\bar{\mathbf{H}} = \left[\bar{\mathbf{h}}_i^T, \ldots, \bar{\mathbf{h}}_j^T\right]_{i,j \in \mathbf{K}}^T.$$

The ZFBF aims at designing $\mathbf{W} = [\mathbf{w}_i, \ldots, \mathbf{w}_j]_{i,j \in \mathbf{K}}$ such that $\mathbf{HW}$ is diagonal.

- Assuming $n_e \leq n_t$ and $\bar{\mathbf{H}}$ is full rank, the precoders can be chosen as the normalized columns of the right pseudo inverse of $\mathbf{H}$

$$\mathbf{F} = \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1} = \bar{\mathbf{F}}\mathbf{D}^{-1} = \bar{\mathbf{H}}^H \left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H\right)^{-1} \mathbf{D}^{-1}.$$

Transmit precoder $\mathbf{w}_q$ for user $q \in \mathbf{K}$: $\mathbf{w}_q = \mathbf{F}(:,q)/\|\mathbf{F}(:,q)\| = \bar{\mathbf{F}}(:,q)/\|\bar{\mathbf{F}}(:,q)\|$ where $\mathbf{F}(:,q)$ is to be viewed as the column of $\mathbf{F}$ corresponding to user q.

Zero-Forcing Beamforming (ZFBF)

- Assuming that $\mathbf{c} = \begin{bmatrix} c_i, \ldots, c_j \end{bmatrix}^T_{i,j \in \mathbf{K}'}$ the received signal of user $q \in \mathbf{K}$ is

$$\mathbf{y}_q = \Lambda_q^{-1/2} \mathbf{h}_q \mathbf{w}_q s_q^{1/2} c_q + n_q = d_q c_q + n_q,$$

  with $d_q = \Lambda_q^{-1/2} \mathbf{h}_q \mathbf{w}_q s_q^{1/2} = \Lambda_q^{-1/2} \frac{\left\|\mathbf{h}_q\right\|}{\left\|\bar{\mathbf{F}}(:,q)\right\|} s_q^{1/2}$.
  Observations: MU-MIMO channel with ZFBF is split into n_e parallel (non-interfering) channels.
- The rate achievable by user q is given by

$$R_q = \log_2 \left(1 + d_q^2 / \sigma_{n,q}^2\right).$$

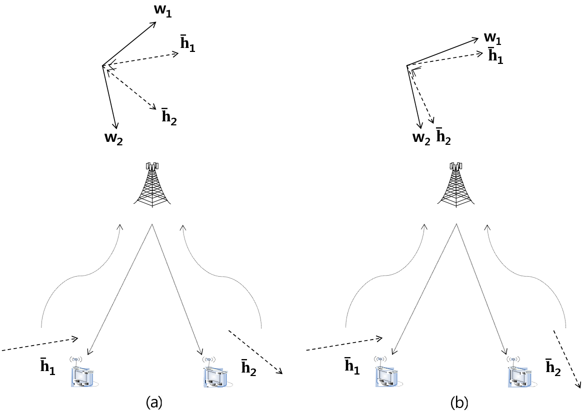
  d_q^2 is low if $\mathbf{H}$ is badly conditioned but would get larger if users' CDI are orthogonal or quasi-orthogonal.
  - reminiscent of the loss caused by noise enhancement incurred by the linear ZF
- For large K, better conditioning of matrix $\mathbf{H}$ through the use of user grouping.
- By uniformly allocating the power across user streams $s_q = E_s / n_e$ and by choosing $n_e = \min\{n_t, K\}$, $d_q^2 / \sigma_{n,q}^2 = \alpha_q^2 \eta_q / n_e$ with $\alpha_q^2 = \left|\mathbf{h}_q \mathbf{w}_q\right|^2 = \left\|\mathbf{h}_q\right\|^2 / \left\|\bar{\mathbf{F}}(:,q)\right\|^2$

$$C_{BF}(\mathbf{H}) = \sum_{q=1}^{\min\{n_t, K\}} \log_2 \left(1 + \alpha_q^2 \frac{\eta_q}{\tilde{n}}\right).$$

  At high SNR with $\eta_q = \eta$, $C_{BF}(\mathbf{H}) \approx \min\{n_t, K\} \log_2(\eta_q)$. The multiplexing gain $\min\{n_t, K\}$ is achieved (same as with DPC).

Zero-Forcing Beamforming (ZFBF)

- Illustration of ZFBF precoding for a two-user scenario: (a) non-orthogonal user set, (b) quasi-orthogonal user set.



Block Diagonalization (BD)

- Extension of ZFBF to multiple receive antennas and multiple streams per user.
- Constraints on the transmit filters targeting user $q \in \mathbf{K}$

$$\Lambda_p^{-1/2} \mathbf{H}_p \mathbf{W}_q = \mathbf{0}, \forall p \neq q, p \in \mathbf{K}$$

- Denoting $\tilde{\mathbf{K}}_q = \mathbf{K} \setminus q$ of size $\tilde{K}_q = \sharp\tilde{\mathbf{K}}_q$, the interference space $\tilde{\mathbf{H}}_q \in {}^{n_r \tilde{K}_q \times n_t}$ is

$$\tilde{\mathbf{H}}_q = \begin{bmatrix} \dots & \Lambda_p^{-1/2} \mathbf{H}_p^T & \dots \end{bmatrix}_{p \in \tilde{\mathbf{K}}_q}^T.$$

- BD filter design forces $\mathbf{W}_q$ to lie in the null space of $\tilde{\mathbf{H}}_q$: null space of $\tilde{\mathbf{H}}_q$ to be strictly larger than 0 $\rightarrow r\left(\tilde{\mathbf{H}}_q\right) < n_t$.
- An orthogonal basis of the null space of $\tilde{\mathbf{H}}_q$ is obtained by taking its SVD

$$\tilde{\mathbf{H}}_q = \tilde{\mathbf{U}}_q \tilde{\mathbf{\Lambda}}_q \begin{bmatrix} \tilde{\mathbf{V}}_q & \tilde{\mathbf{V}}_q' \end{bmatrix}^H$$

  where $\tilde{\mathbf{V}}_q'$ refers to the eigenvectors of $\tilde{\mathbf{H}}_q$ associated with the null singular values.
- Assuming the zero-interference constraint is possible for all users in $\tilde{\mathbf{K}}_q$ and that $r\left(\mathbf{H}_q \tilde{\mathbf{V}}_q'\right) = n_{u,q}$, $\mathbf{W}_q$ writes as a linear combination of columns of $\tilde{\mathbf{V}}_q'$ as

$$\mathbf{W}_q = \tilde{\mathbf{V}}_q' \mathbf{A}_q$$

  with some $n_{u,q} \times n_{u,q}$ unitary matrix $\mathbf{A}_q$.

Block Diagonalization (BD)

- Multi-user interference is eliminated and each user experiences an equivalent single-user MIMO channel $\tilde{\mathbf{H}}_{eq,q} = \mathbf{H}_q \tilde{\mathbf{V}}'_q$, for which the optimal solution is obtained by transmitting along the $n_{u,q}$ dominant right singular vectors of $\tilde{\mathbf{H}}_{eq,q}$

$$\tilde{\mathbf{H}}_{eq,q} = \begin{bmatrix} \tilde{\mathbf{U}}_{eq,q} & \tilde{\mathbf{U}}'_{eq,q} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Lambda}}_{eq,q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_{eq,q} & \tilde{\mathbf{V}}'_{eq,q} \end{bmatrix}^H$$

  where $\tilde{\mathbf{V}}_{eq,q}$ refers to the $n_{u,q}$ dominant right singular vectors.
- The final beamformer for user q writes as

$$\mathbf{W}_q = \tilde{\mathbf{V}}'_q \tilde{\mathbf{V}}_{eq,q}.$$

- Applying $\mathbf{G}_q = \tilde{\mathbf{U}}_{eq,q}^H$, the equivalent channel of each user is

$$\mathbf{z}_q = \mathbf{G}_q \mathbf{y}_q = \Lambda_q^{-1/2} \tilde{\mathbf{\Lambda}}_{eq,q} \mathbf{S}_q^{1/2} \mathbf{c}_q + \mathbf{G}_q \mathbf{n}_q.$$

- Achievable sum-rate (with $\tilde{\lambda}_{eq,q,m}$ diagonal entries of $\tilde{\mathbf{\Lambda}}_{eq,q}^2$)

$$\sum_{q \in \mathbf{K}} \sum_{m=1}^{n_{u,q}} \log_2 \left(1 + s_{q,m} \frac{\Lambda_q^{-1} \tilde{\lambda}_{eq,q,m}}{\sigma_{n,q}^2} \right)$$

  For a sum-power constraint $\sum_{q \in \mathbf{K}} \sum_{m=1}^{n_{u,q}} s_{q,m} = E_s$, the optimal power allocation is obtained by water-filling.

Tomlinson-Harashima Precoding (THP)

- Tomlinson-Harashima Precoding (THP) originally designed to cope with ISI in SISO point-to-point channel when the channel impulse response is known to Tx
  - an alternative to a receiver-based decision-feedback equalizer (DFE).
  - DFE for ISI channels is the analog to the SIC receivers used for MIMO channels

- For a SISO channel, the intended signal at time instant k, $h[0]\,c_k$, is affected by the ISI $i_k = \sum_{l=1}^{L-1} h[l]\,c_{k-l}$ (ignoring the noise)

- If $h[l]\ \forall l$ known to Tx and given that the previous transmitted symbols c_{k-l} are known to Tx, ISI i_k is known and the transmitter can make use of that knowledge.
  - DFE or SIC prone to error propagation, but THP not affected as the previously transmitted symbols c_{k-l} are perfectly known to Tx.

THP

- Assume the received signal

$$y = \tilde{c} + i + n$$

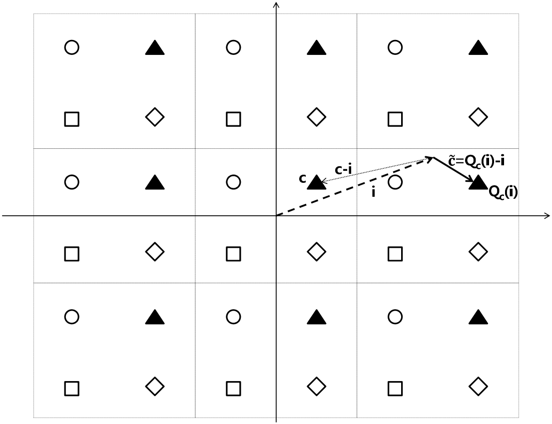
  with
  - $\tilde{c}$ a precoded version of the desired symbol c belonging to a symbol constellation $\mathcal{B}$ with average transmit power E_s,
  - i the interference (e.g. ISI), Gaussian distributed with variance σ_i^2
  - n is the Gaussian noise with variance σ_n^2

- If the Rx treats i as noise, achievable rate given by $\log_2\left(1 + E_s/(\sigma_i^2 + \sigma_n^2)\right)$.

- If i known to Tx and not known at Rx: how to design $\tilde{c}$?
  - Subtracting i to c, i.e. $\tilde{c} = c - i$? No! Rate is $\log_2\left(1 + \left(E_s - \sigma_i^2\right)/\sigma_n^2\right)$ assuming $E_s \geq \sigma_i^2$. Significant power penalty especially for large i.
  - IDEA: THP infinitely replicates the constellation $\mathcal{B}$ and transmits $\tilde{c} = Q_c(i) - i$ where $Q_c(i)$ is the replica of c closest to i.

THP

- Replication of the QPSK constellation made of symbols $\circ$, $\square$, $\lozenge$, $\blacktriangle$: the set of $\blacktriangle$, i.e. $\{\blacktriangle\}$, corresponds to the equivalence class of $\blacktriangle$.



THP

- Set of replicas of a constellation symbol c denoted as the equivalence class of c.
- $Q_c(i)$ is therefore the point in the equivalence class of c closest to i.
- $Q_c(i)$ viewed as a quantizer of i and $\tilde{c} = Q_c(i) - i$ is the quantization error.
- Received signal

  $$y = \tilde{c} + i + n = Q_c(i) + n.$$

  The receiver finds the point in the replicated constellation closest to y and decodes to the equivalence class containing that point.
- The error probability of THP is roughly the same as if $c \in \mathcal{B}$ is transmitted in the absence of interference!
  - The quantization error $Q_c(i) - i$ is always bounded even when i is very large.
  - Constellation points c located at the border of the constellation $\mathcal{B}$ experience a slightly higher decoding error probability due to the presence of the replicas and the probability of confusing c with points belonging to the replicated constellations.
  - In the limit of high SNR, the performance gap is negligible.
  - The power consumption of THP is slightly higher than in the absence of interference: For random interference i, $\mathcal{E}\{|\tilde{c}|^2\}$ is slightly higher than the average power of constellation $\mathcal{B}$.

THP

- At low SNR, enhancement possible by scaling i with a coefficient α

$$\tilde{c} = Q_c\left(\alpha i\right) - \alpha i$$

i.e. the transmitter finds the point in the equivalence class of c closest to αi and transmits the quantization error between that point and αi.

- The receiver scales the received signal by α

$$\alpha y = \alpha\left(\tilde{c} + n\right) + \alpha i$$

and finds the constellation point nearest to αy.

- A suitable value for α is equal to the MMSE scaling factor $E_s/(E_s + \sigma_n^2)$. By doing so, αy is a linear MMSE estimate of $\tilde{c}$ but shifted by αi.

n-dimensional THP and DPC

- System model
$$\mathbf{y} = \tilde{\mathbf{c}} + \mathbf{i} + \mathbf{n}$$
- Choose
$$\tilde{\mathbf{c}} = Q_{\mathbf{c}}\left(\alpha \mathbf{i}\right) - \alpha \mathbf{i}$$
- With high dimensional coding, such precoding technique achieves the same capacity as AWGN channel $\log_2\left(1 + E_s/\sigma_n^2\right)$, i.e. as in the absence of interference.
- This scheme is commonly known as Costa precoding or Dirty-Paper Coding (DPC).
- Alternative representation: view the replicated constellation as a lattice $\mathcal{L}$
  - use a modulo operation such that $\tilde{\mathbf{c}} = [\mathbf{c} - \alpha \mathbf{i}] \bmod \mathcal{L}$ where $[\mathbf{a}] \bmod \mathcal{L} = \mathbf{a} - Q_{\mathcal{L}}\left(\mathbf{a}\right)$ with $Q_{\mathcal{L}}\left(\mathbf{a}\right)$ the lattice vector quantizer based on lattice $\mathcal{L}$.
  - At the receiver, the received signal is passed through the modulo operation $z = \alpha y \bmod \mathcal{L}$ before taking the decision on c.

THP for MU-MIMO

- THP used as a MU-MIMO precoder by precoding sequentially the users data.
  - Assume that the user data are encoded in an increasing order.
  - Any signal generated at step p can be exploited to encode data at step q, for $q > p$.

- Transmit vector as

$$\mathbf{c}' = \mathbf{P}\tilde{\mathbf{c}} = \mathbf{W}\mathbf{S}^{1/2}\tilde{\mathbf{c}}.$$

- Assuming a predefined $\mathbf{K}$ and a predefined user ordering in increasing order of the user index, the signal at user q writes as

$$y_q = \Lambda_q^{-1/2}\mathbf{h}_q\mathbf{w}_q s_q^{1/2}\tilde{c}_q + i_q + \sum_{p>q}\Lambda_q^{-1/2}\mathbf{h}_q\mathbf{w}_p s_p^{1/2}\tilde{c}_p + n_q$$

  where
  - Known interference at Tx: $i_q = \sum_{p<q}\Lambda_q^{-1/2}\mathbf{h}_q\mathbf{w}_p s_p^{1/2}\tilde{c}_p$ (recall that $\tilde{c}_p$, $p < q$, have been previously computed).
  - The interference from $p > q$ is treated as an additional noise.

THP for MU-MIMO

- Relying on THP for SISO channels, we design $\tilde{c}_q$

$$\tilde{c}_q = \left[c_q - \frac{\alpha_q i_q}{\Lambda_q^{-1/2} \mathbf{h}_q \mathbf{w}_q s_q^{1/2}} \right] \bmod \mathcal{L}$$

  where α_q is the scaling factor of user q.
  - Operation repeated sequentially for the n_e data streams to be encoded.
  - THP for MU-MIMO precoding has a very reasonable complexity as it involves only computation of a sequence of n_e complex scalar quantizations (modulo operation).
- At receiver q, the received signal y_q is passed through the modulo operation

$$z_q = \frac{\alpha_q y_q}{\Lambda_q^{-1/2} \mathbf{h}_q \mathbf{w}_q s_q^{1/2}} \bmod \mathcal{L}$$

  before taking the decision on c_q.
- Assuming THP can subtract the interference perfectly, i.e. as DPC, the SINR of user q reads as

$$\rho_q = \frac{\Lambda_q^{-1} \left| \mathbf{h}_q \mathbf{w}_q \right|^2 s_q}{\sum_{p>q} \Lambda_q^{-1} \left| \mathbf{h}_q \mathbf{w}_p \right|^2 s_p + \sigma_{n,q}^2}$$

  and $R_q = \log_2 (1 + \rho_q)$ and $R(\mathbf{K}) = \sum_{q \in \mathbf{K}} w_q R_q$.
  - Note the difference with conventional SINR expression!
  - Note the similarity with capacity of MISO BC
  - Sum-rate further maximized over all unitary beamforming vectors and power allocations satisfying the sum-power constraint.

QR-THP (or ZF-THP)

- Choose the transmit beamformer $\mathbf{W}$ such that the interference due to $p > q$ is eliminated, i.e. zero-forced.
- Let $\mathbf{H} = \mathbf{RQ}$ be the QR decomposition of $\mathbf{H}$
  - $\mathbf{R}$ is a $n_e \times n_e$ lower triangular
  - $\mathbf{Q}$ is a $n_e \times n_t$ matrix with orthonormal rows.
- By selecting $\mathbf{W} = \mathbf{Q}^H$, the system model is simplified as

$$y_q = \Lambda_q^{-1/2} \mathbf{h}_q \mathbf{w}_q s_q^{1/2} \tilde{c}_q + i_q + n_q.$$

  - Since $\mathbf{HW} = \mathbf{R}$ is lower diagonal, the interference caused by users $p > q$ is forced to zero for user q.
- Encoding of $\tilde{c}_q$ can be obtained by successive THP modulo operation.
- The SINR simply boils down to $\rho_q = d_q'^2 / \sigma_{n,q}^2$ with $d_q' = \Lambda_q^{-1/2} \mathbf{h}_q \mathbf{w}_q s_q^{1/2} = \mathbf{R}_{qq} s_q^{1/2}$ with $\mathbf{R}_{q,q}$ the $(q,q)^{th}$ entry of $\mathbf{R}$.
- DPC can be implemented using THP.

Sum-Rate Scaling Laws

- For a large number of users in i.i.d. Rayleigh Fading

Proposition

Assuming $\eta_q = \eta$, $\forall q$, for n_t, n_r and η fixed, the average maximum sum-rates of TDMA, DPC and BF in i.i.d. Rayleigh fading channels (across antennas and users) scales as

$$\bar{C}_{TDMA} \overset{K\nearrow}{\sim} n \log_2\left(1 + \frac{\eta}{n}\log(K)\right)$$

$$\bar{C}_{BC} \overset{K\nearrow}{\sim} \bar{C}_{BF} \overset{K\nearrow}{\sim} n_t \log_2\left(1 + \frac{\eta}{n_t}\log(n_r K)\right)$$

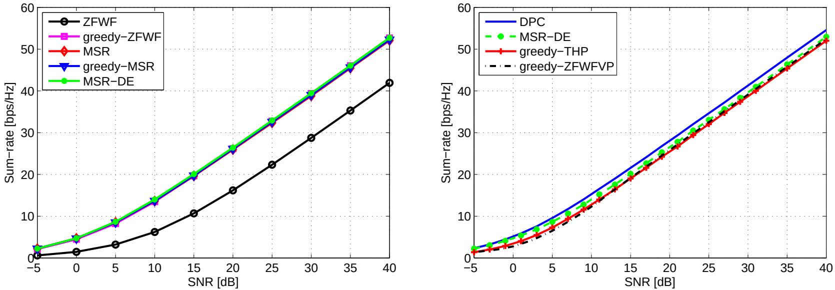
where $n = \min\{n_t, n_r\}$.

Observations:

- $\log(n_r K)$: K users with n_r receive antennas effectively act as a set of $n_r K$ independent users.
- $\bar{C}_{BC} \overset{K\nearrow}{\sim} n_t \log_2 \log K$: full spatial multiplexing gain and MU diversity gain are exploited with DPC.
- $\bar{C}_{BC} \overset{K\nearrow}{\sim} \bar{C}_{BF}$: as K gets large, there is a large probability to find a set of orthogonal user channels to transmit over, each of those channels having a channel gain growing roughly as $\log(K)$.

Global Performance Comparison

- Sum-rate of linear (left) and non-linear (right) MU-MIMO precoders vs SNR in $n_t = 4, K = 20$ i.i.d. Rayleigh fading channels

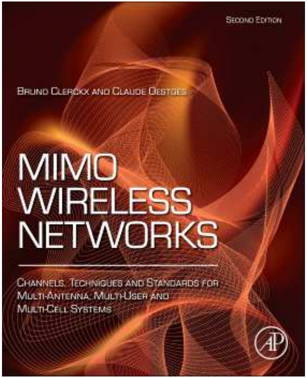


Observations: ZFBF without user selection (ZFWF) performs poorly. ZFBF with user selection (greedy-ZFWF) is a competitive strategy for MU-MIMO broadcast channels, in terms of both performance and complexity.
Keep in mind the assumptions: perfect CSIT, the same average SNR for all users and a max-rate scheduler (i.e. there is no fairness issue involved here).

Multi-User MIMO (Downlink) with Imperfect CSIT

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  – Chapter 12
    - Section: 12.9

Additional References

- M. Maddah-Ali and D. Tse, Completely state transmitter channel state information is still very useful, IEEE Trans. Int. Theory, vol. 58, no. 7, 2012.
- C. Hao and B. Clerckx, Imperfect and Unmatched CSIT is Still Useful for the Frequency Correlated MISO Broadcast Channel, IEEE ICC 2013, June 2013.
- C. Hao and B. Clerckx, MISO Broadcast Channel with Imperfect and (Un)matched CSIT in the Frequency Domain: DoF Region and Transmission Strategies, IEEE PIMRC 2013, September 2013.
- R. Tandon, S. Jafar, S. Shamai Shitz, and H. Poor, "On the synergistic benefits of alternating CSIT for the MISO broadcast channel," IEEE Trans. Inf. Theory., vol. 59, no. 7, pp. 4106-4128, 2013.
- J. Chen and P. Elia, Optimal DoF Region of the two-user MISO BC with general alternating CSIT, ArXiv 2013.

Introduction

- In practice, perfect CSIT is hard to obtain in both FDD and TDD systems.
  - TDD could make use of reciprocity
- Only partial CSIT is available at the transmitter.
- There are two major impairments that prevent from obtaining perfect CSIT:
- inaccurate CSI measurement and feedback (due to channel estimation errors and limited feedback overhead)
- feedback delay (due to processing delay at the mobile and the BS and the frame structure).

Quantized Feedback-Based Precoding

- ZFBF with uniform power allocation, single receive antenna per mobile and quantized feedback $\hat{\mathbf{h}}_q$

$$\mathbf{w}_q = \hat{\mathbf{F}}(:,q)/\|\hat{\mathbf{F}}(:,q)\|$$

  where

$$\hat{\mathbf{F}} = \hat{\mathbf{H}}^H \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^H\right)^{-1} \mathbf{D}^{-1}$$

  with

$$\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_i^T, \ldots, \hat{\mathbf{h}}_j^T\right]^T_{i,j \in \mathbf{K}}$$

- Every user q quantizes its channel using a B_q-bits codebook $\mathcal{W}_q$ of codevectors $\mathbf{v}_{q,i}$, $i = 1, \ldots, n_{p,q} = 2^{B_q}$. the best codevector $\mathbf{v}_q^\star$ for user q is selected as

$$\mathbf{v}_q^\star = \arg \max_{1 \le i \le n_{p,q}} |\mathbf{h}_q \mathbf{v}_{q,i}|^2$$

  The quantized version of the channel direction writes as the $n_t \times 1$ row vector $\hat{\mathbf{h}}_q = (\mathbf{v}_q^\star)^H$.
- With imperfect CSIT, multi-user interference cannot be canceled out perfectly by the ZFBF filter.

Sum-Rate Analysis

- Expected rate achieved by user q with perfect CSIT and uniform power allocation (assume n_e co-scheduled users)

$$\bar{R}_{CSIT,q} = \mathcal{E}_{\mathbf{H}} \left\{ \log_2 \left(1 + \frac{\eta_q}{n_e} \left| \mathbf{h}_q \mathbf{w}_{ZF,q} \right|^2 \right) \right\}$$

- Expected rate achieved by user q with quantized CSIT and uniform power allocation

$$\bar{R}_{LF,q} = \mathcal{E}_{\mathbf{H},\mathcal{W}_q} \left\{ \log_2 \left(1 + \rho_q \right) \right\}$$

where

$$\rho_q = \frac{\frac{\eta_q}{n_e} \left| \mathbf{h}_q \mathbf{w}_q \right|^2}{1 + \frac{\eta_q}{n_e} \sum_{p \in \mathbf{K}, p \neq q} \left| \mathbf{h}_q \mathbf{w}_p \right|^2}$$

- The rate loss for user q incurred by the quantized feedback

$$\Delta \bar{R}_q = \bar{R}_{CSIT,q} - \bar{R}_{LF,q}.$$

- A multi-user MIMO scheme with imperfect CSIT/quantized feedback is affected by the quantization error at two levels:
  - Residual interference term in the SINR ρ_q that does not vanish with the SNR η_q, therefore inducing a ceiling effect as the SNR increases
  - Accurate CQI evaluation becomes challenging

Precoding with Partial Transmit Channel Knowledge

- Assume a predefined set of $n_e > 1$ co-scheduled users

Proposition

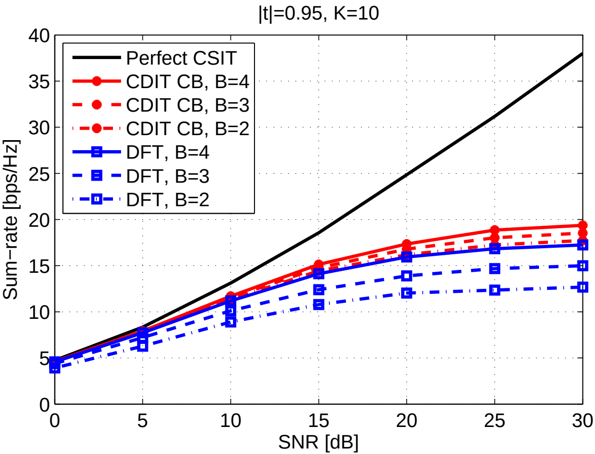
The limited feedback-based ZFBF using a codebook with B_q bits of feedback incurs a rate loss for user q (relative to perfect CSIT-based ZFBF) that is upper bounded as

$$\Delta \bar{R}_q \lesssim \log_2 \left(1 + \frac{\eta_q}{n_e} (n_e - 1)\, d_{f,q}\right)$$

where $d_{f,q}$ is the average distortion function of user q

$$d_f = \mathcal{E}_{\mathbf{h}} \left\{\lambda_{max} - \left\|\mathbf{h}\mathbf{w}^\star\right\|^2\right\} = \mathcal{E}_{\mathbf{h}} \left\{\left\|\mathbf{h}\right\|^2 - \left\|\mathbf{h}\right\|^2 \left|\bar{\mathbf{h}}\mathbf{w}^\star\right|^2\right\}.$$

- Performance of channel statistics-based codebook (CDIT-CB) and DFT codebook with Greedy user selection for $B = 2, 3, 4$, $n_t = 4$, $|t| = 0.95$ and $K = 10$.



Scalable Feedback

- Number of feedback bits necessary to maintain a rate loss of $\Delta\bar{R}_q \leq \log_2(b)$ bps/Hz for user q

Proposition

In order to maintain a rate loss $\Delta\bar{R}_q$ between limited feedback ZFBF and perfect CSIT-based ZFBF smaller than $\log_2(b)$ bps/Hz for user q, the number of feedback bits B_q should scale according to

- *i.i.d. Rayleigh fading channels*

$$B_q \approx (n_t - 1)\log_2(\eta_q) - (n_t - 1)\log_2(b-1).$$

- *spatially correlated Rayleigh fading channels (with CDIT-based codebook)*

$$\begin{aligned} B_q \approx\ & (r_q - 1)\log_2(\eta_q) - (r_q - 1)\log_2(b-1) \\ & + (r_q - 1)\log_2\left(\frac{\sigma_{2,q}^2}{\sigma_{1,q}^2}\right) + (r_q - 1)\log_2\left(\frac{(n_e - 1)\, n_t}{n_e}\right) \end{aligned}$$

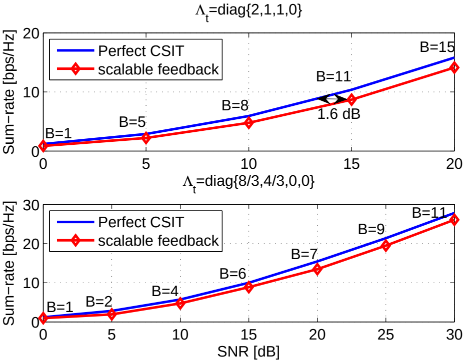
with r_q, $\sigma_{1,q}^2$ and $\sigma_{2,q}^2$ the rank and the two dominant eigenvalues values of user-q transmit correlation matrix.

Scalable Feedback

- Assuming $n_e = n_t$, $\eta_q = \eta$ and $r_q = r\ \forall q$

Deployment	DL throughput	UL overhead
i.i.d.	$n_t \log_2(\eta)$	$n_t(n_t - 1)\log_2(\eta)$
Spatially Correlated	$n_t \log_2(\eta)$	$n_t(r - 1)\log_2(\eta)$

- Performance of ZFBF with perfect CSIT and ZFBF with channel statistics-based codebook and scalable feedback without user selection ($n_t = 4$, $n_e = 4$)



Outdated Feedback-Based Precoding

- Is outdated CSIT useless?
  - No! In a two-user MISO BC with outdated CSIT, a sum DoF of 4/3 can be achieved, a 33% DoF enhancement compared to conventional TDMA approach.
- Assume a two-user two transmit antenna MISO BC with delayed CSIT.
  - The transmission occurs over three coherence times.
  - Each coherence time is made of T time slots over which the channel is constant.
  - Channel vector of user 1 on coherence time k as $\mathbf{h}_k = \begin{bmatrix} h_{k,1} & h_{k,2} \end{bmatrix}$, $\mathbf{g}_k = \begin{bmatrix} g_{k,1} & g_{k,2} \end{bmatrix}$ for user 2.
  - Channel coefficients are assumed constant within a coherence time and change independently from one coherence time to the next one.
  - The CSI is assumed to be available at Tx only at the next coherence time.
- Denoting the transmit signal on time slot t of coherence time k as $\mathbf{x}_{k,t}$, the received signals at user 1 and 2, respectively denoted as $y_{k,t}$ and $z_{k,t}$, write as

$$y_{k,t} = \mathbf{h}_k \mathbf{x}_{k,t} + n_{k,t},$$
$$z_{k,t} = \mathbf{g}_k \mathbf{x}_{k,t} + w_{k,t},$$

where $n_{k,t} \sim \mathcal{CN}(0,1)$ and $w_{k,t} \sim \mathcal{CN}(0,1)$ are AWGN. We consider a long-term power constraint $\mathcal{E}\left\{\mathbf{x}_{k,t}^H \mathbf{x}_{k,t}\right\} \leq \rho_{lt}$.

Outdated Feedback-Based Precoding

- Consider two independent $2 \times T$ codewords, $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_T]$ and $\mathbf{C}' = [\mathbf{c}'_1, \ldots, \mathbf{c}'_T]$ respectively intended for user 1 and user 2.
- Normalized such that $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{C}\mathbf{C}^H\right\}\right\} = \mathcal{E}\left\{\text{Tr}\left\{\mathbf{C}'\mathbf{C}'^H\right\}\right\} = T$.
- MAT strategy transmits codeword $\mathbf{C}$ to user 1 and codeword $\mathbf{C}'$ to user 2 over $3T$ time slots using the MAT strategy. On time slot t, MAT consists in transmitting
  1. $\mathbf{x}_{1,t} = \sqrt{\rho}\mathbf{c}_t$ in coherence time 1,
  2. $\mathbf{x}_{2,t} = \sqrt{\rho}\mathbf{c}'_t$ in coherence time 2,
  3. the overheard interference $\mathbf{x}_{3,t} = \sqrt{\rho}\begin{bmatrix} \mathbf{g}_1\mathbf{c}_t + \mathbf{h}_2\mathbf{c}'_t & 0 \end{bmatrix}^T$ in coherence time 3.
- A long-term average transmit power (where averaging is also taken over the channel realizations) of $\rho_{lt} = 4/3\rho$ is consumed and twice as much power is spent on coherence time 3 as in coherence time 1 and 2. We will refer to ρ as the SNR.

Outdated Feedback-Based Precoding

- The equivalent system model for user 1 at time instant $t = 1, \dots, T$ can then be written as

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \sqrt{\rho} \underbrace{\begin{bmatrix} h_{1,1} & h_{1,2} \\ 0 & 0 \\ h_{3,1}g_{1,1} & h_{3,1}g_{1,2} \end{bmatrix}}_{\text{rank two}} \mathbf{c}_t + \sqrt{\rho} \underbrace{\begin{bmatrix} 0 & 0 \\ h_{2,1} & h_{2,2} \\ h_{3,1}h_{2,1} & h_{3,1}h_{2,2} \end{bmatrix}}_{\text{rank one}} \mathbf{c}'_t + \begin{bmatrix} n_{1,t} \\ n_{2,t} \\ n_{3,t} \end{bmatrix}.$$

- Interference elimination

$$\tilde{\mathbf{y}}_t = \begin{bmatrix} y_{1,t} \\ y_{3,t} - h_{3,1}y_{2,t} \end{bmatrix} = \sqrt{\rho} \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{3,1}g_{1,1} & h_{3,1}g_{1,2} \end{bmatrix} \mathbf{c}_t + \begin{bmatrix} n_{1,t} \\ n_{3,t} - h_{3,1}n_{2,t} \end{bmatrix}.$$

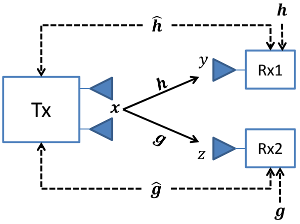
  This is an equivalent 2×2 MIMO channel. User 1 decodes 2 symbols in 3 time slots.

- Similar for user 2. Hence, per user DoF of 2/3 and sum DoF of 4/3.
- Extendable to K-user MISO BC: sum DoF of $\frac{K}{1+\frac{1}{2}+\ldots+\frac{1}{K}} \approx \frac{K}{\ln K}$.

Imperfect CSIT-Based Precoding

- Design MU-MIMO for imperfect CSIT
- Given imperfect feedback in the frequency domain,
  - What is the maximum achievable rate region?
  - What are the optimal/suboptimal transmission and reception strategies?
  - How to optimally make use of feedback resources?

- Consider a two-user MISO-OFDMA Broadcast Channel, with arbitrary values of the CSIT qualities across L subbands.



Qualities of the Imperfect CSIT

	User 1	User 2
Subband 1	a_1	b_1
	⋮	⋮
Subband j	a_j	b_j
	⋮	⋮
Subband L	a_L	b_L

$$a_j = \lim_{P \to \infty} \frac{\log E\left[\left\|h_j - \hat{h}_j\right\|^2\right]}{\log P}$$

$$b_j = \lim_{P \to \infty} \frac{\log E\left[\left\|g_j - \hat{g}_j\right\|^2\right]}{\log P}$$

System Model

- Transmit signal vector in subband j denoted as $\mathbf{x}_j$ subject ot a per-subband based power constraint $\mathcal{E}\left[\,\|\mathbf{x}_j\|^2\,\right]\sim P$ (P is the SNR).
- The observations at user 1 and 2, y_j and z_j respectively, are given by

$$y_j = \mathbf{h}_j^H \mathbf{x}_j + \epsilon_{j1},$$
$$z_j = \mathbf{g}_j^H \mathbf{x}_j + \epsilon_{j2},$$

  where ϵ_{j1} and ϵ_{j2} are unit power AWGN noise.
- $\mathbf{h}_j$ and $\mathbf{g}_j$ are the CSI in subband i of user 1 and user 2, respectively. The CSI are i.i.d. across users and subbands.
- Imperfect CSIT: $\hat{\mathbf{h}}_j$ of user 1 and $\hat{\mathbf{g}}_j$ of user 2. Error vectors $\tilde{\mathbf{h}}_j = \mathbf{h}_j - \hat{\mathbf{h}}_j$ and $\tilde{\mathbf{g}}_j = \mathbf{g}_j - \hat{\mathbf{g}}_j$ with the covariance matrix $\mathbb{E}[\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^H] = \sigma_{j1}^2\mathbf{I}_2$ and $\mathbb{E}[\tilde{\mathbf{g}}_j\tilde{\mathbf{g}}_j^H] = \sigma_{j2}^2\mathbf{I}_2$.
  - $\sigma_{j1}^2 \sim P^{-a_j}$ and $\sigma_{j2}^2 \sim P^{-b_j}$. a_j and b_j are respectively interpreted as the quality of the CSIT of user 1 and user 2 in subband j, given as follows

$$a_j = \lim_{P\to\infty} -\frac{\log \sigma_{j1}^2}{\log P}, \quad b_j = \lim_{P\to\infty} -\frac{\log \sigma_{j2}^2}{\log P}.$$

  - a_j and b_j vary within the range of $[0,1]$. $a_j = 1$ (resp. $b_j = 1$) is equivalent to perfect CSIT and $a_j = 0$ (resp. $b_j = 0$) is equivalent to no CSIT.
- DoF per user and per channel use (assuming S channel uses)

$$d_k \triangleq \lim_{P\to\infty} \frac{R_k}{S \log P}, \quad k = 1, 2,$$

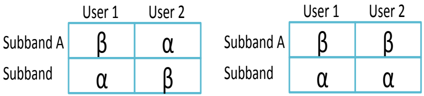
System Model

- Note $\mathcal{E}[|\mathbf{h}_j^H \hat{\mathbf{h}}_j^\perp|^2] = \mathcal{E}[|(\hat{\mathbf{h}}_j + \tilde{\mathbf{h}}_j)^H \hat{\mathbf{h}}_j^\perp|^2] = \mathcal{E}[|\tilde{\mathbf{h}}_j^H \hat{\mathbf{h}}_j^\perp|^2] = \mathcal{E}[\tilde{\mathbf{h}}_j^H \hat{\mathbf{h}}_j^\perp \hat{\mathbf{h}}_j^{\perp H} \tilde{\mathbf{h}}_j] \sim P^{-a_j}$. and $\mathcal{E}[|\mathbf{g}_j^H \hat{\mathbf{g}}_j^\perp|^2] \sim P^{-b_j}$.
- The average CSIT quality of user 1 and user 2 are respectively expressed as $a_e = \frac{1}{L}\sum_{j=1}^{L} a_j$ and $b_e = \frac{1}{L}\sum_{j=1}^{L} b_j$.

Definition

$\mathcal{P}_L$ Problem: Find transmission strategies that maximize the DoF region in a scenario such that $a_e = b_e$.

- $L = 1$: $a_j = b_j \ \forall j$.
- $L = 2$: $a_1 + a_2 = b_1 + b_2$
  - $a_1 = b_1$ and $a_2 = b_2$: two $\mathcal{P}_1$ on each subband
  - $a_1 \neq b_1$ and $a_2 \neq b_2$: $\mathcal{P}_2$ where the transmitted signal in each subband is correlated to each other



(a) Unmatched (b) Matched

Figure: Two-subband based frequency correlated BC.

ZFBF

- ZFBF: designed for perfect CSIT $(a_j = b_j = 1)$
- Transmit signal on subband j with u_j for user 1 and v_j for user 2

$$\mathbf{x}_j = \hat{\mathbf{g}}_j^{\perp} u_j + \hat{\mathbf{h}}_j^{\perp} v_j$$

- Received signals on subband j

$$y_j = \mathbf{h}_j^H \hat{\mathbf{g}}_j^{\perp} u_j + \mathbf{h}_j^H \hat{\mathbf{h}}_j^{\perp} v_j + \epsilon_{j1}$$
$$z_j = \mathbf{g}_j^H \hat{\mathbf{g}}_j^{\perp} u_j + \mathbf{g}_j^H \hat{\mathbf{h}}_j^{\perp} v_j + \epsilon_{j2}$$

  leading to a sum DoF of 2.
- If imperfect CSIT: $a_j = \beta$ for user 1 and $b_j = \alpha$ for user 2

$$y_j = \underbrace{\mathbf{h}_j^H \hat{\mathbf{g}}_j^{\perp} u_j}_{P} + \underbrace{\mathbf{h}_j^H \hat{\mathbf{h}}_j^{\perp} v_j}_{P^{1-\beta}} + \epsilon_{j1}$$
$$z_j = \underbrace{\mathbf{g}_j^H \hat{\mathbf{g}}_j^{\perp} u_j}_{P^{1-\alpha}} + \underbrace{\mathbf{g}_j^H \hat{\mathbf{h}}_j^{\perp} v_j}_{P} + \epsilon_{j2}$$

  leading to a sum DoF of $\beta + \alpha$ (rate of u_j + rate of v_j)

$S_3^{3/2}$

- $S_3^{3/2}$: designed for alternating CSIT in a two subband case, i.e. the transmitter has perfect CSIT of only one user at a time $(a_1 = 1, b_1 = 0, a_2 = 0, b_2 = 1)$
- Transmit signal on subband 1 and 2 with u_0 to be decoded by both users but intended to user 1 or 2 and u_2 intended for user 1 and v_1 intended for user 2.

$$\mathbf{x}_1 = [u_0, 0]^T + \hat{\mathbf{h}}_1^{\perp} v_1,$$
$$\mathbf{x}_2 = [u_0, 0]^T + \hat{\mathbf{g}}_2^{\perp} u_2.$$

- Received signals at each user

$$y_1 = \underbrace{h_{11}^* u_0}_{P} + \underbrace{\mathbf{h}_1^H \hat{\mathbf{h}}_1^{\perp} v_1}_{P^0} + \epsilon_{11}, \qquad z_1 = \underbrace{g_{11}^* u_0}_{P} + \underbrace{\mathbf{g}_1^H \hat{\mathbf{h}}_1^{\perp} v_1}_{P} + \epsilon_{12},$$
$$y_2 = \underbrace{h_{21}^* u_0}_{P} + \underbrace{\mathbf{h}_2^H \hat{\mathbf{g}}_2^{\perp} u_2}_{P} + \epsilon_{21}, \qquad z_2 = \underbrace{g_{21}^* u_0}_{P} + \underbrace{\mathbf{g}_2^H \hat{\mathbf{g}}_2^{\perp} u_2}_{P^0} + \epsilon_{22},$$

leading to a sum DoF of 3/2.

$S_3^{3/2}$

- If imperfect CSIT: $a_1 = b_2 = \beta, b_1 = a_2 = \alpha$

$$y_1 = \underbrace{h_{11}^* u_0}_{P} + \underbrace{\mathbf{h}_1^H \hat{\mathbf{h}}_1^\perp v_1}_{P^{1-\beta}} + \epsilon_{11}, \qquad z_1 = \underbrace{g_{11}^* u_0}_{P} + \underbrace{\mathbf{g}_1^H \hat{\mathbf{h}}_1^\perp v_1}_{P} + \epsilon_{12},$$

$$y_2 = \underbrace{h_{21}^* u_0}_{P} + \underbrace{\mathbf{h}_2^H \hat{\mathbf{g}}_2^\perp u_2}_{P} + \epsilon_{21}, \qquad z_2 = \underbrace{g_{21}^* u_0}_{P} + \underbrace{\mathbf{g}_2^H \hat{\mathbf{g}}_2^\perp u_2}_{P^{1-\beta}} + \epsilon_{22},$$

leading to a sum DoF of $1/2(\beta[\text{rate of} u_0] + 1[\text{rate of} v_1] + 1[\text{rate of} u_2]) = 1 + \beta/2$.

$\mathcal{P}_1$

- Note ZFBF performs a space-only precoding while $S_3^{3/2}$ performs a space-frequency/time precoding
- $\mathcal{P}_1$: Can we do better than ZFBF when $a_j = b_j \ \forall j$?
- Transmit the signal in each subband by superposing a common message with ZFBF-precoded private messages. Focus on subband 1 for simplicity

$$\mathbf{x}_1 = [\underbrace{c_1}_{P - P^{a_1}}, 0]^T + \underbrace{\hat{\mathbf{g}}_1^{\perp} u_1}_{P^{a_1}/2} + \underbrace{\hat{\mathbf{h}}_1^{\perp} v_1}_{P^{a_1}/2},$$

  where c_1 is the common message broadcast to both users and u_1 and v_1 are symbols intended for user 1 and user 2 respectively.
- Integration of broadcasting (ZFBF) and multicating/FDMA.
- Received signal at each user

$$y_1 = \underbrace{h_{11}^* c_1}_{P} + \underbrace{\mathbf{h}_1^H \hat{\mathbf{g}}_1^{\perp} u_1}_{P^{a_1}/2} + \underbrace{\mathbf{h}_1^H \hat{\mathbf{h}}_1^{\perp} v_1}_{P^0} + \underbrace{\epsilon_{11}}_{P^0}, \qquad z_1 = \underbrace{g_{11}^* c_1}_{P} + \underbrace{\mathbf{g}_1^H \hat{\mathbf{g}}_1^{\perp} u_1}_{P^0} + \underbrace{\mathbf{g}_1^H \hat{\mathbf{h}}_1^{\perp} v_1}_{P^{a_1}/2} + \underbrace{\epsilon_{12}}_{P^0},$$

  where the private symbols u_1 and v_1 are drowned by the noise respectively at user 2 and user 1 due to partial ZFBF.

$\mathcal{P}_1$

- Decodability:
  - Both users decode the common message first with rate $(1-a_1)\log P$ by treating the private message as noise,
  - Using SIC, each user can decode their private message with rate $a_1 \log P$ only subject to noise, after removing the common message,
- Assume $a_j = b_j = \beta$, $\mathcal{P}_\infty$ leads to a sum DoF of $1-\beta+\beta+\beta = 1+\beta > 2\beta$ achieved by ZFBF only.
  - The DoF pairs $(1,\beta)$ and $(\beta,1)$ are achieved if we consider the common message is intended for user 1 and user 2 respectively.
- Assume $a_1 = b_1 = \beta$ and $a_2 = b_2 = \alpha$, sum DoF of $1/2(1-\beta+\beta+\beta+1-\alpha+\alpha+\alpha) = 1+\frac{\beta+\alpha}{2} > \beta+\alpha$ achieved by ZFBF only.

$\mathcal{P}_2$

- $\mathcal{P}_2$ problem with $L=2$ such that $a_e = b_2$ with $a_1 \geq b_1$ and $a_2 \leq b_2$: $a_1 = b_2 = \beta$ and $a_2 = b_1 = \alpha$ and $\beta \geq \alpha$
- The transmission blocks in subband 1 and 2 are expressed as

$$\mathbf{x}_1 = [c_1, 0]^T + \hat{\mathbf{g}}_1^{\perp} u_1 + [u_0, 0]^T + \hat{\mathbf{h}}_1^{\perp} v_1,$$
$$\mathbf{x}_2 = [c_2, 0]^T + \hat{\mathbf{h}}_2^{\perp} v_2 + [u_0, 0]^T + \hat{\mathbf{g}}_2^{\perp} u_2.$$

  where
  - Common messages u_0, c_1 and c_2 to be decoded by both users (intended for user 1 and user 2 respectively or exclusively for user 1 or user 2 or for both users).
  - Note that we do not precode common messages as it does not impact the DoF.
  - u_1 and u_2 are symbols intended for user 1, while v_1 and v_2 are symbols intended for user 2.
- Integrate ZFBF, $S_3^{3/2}$ and FDMA/multicasting.
- Power and rate allocation

subband 1	Power	Rate	subband 2	Power	Rate
c_1	$P - P^{\beta}$	$1-\beta$	c_2	$P - P^{\beta}$	$1-\beta$
u_1	$P^{\alpha}/2$	α	u_2	$P^{\beta}/2$	β
u_0	$(P^{\beta} - P^{\alpha})/2$	$\beta-\alpha$	u_0	$(P^{\beta} - P^{\alpha})/2$	$\beta-\alpha$
v_1	$P^{\beta}/2$	β	v_2	$P^{\alpha}/2$	α

$\mathcal{P}_2$

- Received signals at each user

$$y_1 = \underbrace{h_{11}^* c_1}_{P} + \underbrace{\mathbf{h}_1^H \hat{\mathbf{g}}_1^\perp u_1}_{P^\alpha} + \underbrace{h_{11}^* u_0}_{P^\beta} + \underbrace{\mathbf{h}_1^H \hat{\mathbf{h}}_1^\perp v_1}_{P^0} + \epsilon_{11},$$

$$z_1 = \underbrace{g_{11}^* c_1}_{P} + \underbrace{\mathbf{g}_1^H \hat{\mathbf{g}}_1^\perp u_1}_{P^0} + \underbrace{g_{11}^* u_0}_{P^\beta} + \underbrace{\mathbf{g}_1^H \hat{\mathbf{h}}_1^\perp v_1}_{P^\beta} + \epsilon_{12},$$

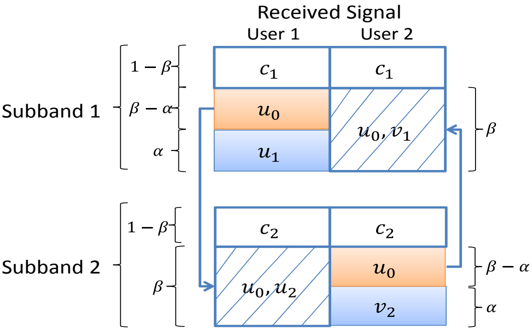
$$y_2 = \underbrace{h_{21}^* c_2}_{P} + \underbrace{\mathbf{h}_2^H \hat{\mathbf{g}}_2^\perp u_2}_{P^\beta} + \underbrace{h_{21}^* u_0}_{P^\beta} + \underbrace{\mathbf{h}_2^H \hat{\mathbf{h}}_2^\perp v_2}_{P^0} + \epsilon_{21},$$

$$z_2 = \underbrace{g_{21}^* c_2}_{P} + \underbrace{\mathbf{g}_2^H \hat{\mathbf{g}}_2^\perp u_2}_{P^0} + \underbrace{g_{21}^* u_0}_{P^\beta} + \underbrace{\mathbf{g}_2^H \hat{\mathbf{h}}_2^\perp v_2}_{P^\alpha} + \epsilon_{22}.$$

- Decodability:
  - c_1 and c_2 are respectively decoded first by treating all the other terms as noise.
  - Afterwards, user 1 decodes u_0 and u_1 from y_1 using SIC. With the knowledge of u_0, u_2 can be recovered from y_2.
  - Similarly, user 2 decodes u_0 and v_2 from z_2 via SIC. v_1 can be decoded from z_1 by eliminating u_0.
- Sum DoF of $1 + \frac{\alpha+\beta}{2}$
  - The DoF pair $(1, \frac{\alpha+\beta}{2})$ and $(\frac{\alpha+\beta}{2}, 1)$ are achieved if we consider the common messages are intended for user 1 and user 2 respectively.
- When $\beta = \alpha$, $\mathcal{P}_2$ degrades to 2 parallel $\mathcal{P}_1$ and no common message u_0 is sent.

$\mathcal{P}_2$

- Received signal and decoding procedure of the optimal scheme for the $\mathcal{P}_2$



- The key point: the transmitter broadcasts u_0 twice, i.e. subband 1 and 2.
  - User 1 (resp. user 2) observes u_0 with higher power than u_1 (resp. v_2) in subband 1 (resp. 2) and receives u_0 with the same power level as u_2 (resp. v_1) in subband 2 (resp. 1).
  - The common message u_0 can be decoded by both users but in different subbands.
  - Can be generalized to solve $\mathcal{P}_L, L \geq 3$ problem by generating multiple streams of u_0 and sending each of them twice.

Weighted-Sum DoF Interpretation of $\mathcal{P}_2$ region

- Decompose the two subbands $\mathcal{P}_2$ with $a_1 = b_2 = \beta$ and $a_2 = b_1 = \alpha$

	User 1	User 2
Subband A	β	α
Subband	α	β

into subchannels

$$\alpha \left\{ \begin{array}{c|c|c} & \text{User 1} & \text{User 2} \\ \hline \bar{A} & 1 & 1 \\ \bar{B} & 1 & 1 \end{array} \right. + \beta\text{-}\alpha \left\{ \begin{array}{c|c|c} & \text{User 1} & \text{User 2} \\ \hline \hat{A} & 1 & 0 \\ \hat{B} & 0 & 1 \end{array} \right. + 1\text{-}\beta \left\{ \begin{array}{c|c|c} & \text{User 1} & \text{User 2} \\ \hline \tilde{A} & 0 & 0 \\ \tilde{A} & 0 & 0 \end{array} \right.$$

where

- $\tilde{A}$, $\tilde{B}$: no CSIT, each with channel use $1-\beta$;
- $\hat{A}$ $(\hat{B})$: perfect CSIT of user 1 (2), with channel use $\beta-\alpha$;
- $\bar{A}$, $\bar{B}$: perfect CSIT of both users, with channel use α.

Weighted-Sum DoF Interpretation of $\mathcal{P}_2$ region

- DoF region of $\mathcal{P}_2$ can be interpreted as a weighted-sum representation of the DoF region of each subchannel

$$\mathcal{D}_u = (1-\beta)\tilde{\mathcal{D}} + (\beta-\alpha)\hat{\mathcal{D}} + \alpha\bar{\mathcal{D}}$$

  - Subchannel $\tilde{A}$ and $\tilde{B}$ are the BC with no CSIT

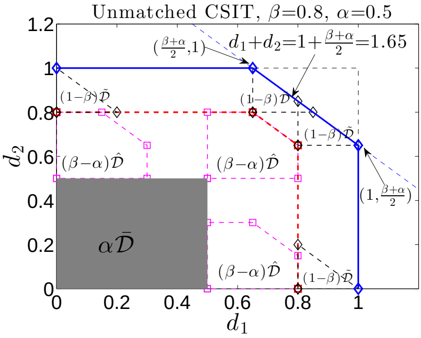
$$\mathcal{D}^{\tilde{A}} = \mathcal{D}^{\tilde{B}} = \tilde{\mathcal{D}} : d_1 + d_2 \leq 1.$$

  - Subchannel $\bar{A}$ and $\bar{B}$ are the BC with perfect CSIT of both users

$$\mathcal{D}^{\bar{A}} = \mathcal{D}^{\bar{B}} = \bar{\mathcal{D}} : d_1 \leq 1, d_2 \leq 1.$$

  - Subchannel $\hat{A}$ and $\hat{B}$ have an alternating CSIT setting with two states: $I_1I_2{=}PN$ and $I_1I_2{=}NP$ (as in $S_3^{3/2}$)

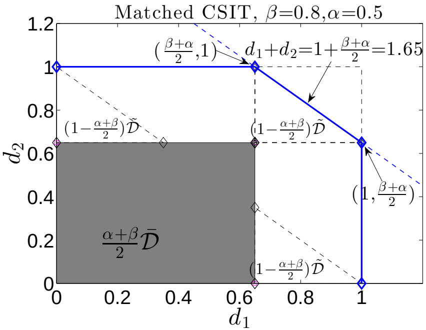
$$(\mathcal{D}^{\hat{A}}+\mathcal{D}^{\hat{B}})/2 = \hat{\mathcal{D}} : d_1 + d_2 \leq 1.5, \quad d_1 \leq 1, d_2 \leq 1.$$



Weighted-Sum DoF Interpretation of $\mathcal{P}_1$ region

- Decompose the subband into subchannels
  - $\tilde{A}$, $\tilde{B}$: no CSIT, each with channel use $1-\beta$ and $1-\alpha$;
  - $\bar{A}$, $\bar{B}$: perfect CSIT of both users, with channel use β and α respectively.
- DoF region of $\mathcal{P}_1$ can be interpreted as a weighted-sum representation of the DoF region of each subchannel

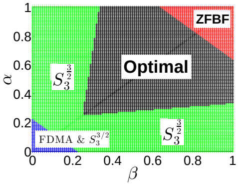
$$\mathcal{D}_m = (1 - \frac{\beta+\alpha}{2})\tilde{\mathcal{D}} + \frac{\beta+\alpha}{2}\bar{\mathcal{D}}.$$



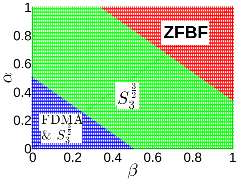
Mode switching among sub-optimal strategies

- $\mathcal{P}_2$ integrates FDMA/multicast, ZFBF, $S_3^{3/2}$. What about a simple switching strategy?
  - FDMA only: sum DoF $d_\Sigma^F{=}1$
  - ZFBF only: sum DoF $d_\Sigma^Z{=}\beta{+}\alpha$
  - $S_3^{3/2}$ only: sum DoF $d_\Sigma^S{=}1{+}\frac{\beta}{2}$
- The best strategy among the 3 sub-optimal strategies can achieve at least 80% of the optimal sum DoF performance as

$$\max(d_\Sigma^F, d_\Sigma^Z, d_\Sigma^S) \geq 0.8 \times d_\Sigma^{opt}, \forall \beta, \alpha \in [0,1].$$



(a) $\rho{=}90\%$



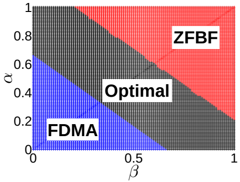
(b) $\rho{=}80\%$

Figure: Unmatched case, switching among FDMA, ZFBF and $S_3^{3/2}$.

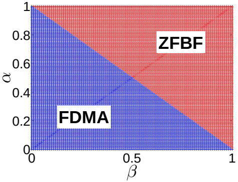
Mode switching among sub-optimal strategies

- $\mathcal{P}_1$ integrates FDMA/multicast, ZFBF.
- The best strategy among the 2 sub-optimal strategies can achieve at least 66.7% of the optimal sum DoF performance as

$$\max(d_\Sigma^F, d_\Sigma^Z) \geq 2/3 \times d_\Sigma^{opt}, \forall \beta, \alpha \in [0,1].$$



(a) ρ=75%



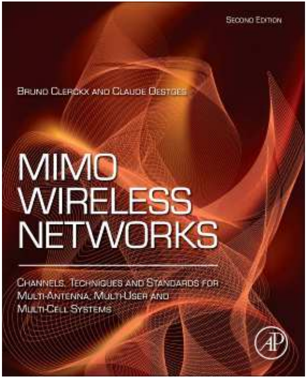
(b) ρ=66.7%

Figure: Matched case, switching among FDMA and ZFBF.

Introduction to Multi-Cell MIMO

Reference Book

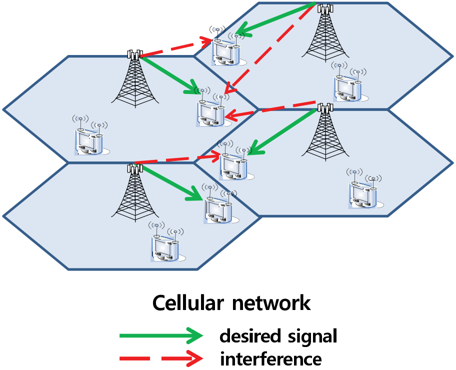
- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 13
  - Section: 13.1, 13.2, 13.3

Introduction

- Current wireless networks primarily operate using a frequency reuse 1 (or close to 1), i.e. all cells share the same frequency band
- Interference is not only made of intra-cell (i.e. multi-user interference), but also of inter-cell (i.e. multi-cell) interference.
- Cell edge performance is primarily affected by the inter-cell interference.

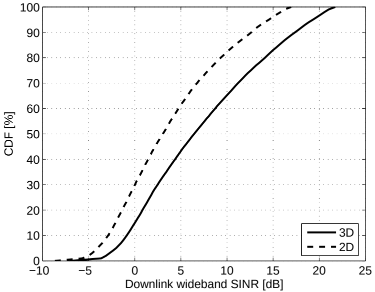


Wideband/long-term SINR

- For user q in cell i, the wideband/long-term SINR is commonly evaluated by ignoring the effect of fading but only account for path loss and shadowing

$$SINR_{w,q} = \frac{\Lambda_{q,i}^{-1} E_{s,i}}{\sigma_{n,q}^2 + \sum_{j \neq i} \Lambda_{q,j}^{-1} E_{s,j}}.$$

- Provides a rough estimate of the network performance. Function of major propagation mechanisms (path loss, shadowing, antenna radiation patterns,...), base stations deployment and user distribution.
- CDF of $SINR_{w,q}$ in a frequency reuse 1 network (cells share the same frequency band) with 2D and 3D antenna patterns in urban macro deployment.



Classical Inter-Cell Interference Mitigation

- Divide-and-conquer approach:
  - fragmenting the network area into small zones independently controlled from each other
  - making progressively use of advanced error correction coding, link adaptation, frequency selective scheduling and lately single-user and multi-user MIMO in each of those zones.

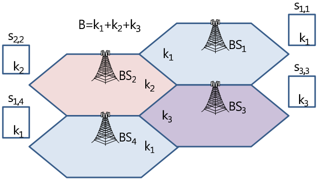


Figure: Frequency Reuse Partitioning.

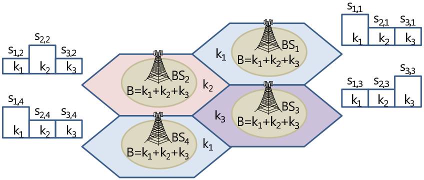


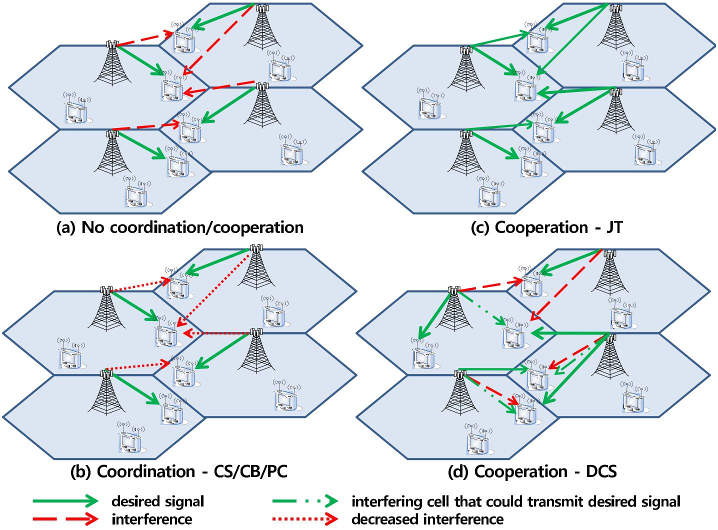
Figure: Static Fractional Frequency Reuse.

Towards Multi-Cell MIMO: Coordination and Cooperation

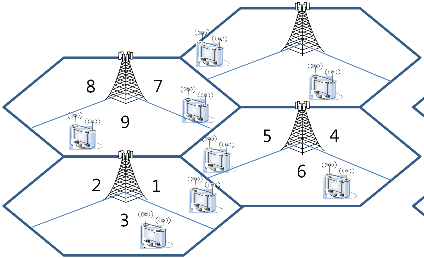
- Jointly allocate resources across the whole network (and not for each cell independently) and use the antennas of multiple cells to improve the received signal quality at the mobile terminal and to reduce the co-channel interferences.

- Two categories:
  - Coordination: No data sharing (user data is available at a single transmsitter) - CSI sharing. Modelled by an Interference Channel and Interfering Broadcast/Multiple Access Channel
  - Cooperation: Data sharing (user data is available at multiple transmsitters) - CSI sharing. Modelled by a Broadcast Channel (for Downlink) and Multiple Access Channel (for Uplink)

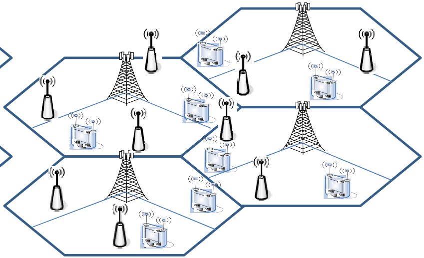
Towards Multi-Cell MIMO: Coordination and Cooperation



Network Deployments



(a) Homogeneous network



(b) Heterogeneous network

System Model - Interference Channel

- Interfering broadcast/multiple access channel
  - For each transmitter i (one per cell), the intended receivers (i.e. users) are in cell i.
  - Each receiver (i.e. user) is only interested in what is being sent by the corresponding transmitter.
  - Transmitters and receivers do not cooperate but only coordinate their transmissions by sharing CSI information. In the downlink, one transmitter does not have access to the codewords sent by other transmitters and cannot perform DPC. In the uplink, one receiver never has access to other received signals and cannot perform SIC.
- General downlink multi-cell multi-user MIMO network with a total number of K_T users distributed in n_c cells.
- K_i users in every cell i, $n_{t,i}$ transmit antennas at BS i, $n_{r,q}$ receive antennas at mobile terminal q.
- The received signal of a given user q in cell i is

$$\mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{c}_i' + \underbrace{\sum_{j \neq i} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{c}_j'}_{\text{inter-cell interference}} + \mathbf{n}_q$$

  where
  - $\mathbf{y}_q \in {}^{n_{r,q}}$,
  - $\mathbf{n}_q$ is a complex Gaussian noise $\mathcal{CN}\left(0, \sigma_{n,q}^2 \mathbf{I}_{n_{r,q}}\right)$,
  - $\Lambda_{q,i}^{-1}$ refers to the path-loss and shadowing between transmitter i and user q,
  - $\mathbf{H}_{q,i} \in {}^{n_{r,q} \times n_{t,i}}$ models the MIMO fading channel between transmitter i and user q.

Linear Precoding

- *scheduled user set* of cell i, denoted as $\mathbf{K}_i$, as the set of users who are actually scheduled by BS i at the time instant of interest

- Transmit $n_{e,i}$ streams in each cell i using MU-MIMO linear precoding

$$\mathbf{c}'_i = \mathbf{P}_i \mathbf{c}_i = \mathbf{W}_i \mathbf{S}_i^{1/2} \mathbf{c}_i = \sum_{q \in \mathbf{K}_i} \mathbf{P}_{q,i} \mathbf{c}_{q,i} = \sum_{q \in \mathbf{K}_i} \mathbf{W}_{q,i} \mathbf{S}_{q,i}^{1/2} \mathbf{c}_{q,i}$$

where

- $\mathbf{c}_i$ is the symbol vector made of $n_{e,i}$ unit-energy independent symbols
- $\mathbf{P}_i \in {}^{n_{t,i} \times n_{e,i}}$ is the precoder made of two matrices, namely a power control diagonal matrix denoted as $\mathbf{S}_i \in {}^{n_{e,i} \times n_{e,i}}$ and a transmit beamforming matrix $\mathbf{W}_i \in {}^{n_{t,i} \times n_{e,i}}$.
- $\mathbf{P}_{q,i} \in {}^{n_{t,i} \times n_{u,q}}$, $\mathbf{W}_{q,i} \in {}^{n_{t,i} \times n_{u,q}}$, $\mathbf{S}_{q,i} \in {}^{n_{u,q} \times n_{u,q}}$, and $\mathbf{c}_{q,i} \in {}^{n_{u,q}}$ are user q's sub-matrices and sub-vector of $\mathbf{P}_i$, $\mathbf{W}_i$, $\mathbf{S}_i$, and $\mathbf{c}_i$, respectively.
- The input covariance matrix at cell i is $\mathbf{Q}_i = \mathcal{E}\{\mathbf{c}'_i \mathbf{c}'^H_i\}$ subject to the transmit power constraint $\mathrm{Tr}\{\mathbf{Q}_i\} \leq E_{s,i}$.

Linear Precoding

- The received signal $\mathbf{y}_q \in {}^{n_{r,q}}$ of user $q \in \mathbf{K}_i$

$$\mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{P}_{q,i}\mathbf{c}_{q,i} + \underbrace{\sum_{p\in\mathbf{K}_i,\, p\neq q} \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{P}_{p,i}\mathbf{c}_{p,i}}_{\text{intra-cell (multi-user) interference}}$$

$$+ \underbrace{\sum_{j\neq i}\sum_{l\in\mathbf{K}_j} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{P}_{l,j}\mathbf{c}_{l,j}}_{\text{inter-cell interference}} + \mathbf{n}_q.$$

- Apply a receive combiner to stream l of user q in cell i

$$z_{q,l} = \mathbf{g}_{q,l}\mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}c_{q,i,l} + \underbrace{\sum_{m\neq l} \Lambda_{q,i}^{-1/2}\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}c_{q,i,m}}_{\text{inter-stream interference}}$$

$$+ \underbrace{\sum_{p\in\mathbf{K}_i,\, p\neq q} \Lambda_{q,i}^{-1/2}\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{P}_{p,i}\mathbf{c}_{p,i}}_{\text{intra-cell (multi-user) interference}} + \underbrace{\sum_{j\neq i}\sum_{l\in\mathbf{K}_j} \Lambda_{q,j}^{-1/2}\mathbf{g}_{q,l}\mathbf{H}_{q,j}\mathbf{P}_{l,j}\mathbf{c}_{l,j}}_{\text{inter-cell interference}} + \mathbf{g}_{q,l}\mathbf{n}_q.$$

Achievable Rate

- By treating all interference as noise, the maximum rate achievable by user q in cell i with linear precoding is

$$R_{q,i} = \sum_{l=1}^{n_{u,q}} \log_2 \left(1 + \rho_{q,l}\right).$$

- The quantity $\rho_{q,l}$ denotes the SINR experienced by stream l of user-q and writes as

$$\rho_{q,l} = \frac{S}{I_l + I_c + I_o + \left\|\mathbf{g}_{q,l}\right\|^2 \sigma_{n,q}^2}.$$

where S refers to the received signal power of the intended stream, I_l the inter-stream interference, I_c the intra-cell interference (i.e. interference from co-scheduled users) and I_o the inter-cell interference and they write as

$$\begin{aligned}
S &= \Lambda_{q,i}^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}\right|^2, \\
I_l &= \sum_{m \neq l} \Lambda_{q,i}^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}\right|^2, \\
I_c &= \sum_{p \in \mathbf{K}_i, p \neq q} \sum_{m=1}^{n_{u,p}} \Lambda_{q,i}^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{p}_{p,i,m}\right|^2, \\
I_o &= \sum_{j \neq i} \Lambda_{q,j}^{-1} \left\|\mathbf{g}_{q,l}\mathbf{H}_{q,j}\mathbf{P}_j\right\|^2.
\end{aligned}$$

Achievable Rate

Example

Given the precoders in all cells, what is the SINR of stream l of user-q in cell i?

- Noise plus interference: $I_l + I_c + I_o + \|\mathbf{g}_{q,l}\|^2 \sigma_{n,q}^2 = \mathbf{g}_{q,l}\mathbf{R}_{\mathbf{n_i}}\mathbf{g}_{q,l}^H$ where

$$\mathbf{R}_{\mathbf{n_i}} = \sum_{m \neq l} \Lambda_{q,i}^{-1}\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}\left(\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}\right)^H + \sum_{p \in \mathbf{K}_i, p \neq q} \sum_{m=1}^{n_{u,p}} \Lambda_{q,i}^{-1}\mathbf{H}_{q,i}\mathbf{p}_{p,i,m}\left(\mathbf{H}_{q,i}\mathbf{p}_{p,i,m}\right)^H + \sum_{j \neq i} \Lambda_{q,j}^{-1}\mathbf{H}_{q,j}\mathbf{P}_j\left(\mathbf{H}_{q,j}\mathbf{P}_j\right)^H + \sigma_{n,q}^2\mathbf{I}_{n_{r,q}}$$

  is the covariance matrix of the noise plus interference.

- MMSE combiner for stream l: $\mathbf{g}_{q,l} = \Lambda_{q,i}^{-1/2}\left(\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}\right)^H \mathbf{R}_{\mathbf{n_i}}{}^{-1}$
- SINR $\rho_{q,l}$ experienced by stream l of user-q

$$\rho_{q,l} = \frac{\Lambda_{q,i}^{-1}\left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}\right|^2}{\mathbf{g}_{q,l}\mathbf{R}_{\mathbf{n_i}}\mathbf{g}_{q,l}^H} = \Lambda_{q,i}^{-1}\left(\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}\right)^H \mathbf{R}_{\mathbf{n_i}}{}^{-1}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}.$$

Extension to OFDMA Networks

- *Scheduled user set* of cell i on subcarrier k, denoted as $\mathbf{K}_{k,i} \subset \mathcal{K}_i$, is the subset of users $\in \mathcal{K}_i$ who are actually scheduled on subcarrier k.
- The received signal after receive filtering of a user $q \in \mathbf{K}_{k,i}$ scheduled in cell i on subcarrier k writes as

$$\begin{aligned}\mathbf{z}_{k,q} = & \Lambda_{q,i}^{-1/2}\mathbf{G}_{k,q}\mathbf{H}_{(k),q,i}\mathbf{W}_{k,q,i}\mathbf{S}_{k,q,i}^{1/2}\mathbf{c}_{k,q,i} \\ & + \sum_{p\in\mathbf{K}_{k,i},\, p\neq q} \Lambda_{q,i}^{-1/2}\mathbf{G}_{k,q}\mathbf{H}_{(k),q,i}\mathbf{W}_{k,p,i}\mathbf{S}_{k,p,i}^{1/2}\mathbf{c}_{k,p,i} \\ & + \sum_{j\neq i}\sum_{l\in\mathbf{K}_{k,j}} \Lambda_{q,j}^{-1/2}\mathbf{G}_{k,q}\mathbf{H}_{(k),q,j}\mathbf{W}_{k,l,j}\mathbf{S}_{k,l,j}^{1/2}\mathbf{c}_{k,l,j} + \mathbf{G}_{k,q}\mathbf{n}_{k,q}.\end{aligned}$$

- The power constraint in an OFDMA network writes as

$$\sum_{k=0}^{T-1} \mathrm{Tr}\{\mathbf{S}_{k,j}\} = \sum_{k=0}^{T-1}\sum_{q\in\mathbf{K}_{k,j}}\sum_{m=1}^{n_{u,k,q}} s_{k,q,j,m} \leq E_{s,j}, \qquad \forall j.$$

System Model - Broadcast and Multiple Access Channel

- If transmitters (resp. receivers) in different cells are allowed to cooperate and can share any information through an ideal backhaul, the MIMO IC effectively becomes the MIMO BC (resp. MIMO MAC).
  - Giant MIMO BC in the downlink and MIMO MAC in the uplink.
- Focus on DL. By stacking up the transmit signal vectors, the received signal at user q is

$$\mathbf{y}_q = \mathbf{H}_q \mathbf{c}' + \mathbf{n}$$

  where

$$\mathbf{c}' = \left[\mathbf{c}_1'^T, \ldots, \mathbf{c}_{n_c}'^T\right]^T, \quad \mathbf{y} = \left[\mathbf{y}_1^T, \ldots, \mathbf{y}_{K_T}^T\right]^T, \quad \mathbf{n} = \left[\mathbf{n}_1^T, \ldots, \mathbf{n}_{K_T}^T\right]^T,$$

$$\mathbf{H}_q = \left[\begin{array}{ccc} \Lambda_{q,1}^{-1/2}\mathbf{H}_{q,1} & \ldots & \Lambda_{q,n_c}^{-1/2}\mathbf{H}_{q,n_c} \end{array}\right],$$

- The transmit signal vector in the DL multi-cell cooperation writes as

$$\mathbf{c}' = \sum_{q=1}^{K_T} \underline{\mathbf{c}}_q'$$

  where $\underline{\mathbf{c}}_q' = \left[\mathbf{c}_{q,1}'^T, \ldots, \mathbf{c}_{q,n_c}'^T\right]^T$.
  - In the MIMO IC, $\mathbf{c}_{q,j}'^T = \mathbf{0}\ \forall j \neq i$ where cell i is the serving cell of user q (i.e. $q \in \mathcal{K}_i$).
- Careful: power constraint per base station $\text{Tr}\{\mathbf{Q}_j\} \leq E_{s,j}\ \forall j$ and not sum-power constraint across base stations $\sum_{j=1}^{n_c} \text{Tr}\{\mathbf{Q}_j\} \leq \sum_{j=1}^{n_c} E_{s,j}$!

Network Architecture: Multi-Cell Measurement, Clustering and Transmission

- All interfering links do not affect equally user q's performance.
  - Dominant interfering links with small path losses/shadowing contribute to a high interference while other interfering links are almost invisible to user q.
  - Only the CSI of the dominant interfering link should actually be measured and made available to the transmitters (the CSI of other links may be ignored).

- The *MC measurement set* of user $q \in \mathcal{K}_i$ whose serving cell is i is defined as the set of cells about which channel state/statistical information related to their link to the MT is reported and is expressed based on long-term channel properties as

$$\mathcal{M}_q = \left\{ j \middle| \frac{\Lambda_{q,i}^{-1} E_{s,i}}{\Lambda_{q,j}^{-1} E_{s,j}} \leq \delta \right\}$$

  for some threshold δ and assuming maximum power transmission.

- *MC user*: a user whose MC measurement set is strictly larger than one (i.e. includes at least the MT's serving cell). The *MC users set* of cell i is defined as $\mathcal{P}_i = \{q \in \mathcal{K}_i \mid \sharp\mathcal{M}_q > 1\}$.

Multi-Cell Measurement, Clustering and Transmission

- The *MC-requested user set* of cell i is defined as the set of MC users that have cell i in their MC measurement set, i.e.

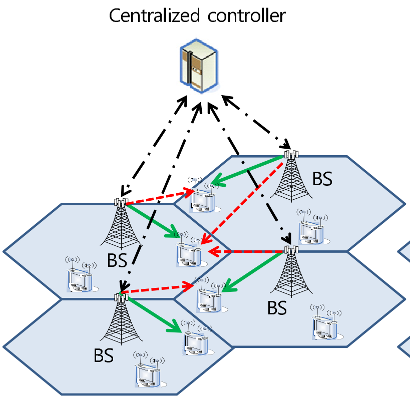
$$\mathcal{R}_i = \{ l | \, i \in \mathcal{M}_l, \sharp \mathcal{M}_l > 1 \} .$$

  Note that the MC-requested user set can also be viewed as the victim user set of cell i as it is the set of users who could be impacted by cell i in the absence of multi-cell cooperation/coordination.

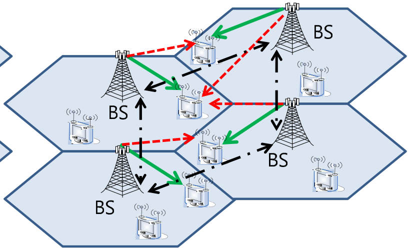
- The *MC clustering set* of user $q \in \mathcal{K}_i$ on subcarrier/time instant k is defined as the set of cells (BS) participating in the multi-cell coordination/cooperation.

- The *MC transmission set* $\mathcal{T}_{k,q}$ is a subset of the MC clustering set and is defined by the BS or set of BSs actively transmitting data to MT q.

Distributed and Centralized Architecture

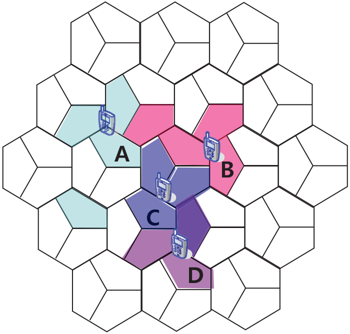


(a) Centralized architecture

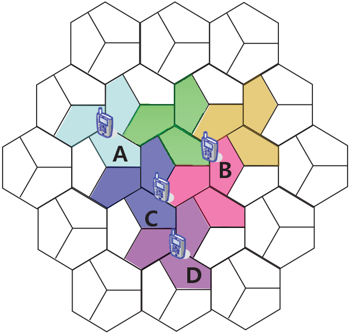


(b) Distributed architecture

User-Centric and Network-Predefined Clustering



(a) User-centric clustering



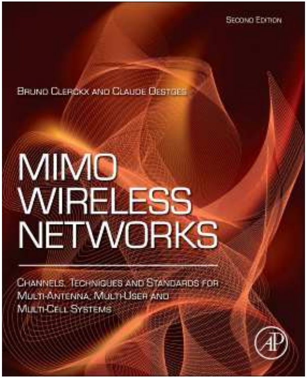
(b) Network predefined clustering

- user-centric clustering: each UE/MT has its owm clustering set. Clustering sets dynamically selected and may overlap.
- network predefined clustering: cells are statically clustered and MTs are only served by one cluster. Clusters do not overlap.

Capacity of the Interference Channel

Reference Book

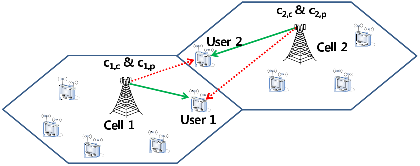
- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 13
  - Section: 13.4

SISO Interference Channel

- What is the capacity region of the two-user SISO IC?



- $\eta_{q,i} = \Lambda_{q,i}^{-1} E_{s,i} / \sigma_{n,q}^2$
  - long-term SNR when user q is served by cell i
  - long-term INR (interference to noise ratio) when q is a victim user of cell i
- $\tilde{\eta}_{q,i} = \eta_{q,i} \left| h_{q,i} \right|^2$ can be thought of as an instantaneous SNR or INR
- two-user SISO IC: transmitter 1 (i.e. cell 1) communicates with user 1 and transmitter 2 (i.e. cell 2) with user 2
  - achievable rate region function of $\tilde{\eta}_{1,1}, \tilde{\eta}_{2,2}, \tilde{\eta}_{1,2}, \tilde{\eta}_{2,1}$
  - symmetric SISO IC characterized by $\tilde{\eta}_{1,1} = \tilde{\eta}_{2,2} = \tilde{\eta}_d$ and $\tilde{\eta}_{1,2} = \tilde{\eta}_{2,1} = \tilde{\eta}_c$
  - symmetric rate: $R_{sym} = \max_{(R_1,R_2) \in \mathcal{C}_{IC}} \min \{R_1, R_2\}$ where R_1 and R_2 are the rates achievable by user 1 and 2 respectively in the two-user SISO IC and $\mathcal{C}_{IC}$ is the capacity region of the SISO IC

Very Weak Interference Regime

- Conditions: $\tilde{\eta}_{2,1} << \tilde{\eta}_{1,1}$ and $\tilde{\eta}_{1,2} << \tilde{\eta}_{2,2}$ (or simply by $\tilde{\eta}_c << \tilde{\eta}_d$ in the symmetric case)
- The interfering signal is treated as noise and encoding/decoding as in the absence of interference is sufficient
  - divide-and-conquer approach mentioned earlier (e.g., with frequency-reuse in cellular systems)

Weak Interference Regime

- Conditions: $\tilde{\eta}_{2,1} < \tilde{\eta}_{1,1}$ and $\tilde{\eta}_{1,2} < \tilde{\eta}_{2,2}$ (or simply by $\tilde{\eta}_c < \tilde{\eta}_d$ in the symmetric case)
- Capacity unknown in general but the best known achievable region has been proposed by Han-Kobayashi.
- Capacity outer-bound lies within 1 bit of the capacity inner-bound achieved by the Han-Kobayashi (HK) scheme
- Main idea behind Han-Kobayashi scheme:
  - split each transmitter information into two parts, i.e., a common and a private message.
  - A codebook shared between both transmitters is used to construct independently the common messages at each transmitter.
  - The private messages are constructed from independent codebooks.
  - Each receiver jointly decodes the common messages (and therefore partially cancel off part of the interference) by treating the private messages as interference, cancels the common messages from the received signal and then decodes the intended private message.
- Assume for simplicity a symmetric interference channel

Weak Interference Regime

- Common and private messages of user i as $c_{i,c}$ and $c_{i,p}$, $i = 1, 2$.
- A fraction x of the transmit power is allocated to the common message while the remaining fraction $1 - x$ is allocated to the private message.
- View SISO IC as formed by two SISO MACs:
  - MAC_1: 3 virtual transmitters respectively sending $c_{1,p}$, $c_{1,c}$ and $c_{2,c}$ to receiver 1, with $c_{2,p}$ treated as noise.
  - MAC_2: 3 virtual transmitters respectively sending $c_{2,p}$, $c_{1,c}$ and $c_{2,c}$ to receiver 2, with $c_{1,p}$ treated as noise.
  - Achievable rate region: intersection of the capacity regions of those two SISO MACs
- Assume for simplicity that $R_{1,c} = R_{2,c} = R_c$ and $R_{1,p} = R_{2,p} = R_p$.
- Private message

$$R_p = \log_2\left(1 + \frac{\tilde{\eta}_d\,(1-x)}{1 + \tilde{\eta}_c\,(1-x)}\right)$$

- Common message

$$R_{1,c} = R_c \leq \log_2\left(1 + \frac{\tilde{\eta}_d x}{1 + \eta_I}\right), \quad R_{2,c} = R_c \leq \log_2\left(1 + \frac{\tilde{\eta}_c x}{1 + \eta_I}\right)$$

$$R_{1,c} + R_{2,c} = 2R_c \leq \log_2\left(1 + \frac{\tilde{\eta}_d x + \tilde{\eta}_c x}{1 + \eta_I}\right)$$

  where $\eta_I = \tilde{\eta}_d\,(1-x) + \tilde{\eta}_c\,(1-x)$ is the interference from private messages.
- In the weak interference regime, $\tilde{\eta}_c < \tilde{\eta}_d$, so that the first inequality can be discarded.

Weak Interference Regime

- Hence

$$R_{sym} = R_p + R_c = \log_2\left(1 + \frac{\tilde{\eta}_d (1-x)}{1 + \tilde{\eta}_c (1-x)}\right)$$
$$+ \min\left\{\log_2\left(1 + \frac{\tilde{\eta}_c x}{1+\eta_I}\right), \frac{1}{2}\log_2\left(1 + \frac{\tilde{\eta}_d x + \tilde{\eta}_c x}{1+\eta_I}\right)\right\}$$

- x chosen so that the interference level caused by the private message has the same level as the other user's noise level.
  - interference caused by the private message has little impact on the other user's performance.
  - does not prevent each user from experiencing a relatively large private message rate as long as $\tilde{\eta}_d > \tilde{\eta}_c$.
  - $\tilde{\eta}_c (1-x) \approx 1$, i.e. $1 - x \approx 1/\tilde{\eta}_c$ and $x \approx \left(\tilde{\eta}_c - 1\right)/\tilde{\eta}_c$
- Assuming $\tilde{\eta}_d >> 1$ and $\tilde{\eta}_c >> 1$ and $\tilde{\eta}_d > \tilde{\eta}_c$

$$R_{sym} \approx \min\left\{\frac{1}{2}\log_2(\tilde{\eta}_d) + \frac{1}{2}\left[\log_2(\tilde{\eta}_d) - \log_2(\tilde{\eta}_c)\right],\right.$$
$$\left.\max\left\{\log_2(\tilde{\eta}_c), \log_2(\tilde{\eta}_d) - \log_2(\tilde{\eta}_c)\right\}\right\}.$$

Mixed Interference Regime

- Conditions: $\tilde{\eta}_{2,1} \geq \tilde{\eta}_{1,1}$ and $\tilde{\eta}_{1,2} < \tilde{\eta}_{2,2}$ or $\tilde{\eta}_{2,1} < \tilde{\eta}_{1,1}$ and $\tilde{\eta}_{1,2} \geq \tilde{\eta}_{2,2}$.
- Not meaningful for the symmetric case
- The capacity is also unknown but the best known achievable region relies on the Han-Kobayashi scheme.

Strong Interference Regime

- Conditions: $\tilde{\eta}_{2,1} \geq \tilde{\eta}_{1,1}$ and $\tilde{\eta}_{1,2} \geq \tilde{\eta}_{2,2}$ (or simply, $\tilde{\eta}_c \geq \tilde{\eta}_d$ in the symmetric case).
- The capacity region has been identified:
  - The interfering signal can be decoded along with the desired signal, i.e. each user is able to decode both messages.
  - By decoding first the interfering signal, the rate of the desired signal is improved. Unfortunately, the decodability of the interfering signal puts a constraint on the other users' rates, therefore resulting in a tradeoff between the interfering signal rate and the desired signal rate.
  - The two-user SISO IC capacity region is expressed as the intersection of the capacity regions of the two SISO MAC formed by the two transmitters and each receiver $q = 1, 2$

$$R_i \leq \log_2\left(1 + \tilde{\eta}_{q,i}\right), i = 1, 2$$
$$R_1 + R_2 \leq \log_2\left(1 + \tilde{\eta}_{q,1} + \tilde{\eta}_{q,2}\right).$$

- Given the strong interference regime $\tilde{\eta}_{2,1} > \tilde{\eta}_{1,1}$ and $\tilde{\eta}_{1,2} > \tilde{\eta}_{2,2}$, the intersection simply writes

Proposition

The capacity region $\mathcal{C}_{IC}$ of the Gaussian two-user SISO IC with strong interference is the set of all achievable rate pair (R_1, R_2) such that

$$R_i \leq \log_2\left(1 + \tilde{\eta}_{i,i}\right), i = 1, 2$$
$$R_1 + R_2 \leq \min\left\{\log_2\left(1 + \tilde{\eta}_{1,1} + \tilde{\eta}_{1,2}\right), \log_2\left(1 + \tilde{\eta}_{2,2} + \tilde{\eta}_{2,1}\right)\right\}.$$

Strong Interference Regime

- Symmetric

Corollary

The capacity region $\mathcal{C}_{IC}$ of the symmetric Gaussian two-user SISO IC with strong interference is the set of all achievable rate pair (R_1, R_2) such that

$$R_i \leq \log_2(1 + \tilde{\eta}_d), i = 1, 2$$
$$R_1 + R_2 \leq \log_2(1 + \tilde{\eta}_d + \tilde{\eta}_c).$$

- In the strong interference regime, the capacity region of the two-user SISO IC is a pentagon (as in two-user SISO MAC).
- The symmetric rate (it is actually the symmetric capacity) simply writes as

$$\begin{aligned} R_{sym} &= \frac{1}{2}\log_2(1 + \tilde{\eta}_d + \tilde{\eta}_c), \\ &\approx \frac{1}{2}\max\{\log_2(\tilde{\eta}_d), \log_2(\tilde{\eta}_c)\}, \\ &\approx \frac{1}{2}\log_2(\tilde{\eta}_c). \end{aligned}$$

Very Strong Interference Regime

- Can the capacity region, under some interference conditions, become a square only determined by the inequalities $R_i \leq \log_2(1 + \tilde{\eta}_{i,i}), i = 1, 2$?
  - i.e. each transmitter can communicate with its receiver at a rate equal to the one achievable without any interference
- Possible whenever

$$\log_2(1 + \tilde{\eta}_{1,1}) + \log_2(1 + \tilde{\eta}_{2,2}) \leq \min\{\log_2(1 + \tilde{\eta}_{1,1} + \tilde{\eta}_{1,2}), \log_2(1 + \tilde{\eta}_{2,2} + \tilde{\eta}_{2,1})\}.$$

- Very strong interference regime conditions:
  - If $\tilde{\eta}_{1,1} + \tilde{\eta}_{1,2} \leq \tilde{\eta}_{2,2} + \tilde{\eta}_{2,1}$, $\tilde{\eta}_{1,2} \geq \tilde{\eta}_{2,2} + \tilde{\eta}_{1,1}\tilde{\eta}_{2,2}$
  - If $\tilde{\eta}_{2,2} + \tilde{\eta}_{2,1} \leq \tilde{\eta}_{1,1} + \tilde{\eta}_{1,2}$, $\tilde{\eta}_{2,1} \geq \tilde{\eta}_{1,1} + \tilde{\eta}_{1,1}\tilde{\eta}_{2,2}$
- The interference is so strong that each user performs SIC by decoding the interfering message first and subtracting it from the received signal before decoding its own message.
- Each transmitter can communicate with its receiver at a rate $R_i = \log_2\left(1 + \tilde{\eta}_{i,i}\right)$ for $i = 1, 2$, as in the absence of any interference.
- The symmetric rate (symmetric capacity) simply writes as

$$R_{sym} = \log_2(1 + \tilde{\eta}_d).$$

Very Strong Interference Regime

- The very strong interference conditions can be viewed from another angle that is reminiscent of the SIC behavior in SISO BC.
- When user 1 decodes user 2's signal in the very strong interference regime, it treats its own signal as noise. Hence, for user 1 to be able to cancel correctly user 2's signal, the interfering channel between transmitter 2 and user 1 has to be strong enough to support R_2, i.e.

$$R_2 \leq \log_2\left(1+\frac{\Lambda_{1,2}^{-1}\left|h_{1,2}\right|^2 E_{s,2}}{\sigma_{n,1}^2+\Lambda_{1,1}^{-1}\left|h_{1,1}\right|^2 E_{s,1}}\right)=\log_2\left(1+\frac{\tilde{\eta}_{1,2}}{1+\tilde{\eta}_{1,1}}\right).$$

Given that user 2 wants to receive its message at a rate $R_2=\log_2\left(1+\tilde{\eta}_{2,2}\right)$, this puts the constraints

$$\log_2\left(1+\tilde{\eta}_{2,2}\right) \leq \log_2\left(1+\frac{\tilde{\eta}_{1,2}}{1+\tilde{\eta}_{1,1}}\right),$$

which equivalently writes as $\tilde{\eta}_{1,2} \geq \tilde{\eta}_{2,2}+\tilde{\eta}_{1,1}\tilde{\eta}_{2,2}$. The other condition is obtained similarly by looking at user 2's requirement to decode user 1's message correctly.

Degrees of Freedom - Multiplexing Gain

Definition

The number of generalized degrees of freedom (or multiplexing gain) is defined as

$$g_s(\alpha) = \lim_{\tilde{\eta}_d, \tilde{\eta}_c \to \infty : \frac{\log_2(\tilde{\eta}_c)}{\log_2(\tilde{\eta}_d)} = \alpha} \frac{R_{sym}(\tilde{\eta}_d, \tilde{\eta}_c)}{\log_2(\tilde{\eta}_d)}.$$

Proposition

The achievable number of generalized degrees of freedom (i.e. per-user multiplexing gain) of the two-user Gaussian SISO IC is given by

$$g_s(\alpha) = \begin{cases} 1-\alpha & 0 \leq \alpha < \frac{1}{2} \\ \alpha & \frac{1}{2} \leq \alpha < \frac{2}{3} \\ 1-\frac{\alpha}{2} & \frac{2}{3} \leq \alpha < 1 \\ \frac{\alpha}{2} & 1 \leq \alpha < 2 \\ 1 & 2 \leq \alpha. \end{cases}$$

where $\frac{\log_2(\tilde{\eta}_c)}{\log_2(\tilde{\eta}_d)} = \alpha$.

Degrees of Freedom - Multiplexing Gain

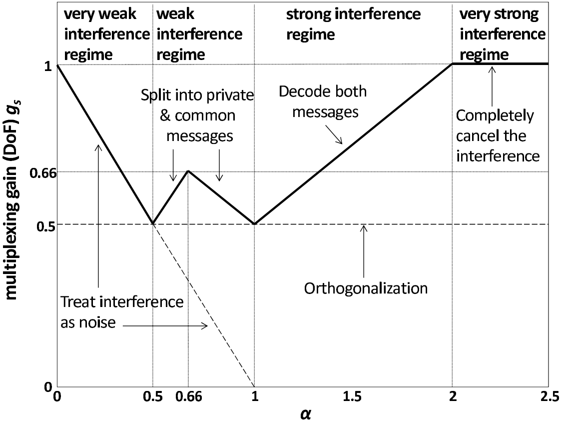


Figure: Achievable multiplexing gain per user of the two-user Gaussian SISO IC ($\alpha =$ INR/SNR).

More than Two-User SISO Interference Channels

- The extension to more than two users is far from being clear.
  - In the very weak interference regime, the optimality of treating the interference as noise has been established for scenarios with more than two users.
  - In the strong interference regime, the extension of the two-user strategy to more than two users is not straightforward in general.
  - In the weak and mixed interference regimes, the situation is even less clear.

- In the n_c-user case (or n_c-cell case), is the degree of freedom be of the order of $1/n_c$?
- Fortunately not! In a n_c-user interference channel where the intended and interfering signals are of comparable strength (i.e. medium interference regime), it is possible with Interference Alignment to achieve a multiplexing gain per user of $1/2$ despite the presence of n_c interfering users!
  - As the transmit power of each base station increases, every user will be able to simultaneously achieve half of the capacity he could achieve in the absence of the interference from other users.

More than Two-User SISO Interference Channels

- In a general n_c-user SISO IC, it is challenging to characterize the achievable multiplexing gain as a function of the SNR and INR of all links.
- The multiplexing gain is therefore commonly evaluated by taking the transmit powers to inifinity, leading to infinite SNR and INR, but without constraining the ratio between SNR and INR.

Definition

The achievable multiplexing gain of user i is defined as

$$\lim_{E_s \longrightarrow \infty} \frac{R_i}{\log_2(\eta_i)} = g_{s,i}$$

where $\eta_i = E_s \Lambda_{i,i}^{-1} / \sigma_{n,i}^2$. The total achievable multiplexing gain at the network level is defined as

$$\lim_{E_s \longrightarrow \infty} \sum_{j=1}^{n_c} \frac{R_j}{\log_2(\eta_j)} = g_{s,sum}.$$

- Meaningful at asymptotically high SNR and INR, not necessarily at finite SNR and INR!

More than Two-User SISO Interference Channels

- Time/frequency-varying interference networks are not fundamentally interference-limited

Proposition

In the n_c-user time/frequency-varying SISO IC with an infinite number of symbol extensions, the total multiplexing gain $g_{s,sum}$ (or number of degrees of freedom) is $n_c/2$.

  - time-varying: channel coefficients vary from one channel use to the another
  - beamforming over multiple symbol extensions of the time-varying channel.
  - every user must be able to partition its observed signal space into two subspaces of equal size: 1) one for the desired signals, 2) one for the waste basket for all the interference terms, under the constraint that the vector spaces corresponding to the interference must exactly align at every user receiver within the waste basket.

MIMO Interference Channels

- Use multiple antennas over static channels rather than time-varying channels
- Two-user MIMO IC

Proposition

The two-user MIMO IC with $n_{t,1}$, $n_{t,2}$ antennas at the two transmitters and $n_{r,1}$, $n_{r,2}$ antennas at their respective receivers has a maximum multiplexing gain

$$g_{s,sum} = \min\left\{n_{t,1} + n_{t,2}, n_{r,1} + n_{r,2}, \max\left\{n_{t,1}, n_{r,2}\right\}, \max\left\{n_{t,2}, n_{r,1}\right\}\right\}.$$

  - If $n_{t,1} = n_{t,2} = n_{r,1} = n_{r,2} = n$, $g_{s,sum} = n$.
  - The way antennas are distributed at both ends significantly impacts the multiplexing gain of the MIMO interference channel. A $(n_{t,1}, n_{t,2}, n_{r,1}, n_{r,2}) = (1, n-1, n-1, 1)$ MIMO IC with a total of n transmit antennas and n receive antennas would only achieve a maximum multiplexing gain of 1. Distributed processing at both ends severely limits the multiplexing gain!

- In a three-user ($n_c = 3$) MIMO IC with $n > 1$ antennas at each transmitter and each receiver and static (constant) channels, $g_{s,sum} = 3n/2$.
- Extension to more general settings but but achievable multiplexing gains commonly known only for specific antenna configurations.

Capacity of Multiple Access and Broadcast Channels

- Two-user MIMO BC and MAC

Proposition

The two-user MIMO BC with n_t transmit antennas and $n_{r,1}$, $n_{r,2}$ receive antennas has a total multiplexing gain

$$g_{s,sum} = \min\left\{n_{r,1} + n_{r,2}, n_t\right\}.$$

The two-user MIMO MAC with n_r receive antennas and $n_{t,1}$, $n_{t,2}$ transmit antennas at the two transmitters has a total multiplexing gain

$$g_{s,sum} = \min\left\{n_{t,1} + n_{t,2}, n_r\right\}.$$

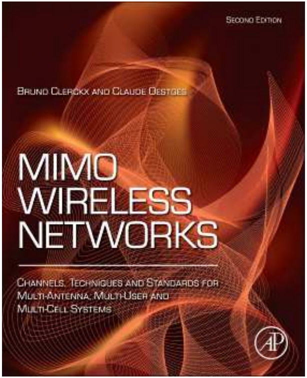
  - If $n_{t,1} = n_{t,2} = n_{r,1} = n_{r,2} = n$, $g_{s,sum} = 2n$ with BC and MAC. Twice as much as MIMO IC!
  - The way antennas are distributed at both ends does not affect the multiplexing gain in MIMO BC and MIMO MAC, contrary to the MIMO IC!

- Note: In multi-cell cooperation, no sum power constraint anymore but per BS power constraint

Coordinated Scheduling and Power Control

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 13
  - Section: 13.5, 13.6

Multi-Cell Multi-User Diversity

- Benefit of a large number of users per cell in a multi-cell network?
  - Reminiscent of the single-cell multi-user diversity.
- Assume SISO with a single user scheduled on any given spectral resource slot.
- The SINR of user q in cell i in a SISO IC simply writes as

$$\rho_q = \frac{\Lambda_{q,i}^{-1} \left|h_{q,i}\right|^2 s_i}{\sum_{j \neq i} \Lambda_{q,j}^{-1} \left|h_{q,j}\right|^2 s_j + \sigma_{n,q}^2}$$

  with s_i and s_j the transmit powers.
- Upper- and lower-bounds $\rho_{q,lb} \overset{(a)}{\leq} \rho_q \overset{(b)}{\leq} \rho_{q,ub}$

$$\rho_{q,lb} = \frac{\Lambda_{q,i}^{-1} \left|h_{q,i}\right|^2 E_s}{\sum_{j \neq i} \Lambda_{q,j}^{-1} \left|h_{q,j}\right|^2 E_s + \sigma_{n,q}^2}, \quad \rho_{q,ub} = \max_{s_i} \frac{\Lambda_{q,i}^{-1} \left|h_{q,i}\right|^2 s_i}{\sigma_{n,q}^2} = \eta_q \left|h_{q,i}\right|^2.$$

- With a rate maximization policy, the scheduler in cell i picks up the user whose SINR is the largest.

$$C_{n,lb} \leq C_n = \max_{\mathbf{S},\mathbf{K}} \sum_{i=1}^{n_c} R_{q,i} \leq C_{n,ub}$$

  where

$$C_{n,lb} = \sum_{i=1}^{n_c} \log_2\left(1 + \max_{q \in \mathcal{K}_i} \rho_{q,lb}\right), \quad C_{n,ub} = \sum_{i=1}^{n_c} \log_2\left(1 + \max_{q \in \mathcal{K}_i} \rho_{q,ub}\right).$$

Multi-Cell Multi-User Diversity

- In the *symmetric configuration*, the users within a cell i are assumed to be located at the same distance from the base station i, i.e. $\Lambda_{q\in\mathcal{K}_i,i} = \Lambda\ \forall i$, such that they experience the same average SNR η.

Proposition

In a symmetric network configuration with the fading $h_{q,i}$ independent and identically Rayleigh distributed across users, for a fixed number of cells n_c and an asymptotically large number of users per cell $K_i = K$, $\forall i$, the upper and lower bounds on the average SINR in cell i and on the network capacity scale as

$$\bar{\rho}_{ub} = \mathcal{E}\left\{\max_{q\in\mathcal{K}_i} \rho_{q,ub}\right\} \overset{K\nearrow}{\sim} \eta \log K, \qquad \bar{C}_{n,ub} \overset{K\nearrow}{\sim} n_c \log\log K,$$

$$\bar{\rho}_{lb} = \mathcal{E}\left\{\max_{q\in\mathcal{K}_i} \rho_{q,lb}\right\} \overset{K\nearrow}{\sim} \eta \log K, \qquad \bar{C}_{n,lb} \overset{K\nearrow}{\sim} n_c \log\log K.$$

- Same scaling law for ub and lb: degradation created by inter-cell interference becomes negligible when the number of users is large!
- With a rate maximization scheduler, the network is not interference limited as long as K is large enough.

Multi-Cell Multi-User Diversity

- In the *asymmetric configuration*, the users are uniformly distributed in each cell such that the path-loss is determined by the distance between the user and its serving cell. $\Lambda_{q,i}^{-1}$ is a random variable i.i.d. across users and cells.

Proposition

In an asymmetric network configuration with the fading $h_{q,i}$ independent and identically Rayleigh distributed across users, for a fixed number of cells n_c and an asymptotically large number of users per cell $K_i = K$, $\forall i$, the upper and lower bounds on the SINR in cell i and on the network capacity scale as

$$\bar{\rho}_{ub} = \mathcal{E}\left\{\max_{q\in\mathcal{K}_i} \rho_{q,ub}\right\} \overset{K\nearrow}{\sim} \kappa_{ub} K^{\frac{\epsilon}{2}}, \quad \bar{C}_{n,ub} \overset{K\nearrow}{\sim} n_c \frac{\epsilon}{2} \log K,$$
$$\bar{\rho}_{lb} = \mathcal{E}\left\{\max_{q\in\mathcal{K}_i} \rho_{q,lb}\right\} \overset{K\nearrow}{\sim} \kappa_{lb} K^{\frac{\epsilon}{2}}, \quad \bar{C}_{n,lb} \overset{K\nearrow}{\sim} n_c \frac{\epsilon}{2} \log K$$

where κ_{ub} and κ_{lb} are scaling factors.

- Same scaling law for ub and lb but the presence of unequal path-losses among users enhances the multi-user diversity, therefore resulting in a larger growth rate in the asymmetric case $(\log K)$ compared to the symmetric case $(\log\log K)$.

Multi-Cell Multi-User Diversity

- Similar behavior was already observed in single-cell MU-MIMO where the intra-cell interference is shown to have a negligible impact on the capacity when random beamforming and opportunistic scheduling are performed under the assumption that the number of users in the cell is large enough.
- Possible to achieve the optimal network capacity under a totally distributed network architecture that does not require any exchange of CSI and coordination among cells. Indeed, each cell can perform single-cell scheduling simply relying on the report of a CQI $\rho_{q,lb}$.

Multi-Cell Resource Allocation

- Maximize a network utility metric rather than a cell utility metric
- Narrowband transmission

$$\{\mathbf{S}^\star, \mathbf{W}^\star, \mathbf{G}^\star, \mathbf{K}^\star\} = \arg \max_{\mathbf{S},\mathbf{W},\mathbf{G},\mathbf{K}\subset\mathcal{K}} \sum_{i=1}^{n_c} \sum_{q\in\mathbf{K}_i} w_q R_{q,i}$$

  - Weights w_q account for rate maximization or network-wide proportional fairness

- Multi-carrier (MIMO-OFDMA) transmission

$$\{\mathbf{S}^\star, \mathbf{W}^\star, \mathbf{G}^\star, \mathbf{K}^\star\} = \arg \max_{\mathbf{S},\mathbf{W},\mathbf{G},\mathbf{K}\subset\mathcal{K}} \frac{1}{T} \sum_{i=1}^{n_c} \sum_{k=0}^{T-1} \sum_{q\in\mathbf{K}_{k,i}} w_q R_{(k),q,i}$$

Coordinated Power Control

- Assume SISO narrowband transmissions, i.e. $n_t = n_r = 1$
- What are the benefits of performing power control (as well as joint power control and user scheduling) to mitigate inter-cell interference ?

$$\{\mathbf{S}^\star, \mathbf{K}^\star\} = \arg \max_{\mathbf{S}, \mathbf{K} \subset \mathcal{K}} \sum_{i=1, q \in \mathbf{K}_i}^{n_c} w_q R_{q,i}.$$

- Same scaling laws for lower- and upper-bounds on the network capacity (a)symmetric configurations for large K and with a rate maximization approach.
  - $\rightarrow$ The use of power control, even though optimal, does not further increase the network capacity, and transmitting at full power is optimal in the asymptotic case of a large number of users.

Proposition

Assuming a fading channel $h_{q,i}$ to be independent and identically Rayleigh distributed across users, for a fixed number of cells n_c and an asymptotically large number of users per cell $K_i = K$, $\forall i$, the network capacity with optimal power control and rate-maximization based scheduling scales as $\bar{C}_n \overset{K \nearrow}{\sim} n_c \log \log K$ in a symmetric network configuration, and as $\bar{C}_n \overset{K \nearrow}{\sim} n_c \frac{\epsilon}{2} \log K$ in an asymmetric network configuration.

High and Low SINR Regimes

- Imagine that the users to be scheduled $\mathbf{K}$ have been selected, and focus on optimal power allocation.

$$\{\mathbf{S}^\star\} = \arg\max_{\mathbf{S}\in\mathcal{S}} \sum_{i=1, q\in\mathbf{K}_i}^{n_c} w_q R_{q,i}$$

where $\mathcal{S} = \{\mathbf{S} \mid e_{s,i} \leq s_i \leq E_{s,i}, i = 1, \ldots, n_c\}$ is the feasible set of power allocation strategies.

Proposition

In the high and low SINR regimes, the optimal power control $S^\star$, is binary, i.e. $\mathbf{S}^\star \in \mathcal{S}_b^{n_c}$, where $\mathcal{S}_b^{n_c}$ is the set of $2n_c - 1$ corner points of $\mathcal{S}$, excluding the all-$e_{s,i}$ point ($i = 1, \ldots, n_c$).

- The original problem can be converted into the following user scheduling and power control exhaustive search problem

$$\{\mathbf{S}^\star, \mathbf{K}^\star\} = \arg\max_{\mathbf{S}\subset\mathcal{S}_b^{n_c}, \mathbf{K}\subset\mathcal{K}} \sum_{i=1, q\in\mathbf{K}_i}^{n_c} w_q R_{q,i}.$$

An exhaustive search is conducted over the sets $\mathcal{S}_b^{n_c}$ and $\mathcal{K}$ to find the optimal $\mathbf{S}^\star$ and $\mathbf{K}^\star$.

Two-Cell Clusters

- Interestingly, in the specific two-cell network $(n_c = 2)$, binary power control is not only optimal at low SINR but in the whole SINR range for a network relying on a rate maximization policy

Proposition

In the two-cell case, the network sum-rate maximizing power allocation $\left(s_1^\star, s_2^\star\right)$ *is binary and always takes one of the following three power allocation candidates:* $(E_{s,1}, e_{s,2})$, $(e_{s,1}, E_{s,2})$ *and* $(E_{s,1}, E_{s,2})$.

- Implications:
  - The transmit power can be quantized to two values without loss of capacity.
  - This makes the power allocation strategy particularly simple and of very low overhead.
  - The decision cannot be taken based on local CSI only as the optimal decision requires simultaneous CSI from both cells, therefore requiring some form of centralized scheduler.
  - Popular in LTE-A in the name of on-off power control or coordinated silencing. Robust to CSI measurement and CSI feedback impairments once implemented at the subframe level in OFDMA networks
- The binary allocation is not optimal anymore for a network-wide proportional criterion for which the weights w_1 and w_2 are different, and for a more general set-up containing more than two cells.

OFDMA Networks

- Multi-cell coordinated OFDMA networks consists in deriving a joint scheduling and power allocation scheme that decides, on each subcarrier, upon the transmit power level and the user to be scheduled in each cell.
  - one more dimension since: multiple users can be allocated different frequency resources.
  - The rate of a given user would typically be improved by increasing its bandwidth allocation or transmit power:
    - the former leading to a bandwidth allocation loss for other users in the cell,
    - the latter leading to an increase of the inter-cell interference (analogous to the narrowband system).
  - Any improvement of the rate of one user affects the rate of the other users in the network

- Assume SISO

- User assignments for all sub-carriers and all cells: $\mathbf{K} = \{\mathbf{K}_i\}_{i=1}^{n_c}$ where $\mathbf{K}_i = \{\mathbf{K}_{k,i}\}_{\forall k}$.

- Power allocation: $\mathbf{S} = \{\mathbf{S}_i\}_{i=1}^{n_c}$ where $\mathbf{S}_i = \{s_{k,i}\}_{\forall k}$

Objective function

- Coordinated power control and scheduling in OFDMA Networks

$$\{\mathbf{S}^\star, \mathbf{K}^\star\} = \arg \max_{\mathbf{S}, \mathbf{K} \subset \mathcal{K}} \frac{1}{T} \sum_{i=1}^{n_c} \sum_{k=0, q \in \mathbf{K}_{k,i}}^{T-1} w_q R_{(k),q,i}$$

where

$$R_{(k),q,i} = \log_2 (1 + \rho_{k,q})$$

and

$$\rho_{k,q} = \frac{\Lambda_{q,i}^{-1} \left| h_{(k),q,i} \right|^2 s_{k,i}}{\sum_{j \neq i} \Lambda_{q,j}^{-1} \left| h_{(k),q,j} \right|^2 s_{k,j} + \sigma_{n,k,q}^2},$$

under the constraint

$$\sum_{k=0}^{T-1} s_{k,i} \leq E_{s,i}, \ \forall i.$$

- Non-convex problem. The globally optimal solution may not be found but near-optimal solutions can be obtained using iterative algorithms. The key idea relies on an iterative optimization of scheduling and power allocation as discussed below.

Optimality Conditions

- For a predefined set of scheduled users, the optimal power allocation problem must satisfy the Karush-Kuhn-Tucker (KKT) conditions
- Lagrangian of the optimization problem

$$\mathcal{L}\left(\mathbf{S},\mathbf{K},\nu\right)=\sum_{i=1}^{n_c}\sum_{k=0,q\in\mathbf{K}_{k,i}}^{T-1} w_q R_{(k),q,i}+\sum_{i=1}^{n_c}\nu_i\left(E_{s,i}-\sum_{k=0}^{T-1}s_{k,i}\right)$$

  where $\nu=\{\nu_i\}_{i=1}^{n_c}$ is the set of Lagrange multipliers associated with the power constraint in each cell.

- The solution should satisfy

$$\frac{\partial\mathcal{L}}{\partial s_{k,i}}=0$$

  and

$$\nu_i\left(E_{s,i}-\sum_{k=0}^{T-1}s_{k,i}\right)=0,$$

  under the constraints $\nu_i\geq 0$, $s_{k,i}\geq 0$ and $\sum_{k=0}^{T-1}s_{k,i}\leq E_{s,i}$, for $i=1,\ldots,n_c$ and $k=0,\ldots,T-1$.

Optimality Conditions

- $\frac{\partial \mathcal{L}}{\partial s_{k,i}} = 0$ leads to

$$w_q \frac{\partial R_{(k),q,i}}{\partial s_{k,i}} + \sum_{m \neq i} w_{q',m} \frac{\partial R_{(k),q',m}}{\partial s_{k,i}} = \nu_i,$$

with $q \in \mathbf{K}_{k,i}$ and $q' \in \mathbf{K}_{k,m}$. Equivalently we can write

$$w_q \frac{\partial R_{(k),q,i}}{\partial s_{k,i}} - \Pi_{k,i} = \nu_i$$

where we define $\Pi_{k,i} = \sum_{m \neq i} \Pi_{k,i,m}$ with

$$\Pi_{k,i,m} = -w_{q',m} \frac{\partial R_{(k),q',m}}{\partial s_{k,i}} = -w_{q',m} \underbrace{\frac{\partial R_{(k),q',m}}{\partial I_{k,q',m}}}_{-\pi_{k,q',m}} \frac{\partial I_{k,q',m}}{\partial s_{k,i}}$$

$$\stackrel{(a)}{=} w_{q',m} \pi_{k,q',m} \Lambda_{q',i}^{-1} \left| h_{(k),q',i} \right|^2$$

where $I_{k,q',m} = \sum_{l \neq m} \Lambda_{q',l}^{-1} \left| h_{(k),q',l} \right|^2 s_{k,l}$ is the total interference received by user q' in cell m and $\pi_{k,q',m}$ is defined as the non-negative quantity that represents the marginal increase in rate of user q' in cell m per unit decrease in total interference on subcarrier k.

Optimality Conditions

- Power allocation

$$\frac{1}{\log 2}\frac{w_q}{(\nu_i+\Pi_{k,i})}=s_{k,i}+\frac{\sum_{j\neq i}\Lambda_{q,j}^{-1}\left|h_{(k),q,j}\right|^2 s_{k,j}+\sigma_{n,k,q}^2}{\Lambda_{q,i}^{-1}\left|h_{(k),q,i}\right|^2}$$

for $i=1,\ldots,n_c$ and $k=0,\ldots,T-1$, where $\Pi_{k,i}=\sum_{m\neq i}\Pi_{k,i,m}$ and

$$\pi_{k,q',m}=\frac{1}{\log 2}\frac{1}{\Lambda_{q',m}^{-1}\left|h_{(k),q',m}\right|^2 s_{k,m}}\left(\frac{\rho_{k,q'}^2}{1+\rho_{k,q'}}\right).$$

Interference Pricing

- View $\pi_{k,q',m}$ as a *price* charged to other cells for generating interference to user q'
- $$w_q \frac{\partial R_{(k),q,i}}{\partial s_{k,i}} - \Pi_{k,i} = \nu_i$$

  is a necessary and sufficient optimality condition for the problem in which each cell i specifies a power level $s_{k,i}$ on subcarrier k to maximize the following surplus function

  $$\Upsilon_{(k),i} = w_q R_{(k),q,i} - s_{k,i}\Pi_{k,i},$$

  assuming fixed $s_{k,j}$ with $j \neq i$ and $\pi_{k,q',m}$ with $m \neq i$.
- Rather than maximizing selfishly its own utility metric (i.e. weighted sum-rate), cell i maximizes the difference between its utility and its payment owed to the interference created to the victim users in the neighboring cells:
  - The payment is given by the transmit power $s_{k,i}$ times $\Pi_{k,i}$. $\Pi_{k,i}$ writes as a weighted sum of victim users' prices, with the weights equal to QoS weights times the channel gains between cell i and the victim users.
  - Representative of the effect of allocating additional transmit power at cell i on the weighted rate of all victim users in neighboring cells.
  - A high value of $\Pi_{k,i}$ suggests that cell i must pay a high price for assigning power on subcarrier k.

Iterative Scheduler

- In the downlink, the inter-cell interference is only function of the power levels and is independent of the user scheduling decisions.
- This suggests that the user scheduling and the power allocation can be carried out separately.
- An iterative scheduler can be derived so that the best user to schedule are first found assuming a fixed power allocation, then the best power allocation are computed for the fixed scheduled users.
- Assuming a fixed power allocation, given the independence of the inter-cell interference on the scheduled users,

$$q^{\star} \in \mathbf{K}_{k,i}^{\star} = \arg\max_{\mathcal{K}_i} \Upsilon_{(k),i} = \arg\max_{\mathcal{K}_i} w_q R_{(k),q,i}, \ \forall i, k.$$

Iterative Scheduler

- *Initialization step:* We first fix the maximum number of iteration N_{max} and fix $n = 0$. We initialize $\mathbf{S}^{(0)}$ using e.g. a binary power control strategy, and compute $\mathbf{K}^{(0)}$ and $\mathbf{\Pi}^{(0)}$.
- *Iteration-n:* For each cell $i = 1, \ldots, n_c$, we update the power allocation $\mathbf{S}_i^{(n)}$ based on $\mathbf{K}^{(n-1)}$, $\mathbf{\Pi}^{(n-1)}$ and $\mathbf{S}_j^{(n-1)}$ for $\forall j \neq i$ (i.e. assuming the transmit powers in the other cells remain fixed) as follows

$$s_{k,i}^{(n)} = \left(\frac{1}{\log 2} \frac{w_q}{\left(\nu_i + \Pi_{k,i}^{(n-1)}\right)} - \frac{\sigma_{I,k,q}^2 + \sigma_{n,k,q}^2}{\Lambda_{q,i}^{-1} \left| h_{(k),q,i} \right|^2} \right)^+ , \quad q \in \mathbf{K}_{k,i}^{(n-1)}$$

where

$$\sigma_{I,k,q}^2 = \sum_{j \neq i} \Lambda_{q,j}^{-1} \left| h_{(k),q,j} \right|^2 s_{k,j}^{(n-1)}.$$

The parameters ν_i are obtained from the power constraint $\sum_{k=0}^{T-1} s_{k,i}^{(n)} \leq E_{s,i}$. After obtaining the power allocation $\mathbf{S}^{(n)}$, the user selection $\mathbf{K}^{(n)}$ and finally $\mathbf{\Pi}^{(n)}$ are obtained. The procedure is repeated till convergence or till the number of iterations reaches N_{max}.

Modified Iterative Water-Filling

- Modified iterative water-filling

$$s_{k,i}^{(n)} = \left(\frac{1}{\log 2} \frac{w_q}{\left(\nu_i + \Pi_{k,i}^{(n-1)}\right)} - \frac{\sigma_{I,k,q}^2 + \sigma_{n,k,q}^2}{\Lambda_{q,i}^{-1} \left| h_{(k),q,i} \right|^2} \right)^+ , \quad q \in \mathbf{K}_{k,i}^{(n-1)}$$

- Similar to the point-to-point water-filling algorithm
  - The main difference lies in the fact that the power is allocated accounting for the combined noise and interference and that the water-filling level ν_i is modified by the additional pricing term $\Pi_{k,i}$.
  - As the inter-cell interference increases, the water-filling level decreases, which results in a lower power allocation.
  - The water-filling level is also affected by the proportional fairness weights w_q in such a way that the water-filling level gets higher as the weight increases.
  - As a result, BSs allocate more power on subcarriers that serve users with either high priorities or better channel qualities but transmit power is decreased on subcarriers where transmission causes excessive interference to victim users in adjacent cells.

Feedback and Message Passing Requirements

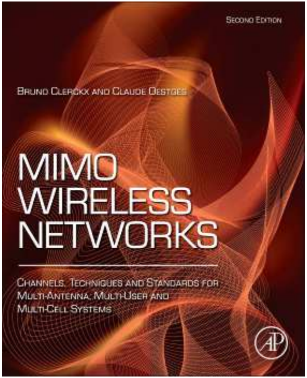
- Centralized implementation:
  - Each user q in cell i to report $n_c T$ channel measurements as the channel from user q to any base station j over all subcarriers must be known.
  - Each cell i forwards the CSI to a centralized controller.
  - The central controller performs the modified iterative water-filling and informs each cell about the scheduled user and the transmit power on each subcarrier.

- Distributed implementation:
  - Cells exchange with each other messages and rely on the feedback information from the users
  - Each cell is assumed to be aware of local CSI, i.e. CSI that can be measured by its user and reported, plus the messages exchanged between cells.
  - From cell i perspective:
    - the report to cell i from users $q \in \mathbf{K}_{k,i}^{(n-1)}$ of $\sigma_{I,k,q}^2 + \sigma_{n,k,q}^2$ and $\Lambda_{q,i}^{-1} \left| h_{(k),q,i} \right|^2$
    - the reception by cell i from each cell $m \neq i$ in the MC clustering set of user q of the tax information $\Pi_{k,i,m}^{(n-1)}$ at iteration $n-1$
    - the transfer of the tax information $\Pi_{k,j,i}^{(n-1)}$ at iteration $n-1$ to cooperating cell j.

Coordinated Beamforming and Interference Alignment

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 13
  - Section: 13.7

Coordinated Beamforming

- Rate achievable by user q in cell i with linear precoding is

$$R_{q,i} = \sum_{l=1}^{n_{u,q}} \log_2 \left(1 + \rho_{q,l}\right).$$

- The quantity $\rho_{q,l}$ denotes the SINR experienced by stream l of user-q

$$\rho_{q,l} = \frac{S}{I_l + I_c + I_o + \left\|\mathbf{g}_{q,l}\right\|^2 \sigma_{n,q}^2}.$$

where S refers to the received signal power of the intended stream, I_l the inter-stream interference, I_c the intra-cell interference (i.e. interference from co-scheduled users) and I_o the inter-cell interference

$$S = \Lambda_{q,i}^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{w}_{q,i,l}\right|^2 s_{q,i,l},$$

$$I_l = \sum_{m \neq l} \Lambda_{q,i}^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{w}_{q,i,m}\right|^2 s_{q,i,m},$$

$$I_c = \sum_{p \in \mathbf{K}_i, p \neq q} \sum_{m=1}^{n_{u,p}} \Lambda_{q,i}^{-1} \left|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{w}_{p,i,m}\right|^2 s_{p,i,m},$$

$$I_o = \sum_{j \neq i} \Lambda_{q,j}^{-1} \left\|\mathbf{g}_{q,l}\mathbf{H}_{q,j}\mathbf{W}_j\mathbf{S}_j^{1/2}\right\|_F^2.$$

Zero-Forcing Beamforming and Block Diagonalization

- Zero-Forcing Beamforming (ZFBF) based coordination is a natural extension of MU-MIMO precoding based on ZFBF or BD.
- Forcing the interference to zero at either the input or the output of the receiver.
- If we want to zero-force at the input of the receiver, the constraints on the ZFBF transmit filters, targeting user $q \in \mathbf{K}_i$, follow from (**??**) as

$$\Lambda_{s,i}^{-1/2}\mathbf{H}_{s,i}\mathbf{W}_{q,i} = \mathbf{0}, \forall s \neq q, s \in \mathbf{K}_i,$$
$$\Lambda_{l,i}^{-1/2}\mathbf{H}_{l,i}\mathbf{W}_{q,i} = \mathbf{0}, \forall l \in \mathbf{K}_j \cap \mathcal{R}_i.$$

- Denoting the set of user indices

$$\tilde{\mathbf{K}}_{q,i} = \{\mathbf{K}_i, \mathbf{K}_j \cap \mathcal{R}_i\}_{\forall j \neq i} \setminus q$$

whose size is $\tilde{K}_{q,i} = \sharp\tilde{\mathbf{K}}_{q,i}$, we define the interference space $\tilde{\mathbf{H}}_{q,i} \in {}^{n_r\tilde{K}_{q,i}\times n_t}$ as

$$\tilde{\mathbf{H}}_{q,i} = \left[\begin{array}{ccc} \dots & \Lambda_{s,i}^{-1/2}\mathbf{H}_{s,i}^T & \dots \end{array} \right]^T_{s \in \tilde{\mathbf{K}}_{q,i}}.$$

The zero forcing constraint forces $\mathbf{W}_{q,i}$ to lie in the null space of $\tilde{\mathbf{H}}_{q,i}$.
- Serving cell channel (between the user and the serving BS) but also the interfering cells channels (between the user and the interfering BSs in the MC measurement set) need to be known at the transmitter.

Interference Alignment

- Extends the coordinated zero-forcing beamforming (or block diagonalization) to jointly design the transmit precoders and the receive combiners in every cell and for every user.
- IA restricts the interference at every receiver input to a subset of the received signal space (i.e. the interference is aligned in that subset) and arranges the desired signal in the complementary subset such that it can be perceived as interference-free at the receiver output.
- Such alignment is receiver specific in the sense that some signals may appear aligned in a given space at receivers where they constitute interference while they remain distinguishable at other receivers where they are desired.
- Focus on maximizing the degrees of freedom in the network.
- An underlying assumption is that the SNR/INR are high enough.
- Assume a predefined set of scheduled users and a single user transmission, where every cell schedules only a single user at a given time instant.
  - The scheduled user index q is chosen as i and $\mathbf{c}_{q,i}$ and $\mathbf{W}_{q,i}$ write as $\mathbf{c}_i$ and $\mathbf{W}_i$

Conditions for Interference Alignment

- Assumptions:
  - n_c cells fully connected with n_c users with n_t transmit antennas at each BS ($n_{t,i} = n_t$ $\forall i$) and $n_r \leq n_t$ ($n_{r,i} = n_r$ $\forall i$) receive antennas at each MT.
  - Each MT receives $n_e < n_r$ data streams from its serving BS.

- Divide the n_r-dimensional observation space at the receiver into a n_e-dimensional signal space and a $n_r - n_e$-dimensional interference space and design jointly the transmit and receive filters such that every interference is aligned into the $n_r - n_e$-dimensional interference space.

- Interference alignment possible if

$$
\begin{aligned}
\mathcal{C}\left(\mathbf{H}_{1,2}\mathbf{W}_2\right) &= \mathcal{C}\left(\mathbf{H}_{1,3}\mathbf{W}_3\right) = \cdots = \mathcal{C}\left(\mathbf{H}_{1,n_c}\mathbf{W}_{n_c}\right), \\
\mathcal{C}\left(\mathbf{H}_{2,1}\mathbf{W}_1\right) &= \mathcal{C}\left(\mathbf{H}_{2,3}\mathbf{W}_3\right) = \cdots = \mathcal{C}\left(\mathbf{H}_{2,n_c}\mathbf{W}_{n_c}\right), \\
&\vdots \\
\mathcal{C}\left(\mathbf{H}_{n_c,1}\mathbf{W}_1\right) &= \mathcal{C}\left(\mathbf{H}_{n_c,2}\mathbf{W}_2\right) = \cdots = \mathcal{C}\left(\mathbf{H}_{n_c,n_c-1}\mathbf{W}_{n_c-1}\right),
\end{aligned}
$$

where $\mathcal{C}\left(\mathbf{A}\right)$ is the column space of a matrix, i.e., the vector space spanned by the column vectors of matrix $\mathbf{A}$.

Conditions for Interference Alignment

- Given $n_e < n_r$, $\mathcal{C}\left(\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}\right) = \mathcal{C}\left(\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix}\right)$ is satisfied if $\exists\ \mathbf{G}$ such that $\mathbf{G}\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \mathbf{G}\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix} = \mathbf{0}$
- Interference is aligned before receive shaping in such a way that after receive shaping, it is completely canceled out and the received signal $\mathbf{y}_i$ of the scheduled user in cell i lies in the d-dimensional signal space.
- The transmit filter $\mathbf{W}_i$ and the receive shaping $\mathbf{G}_l$ are obtained as solutions of the set of $n_c(n_c - 1)$ equations

$$\Lambda_{l,i}^{-1/2}\mathbf{G}_l\mathbf{H}_{l,i}\mathbf{W}_i = \mathbf{0}, \qquad l \neq i, \forall i, l = 1, ..., n_c.$$

- Knowledge of global CSI required. Strictly speaking, $\mathbf{H}_{i,i}\ \forall i$ not needed.

Closed Form Solutions

- Assume
  - $n_c = 3$, $n = n_t = n_r$ with n even for simplicity.
  - The precoder $\mathbf{W}_i$ is of dimension $n \times n/2$.
  - The channel matrices are full rank.
- Show that there exist $n/2$ non-interfering paths between transmitter i and receiver i for each $i = 1, 2, 3$, i.e. a total multiplexing gain of $3n/2$.
- The interference can be zero-forced if the dimension of the interference space is $\leq n/2$
  - e.g. at receiver 1, $r\left(\begin{bmatrix} \mathbf{H}_{1,3}\mathbf{W}_3 & \mathbf{H}_{1,2}\mathbf{W}_2 \end{bmatrix}\right) = n/2$. Recall that $r\left(\mathbf{H}_{1,2}\mathbf{W}_2\right) = n/2$ and $r\left(\mathbf{H}_{1,3}\mathbf{W}_3\right) = n/2$!
- The conditions for IA

$$\begin{aligned}
\mathcal{C}\left(\mathbf{H}_{1,2}\mathbf{W}_2\right) &= \mathcal{C}\left(\mathbf{H}_{1,3}\mathbf{W}_3\right), \\
\mathcal{C}\left(\mathbf{H}_{2,1}\mathbf{W}_1\right) &= \mathcal{C}\left(\mathbf{H}_{2,3}\mathbf{W}_3\right), \\
\mathcal{C}\left(\mathbf{H}_{3,1}\mathbf{W}_1\right) &= \mathcal{C}\left(\mathbf{H}_{3,2}\mathbf{W}_2\right),
\end{aligned}$$

- Given that channel matrices are invertible,

$$\begin{aligned}
\mathcal{C}\left(\mathbf{W}_2\right) &= \mathcal{C}\left(\mathbf{H}_{3,2}^{-1}\mathbf{H}_{3,1}\mathbf{W}_1\right), \\
\mathcal{C}\left(\mathbf{W}_3\right) &= \mathcal{C}\left(\mathbf{H}_{2,3}^{-1}\mathbf{H}_{2,1}\mathbf{W}_1\right). \\
\mathcal{C}\left(\mathbf{W}_1\right) &= \mathcal{C}\left(\mathbf{T}\mathbf{W}_1\right),
\end{aligned}$$

where $\mathbf{T} = \mathbf{H}_{3,1}^{-1}\mathbf{H}_{3,2}\mathbf{H}_{1,2}^{-1}\mathbf{H}_{1,3}\mathbf{H}_{2,3}^{-1}\mathbf{H}_{2,1}$.

Closed Form Solutions

- Set $\mathbf{W}_1 = eig(\mathbf{T})$ where $eig(\mathbf{T})$ refers to $n/2$ dominant eigenvectors of $\mathbf{T}$. Hence $\mathbf{W}_1 = \begin{bmatrix} \mathbf{t}_1 & \dots & \mathbf{t}_{n/2} \end{bmatrix}$.
- Stricter conditions

$$\mathbf{H}_{2,1}\mathbf{W}_1 = \mathbf{H}_{2,3}\mathbf{W}_3,$$
$$\mathbf{H}_{3,1}\mathbf{W}_1 = \mathbf{H}_{3,2}\mathbf{W}_2,$$

  leding to $\mathbf{W}_2 = \mathbf{H}_{3,2}^{-1}\mathbf{H}_{3,1}\mathbf{W}_1$ and $\mathbf{W}_3 = \mathbf{H}_{2,3}^{-1}\mathbf{H}_{2,1}\mathbf{W}_1$.
- To zero-force interference, the desired signal must be linearly independent of the interference at the receivers.
  - e.g. at receiver 1, we need a full rank matrix $\begin{bmatrix} \mathbf{H}_{1,1}\mathbf{W}_1 & \mathbf{H}_{1,2}\mathbf{W}_2 \end{bmatrix}$.
  - Multiplying by $\mathbf{H}_{1,1}^{-1}$, columns of $\begin{bmatrix} \mathbf{t}_1 & \dots & \mathbf{t}_{n/2} & \mathbf{A}\mathbf{t}_1 & \dots & \mathbf{A}\mathbf{t}_{n/2} \end{bmatrix}$ should be linearly independent, with $\mathbf{A} = \mathbf{H}_{1,1}^{-1}\mathbf{H}_{1,2}\mathbf{H}_{3,2}^{-1}\mathbf{H}_{3,1}$,
  - Satisfied given that $\mathbf{A}$ is a random full rank linear transformation.
  - Similar observations hold true for receivers 2 and 3,
- All 3 receivers can decode $n/2$ streams using zero-forcing.

Closed Form Solutions

- At receiver 1, the eigenvalue decomposition of the interference matrix is given by

$$\left[\begin{array}{cc} \mathbf{H}_{1,2}\mathbf{W}_2 & \mathbf{H}_{1,3}\mathbf{W}_3 \end{array}\right] = \left[\begin{array}{cc} \mathbf{U}^{(1)} & \mathbf{U}^{(0)} \end{array}\right] \mathbf{\Lambda}\mathbf{V}^H.$$

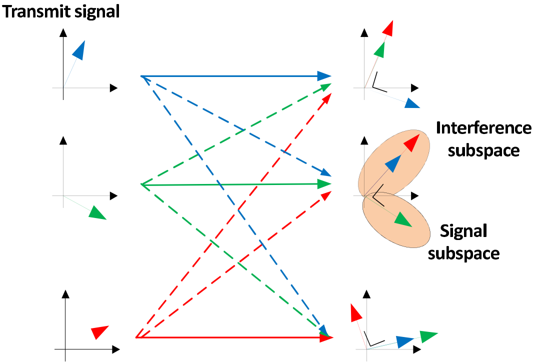
- Given the alignment of $\mathbf{H}_{1,2}\mathbf{W}_2$ and $\mathbf{H}_{1,3}\mathbf{W}_3$, $\mathbf{U}^{(0)}$ refers to the $n/2$ singular vectors corresponding to zero singular values. Hence by selecting $\mathbf{G}_1 = \left(\mathbf{U}^{(0)}\right)^H$, we obtain

$$\begin{aligned} \mathbf{y}_1 &= \Lambda_{1,1}^{-1/2}\mathbf{G}_1\mathbf{H}_{1,1}\mathbf{W}_1\mathbf{S}_1^{1/2}\mathbf{x}_1 + \sum_{j=2,3} \Lambda_{1,j}^{-1/2}\mathbf{G}_1\mathbf{H}_{1,j}\mathbf{W}_j\mathbf{S}_j^{1/2}\mathbf{x}_j + \mathbf{G}_1\mathbf{n}_1, \\ &= \Lambda_{1,1}^{-1/2}\mathbf{G}_1\mathbf{H}_{1,1}\mathbf{W}_1\mathbf{S}_1^{1/2}\mathbf{x}_1 + \mathbf{G}_1\mathbf{n}_1. \end{aligned}$$

  User 1 perceives an equivalent channel given by $\mathbf{H}_{eq,1,1} = \Lambda_{1,1}^{-1/2}\mathbf{G}_1\mathbf{H}_{1,1}\mathbf{W}_1$ and is not affected by multi-cell interference.
- The interference alignment presented so far creates an interference-free subspace but does not attempt to maximize the desired signal strength within the desired signal subspace.
  - IA solution is not a function of the direct channels $\mathbf{H}_{i,i}$
  - sub-optimal at low and medium SNR
  - Enhancement: precoding in a second stage (IA being the first stage) along the eigenvectors of $\mathbf{H}_{eq,1,1}$ and applying water-filling based on its singular values.

IA Illustration

- Illustration of interference alignment on a three-user interference channel



Iterative Solution

- Closed form solutions for IA have been found for specific settings only.
- In general, with $n_c > 3$, $n_t \neq n_r$ and $n_{e,i}$ streams for transmitter i, analytical solutions to the IA problem are difficult to obtain.
- Iterative solution:
  - Assuming the transmit precoders $\mathbf{W}_i^{(n-1)}$ at iteration $n-1$, the receiver shaping $\mathbf{G}_i^{(n)}$ are first computed at iteration n.
  - The updated transmit precoders $\mathbf{W}_i^{(n)}$ can then be computed based on all $\mathbf{G}_i^{(n)}$.
  - The process iterates until convergence.
- The iterative algorithm alternates between the original and reciprocal networks.
  - The reciprocal network consists in switching the roles of transmitter and receiver.
  - We denote a variable in the reciprocal network with a bar on top.
  - In the reciprocal network, the channel matrix writes as $\bar{\mathbf{H}}_{j,i} = \mathbf{H}_{i,j}^H$.
  - IA conditions in the reciprocal network write as $\bar{\mathbf{G}}_j \bar{\mathbf{H}}_{j,i} \bar{\mathbf{W}}_i = 0$, $\forall j \neq i$.
  - $\bar{\mathbf{W}}_i = \mathbf{G}_i^H$ and $\bar{\mathbf{G}}_i = \mathbf{W}_i^H$.

Iterative Solution

- Within each network only the receive filters are updated to minimize the total inter-cell leakage interference.
  - In the original network, the total inter-cell interference leakage at receiver i due to all interfering transmitters is given by

$$I_{o,i} = \mathsf{Tr}\left\{\mathbf{G}_i\mathbf{Q}_i\mathbf{G}_i^H\right\}$$

    with

$$\mathbf{Q}_i = \sum_{j \neq i} \Lambda_{i,j}^{-1}\mathbf{H}_{i,j}\mathbf{W}_j\mathbf{S}_j\mathbf{W}_j^H\mathbf{H}_{i,j}^H.$$

- Design the receive shaping $\mathbf{G}_i$ such that it lies in the space spanned by the $n_{e,i}$ eigenvectors corresponding to the $n_{e,i}$ smallest eigenvalues of $\mathbf{Q}_i$. Writing $\mathbf{Q}_i = \mathbf{U}_{\mathbf{Q}_i}\mathbf{\Lambda}_{\mathbf{Q}_i}\mathbf{U}_{\mathbf{Q}_i}^H$ with the entries of $\mathbf{\Lambda}_{\mathbf{Q}_i}$ ranked by increasing order of magnitude, $\mathbf{G}_i = \mathbf{U}_{\mathbf{Q}_i}^H(:, 1:n_{e,i})$.
- At every iteration, the computation of the receive filter $\mathbf{G}_i$ is performed in the original network while the computation of the transmit filter $\mathbf{W}_i$ (or equivalently the receive shaping $\bar{\mathbf{G}}_i$) is computed in the reciprocal network.

Iterative Solution

- *Initialization step:* Start with arbitrary precoding matrices $\mathbf{W}_i^{(0)}$ with $\left(\mathbf{W}_i^{(0)}\right)^H \mathbf{W}_i^{(0)} = \mathbf{I}_{n_{e,i}}$.
- *Iteration-n:* Alternate between the original and the reciprocal networks:
  1. In the original network, compute the interference covariance matrix at each receiver i

$$\mathbf{Q}_i^{(n)} = \sum_{j \neq i} \Lambda_{i,j}^{-1} \mathbf{H}_{i,j} \mathbf{W}_j^{(n-1)} \mathbf{S}_j \left(\mathbf{W}_j^{(n-1)}\right)^H \mathbf{H}_{i,j}^H,$$

  and fix the receive shaping in the original network and the transmit beamformer in the reciprocal network respectively as

$$\mathbf{G}_i^{(n)} = \mathbf{U}_{\mathbf{Q}_i^{(n)}}^H \left(:, 1 : n_{e,i}\right), \ \bar{\mathbf{W}}_i^{(n)} = \left(\mathbf{G}_i^{(n)}\right)^H.$$

  2. In the reciprocal network, compute the interference covariance matrix at each receiver j,

$$\bar{\mathbf{Q}}_j^{(n)} = \sum_{i \neq j} \bar{\Lambda}_{j,i}^{-1} \bar{\mathbf{H}}_{j,i} \bar{\mathbf{W}}_i^{(n)} \bar{\mathbf{S}}_i \left(\bar{\mathbf{W}}_i^{(n)}\right)^H \bar{\mathbf{H}}_{j,i}^H,$$

  and fix the receive shaping in the reciprocal network and the transmit beamformer in the original network respectively as

$$\bar{\mathbf{G}}_j^{(n)} = \mathbf{U}_{\bar{\mathbf{Q}}_j^{(n)}}^H \left(:, 1 : n_{e,i}\right), \ \mathbf{W}_j^{(n)} = \left(\bar{\mathbf{G}}_j^{(n)}\right)^H.$$

For simplicity, uniform power allocation ($\mathbf{S}_j = \bar{\mathbf{S}}_j = E_{s,j}/n_{e,j} \mathbf{I}_{n_{e,j}}$) is often assumed.

Iterative Solution

- $n_{r,i} = n_{t,i} = n \ \forall i$, $n_{e,i} = n_e \ \forall i$, uniform power allocation

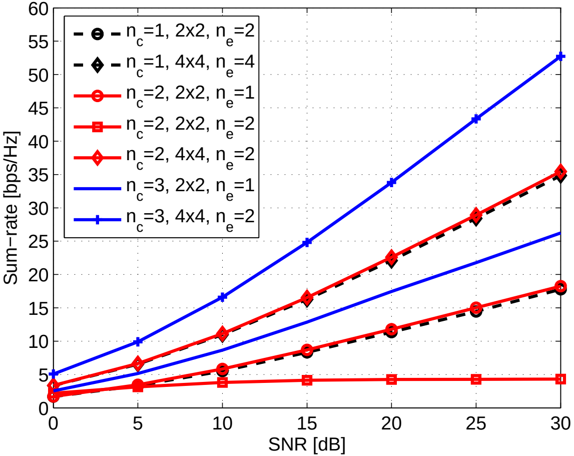


Figure: Sum-rate of IA vs SNR in various configurations (n_c, $n_r \times n_t$, n_e) in i.i.d. Rayleigh fading channels.

MIMO Interfering Broadcast/Multiple Access Channels

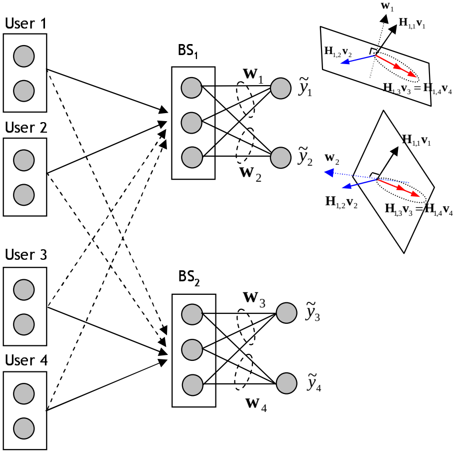
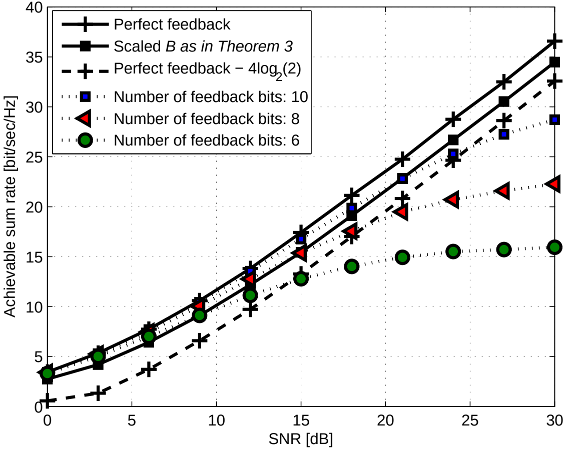


Figure: The system model of the two-cell interfering MIMO-MAC when $n_t = 2$, $n_r = 3$, and two cells with two users in each cell.

CSI Feedback and Message Exchange

- IA relies heavily on accurate CSI knowledge!
- Recall the sensitivity of MU-MIMO to inaccurate CSIT.
- MIMO Interfering MAC with quantized feedforward



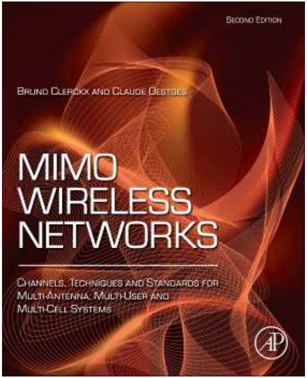
Other Beamformers

- Matched Beamforming
  - important when it comes to Massive MIMO
- Joint Leakage Suppression
- Maximum Network Sum-Rate Beamforming
- Beamforming with Assigned Target SINR
- Balancing Competition and Coordination
- Opportunistic Beamforming

Coordinated Scheduling, Beamforming and Power Control

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



- Chapter 13
  - Section: 13.8

Coordinated Scheduling, Beamforming and Power Control

- Coordinated scheduler, beamformer and power control in MIMO-OFDMA

$$\{\mathbf{S}^\star, \mathbf{W}^\star, \mathbf{K}^\star\} = \arg \max_{\mathbf{S},\mathbf{W},\mathbf{K}\subset\mathcal{K}} \frac{1}{T} \sum_{i=1}^{n_c} \sum_{k=0}^{T-1} \sum_{q\in\mathbf{K}_{k,i}} w_q R_{(k),q,i}$$

where

$$R_{(k),q,i} = \log_2\left(1 + \rho_{k,q}\right)$$

with

$$\rho_{k,q} = \frac{\Lambda_{q,i}^{-1} \left|\mathbf{h}_{(k),q,i}\mathbf{w}_{k,q,i}\right|^2 s_{k,q,i}}{\sum_{j=1}^{n_c} \sum_{\substack{u\in\mathbf{K}_{k,j} \\ (u,j)\neq(q,i)}} \Lambda_{q,j}^{-1} \left|\mathbf{h}_{(k),q,j}\mathbf{w}_{k,u,j}\right|^2 s_{k,u,j} + \sigma_{n,k,q}^2},$$

under the constraint

$$\sum_{k=0}^{T-1} \sum_{q\in\mathbf{K}_{k,i}} s_{k,q,i} \leq E_{s,i},\ \forall i.$$

Optimality Conditions

- Assuming a fixed user schedule and transmit beamformers

$$\mathcal{L}\left(\mathbf{S},\mathbf{W},\mathbf{K},\nu\right)=\sum_{i=1}^{n_c}\sum_{k=0}^{T-1}\sum_{q\in\mathbf{K}_{k,i}} w_q R_{(k),q,i}+\sum_{i=1}^{n_c}\nu_i\left(E_{s,i}-\sum_{k=0}^{T-1}\sum_{q\in\mathbf{K}_{k,i}} s_{k,q,i}\right)$$

  where $\nu=\{\nu_i\}_{i=1}^{n_c}$ is the set of Lagrange multipliers associated with the power constraint in each cell.

- The solution should satisfy

$$\frac{\partial\mathcal{L}}{\partial s_{k,q,i}}=0$$

  and

$$\nu_i\left(E_{s,i}-\sum_{k=0}^{T-1}\sum_{q\in\mathbf{K}_{k,i}} s_{k,q,i}\right)=0,$$

  under the constraints $\nu_i\geq 0$, $s_{k,q,i}\geq 0$ and $\sum_{k=0}^{T-1}\sum_{q\in\mathbf{K}_{k,i}} s_{k,q,i}\leq E_{s,i}$, for $i=1,\ldots,n_c$ and $k=0,\ldots,T-1$.

Iterative Scheduler

- Power allocation follows the modified iterative water-filling (accounting for beamforming)

$$s_{k,q,i} = \left(\frac{1}{\ln 2} \frac{w_q}{\nu_i + \Pi_{k,q,i}} - \frac{\sigma_{I,k,q}^2 + \sigma_{n,k,q}^2}{\Lambda_{q,i}^{-1} \left| \mathbf{h}_{(k),q,i} \mathbf{w}_{k,q,i} \right|^2} \right)^+$$

where

$$\sigma_{I,k,q}^2 = \sum_{j=1}^{n_c} \sum_{\substack{u \in \mathbf{K}_{k,j} \\ (u,j) \neq (q,i)}} \Lambda_{q,j}^{-1} \left| \mathbf{h}_{(k),q,j} \mathbf{w}_{k,u,j} \right|^2 s_{k,u,j}.$$

- Interference pricing interpretation: each cell i attempts to maximize on subcarrier k the following surplus function for every single user q which this cell i aims to schedule

$$\Upsilon_{(k),q,i} = w_q R_{(k),q,i} - s_{k,q,i} \Pi_{k,q,i},$$

assuming fixed $s_{k,u,j}$ and $\Pi_{k,u,j}$ $\forall (u,j) \neq (q,i)$.

  - pricing mechanism that accounts for the impact of beamforming and power allocation on the interference created to co-scheduled users and adjacent cells.
  - any variation of the transmit beamformer of a given user in a cell alters the interference created to other users in the network.

Iterative Scheduler

- Scheduler:
  - For a fixed user schedule and transmit power per beam, optimize the beamforming vectors.
  - For fixed beamforming vectors and power allocation, the user scheduling is done per beam by finding the user that maximizes $w_q R_{(k),q,i}$.
  - For a fixed beamformers and user schedule, the power levels are updated.
- Recall that SINR of each user needs to be accurately computed by the BS at every iteration!
  - Inaccurate SINR prediction hampers the appropriate selection of the users, the transmission ranks and the beamformers at every iteration of scheduler and ultimately the whole link adaptation and the convergence of the scheduler

A General Framework of Coordination

- Previous iterative scheduler motivates the design of a general framework of coordination as used in CoMP
  - CSI exchange between cells
  - Dynamically and iteratively take decisions on the users to schedule, on the appropriate subcarriers, on their corresponding beamformers and on the power levels to maximize a network utility metric

- *Initialization step:* Each cell decides upon which users to schedule in SU or MU-MIMO mode and the corresponding transmit precoders and power levels assuming no coordination between cells. In cell i,

$$\left\{\mathbf{S}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{G}_i^{(0)}, \mathbf{K}_i^{(0)}\right\} = \arg\max \mathcal{U}_i^{(0)}$$

with

$$\mathcal{U}_i^{(0)} = \frac{1}{T} \sum_{k=0}^{T-1} \sum_{q \in \mathbf{K}_{k,i}^{(0)}} w_q R_{(k),q,i}\left(\mathbf{P}_{k,q,i}^{(0)}\right)$$

A General Framework of Coordination

- *Iteration-n:* Each cell revisits its decision regarding the users to be scheduled and their transmit precoders, based on the decisions taken by other cells in iteration $n-1$.
  - The scheduling decisions in a given cell i are not only function of the utility metric of users scheduled by that cell but also of the utility metric of victim users that have been tentatively scheduled by other cells in iteration $n-1$.
  - Cell i allocates resources such that

$$\left\{\mathbf{S}_i^{(n)}, \mathbf{W}_i^{(n)}, \mathbf{G}_i^{(n)}, \mathbf{K}_i^{(n)}\right\} = \arg \max_{\substack{\mathbf{K}_i \in \mathcal{K}_i \\ \mathbf{P}_{k,i}(\mathcal{K}_i, \mathcal{R}_i^{(n-1)})}} \mathcal{U}_i^{(n)}(\mathcal{K}_i, \mathcal{R}_i^{(n-1)}).$$

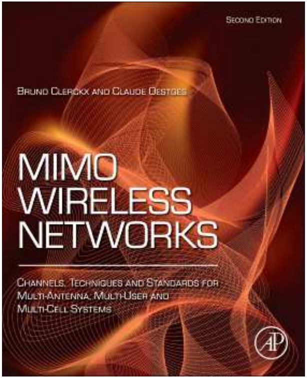
    - $(\mathcal{K}_i, \mathcal{R}_i^{(n-1)})$: function of its served user set and its victim user set at iteration $n-1$!
  - Utility metric of cell i at iteration n

$$\mathcal{U}_i^{(n)} = \underbrace{\frac{1}{T}\sum_{k=0}^{T-1}\sum_{q\in\mathbf{K}_{k,i}^{(n)}} w_q R_{(k),q,i}\left(\mathbf{P}_{k,q,i}^{(n)}, \mathbf{P}_{k,j\in\mathcal{M}_q^{(n-1)}}^{(n-1)}\right)}_{\text{Single-cell weighted sum-rate}} - \underbrace{\Pi_i\left(\mathcal{R}_i^{(n-1)}\right)}_{\text{Tax to be paid due to the interference created to vicitm users in adjacent cells}}.$$

Massive MIMO

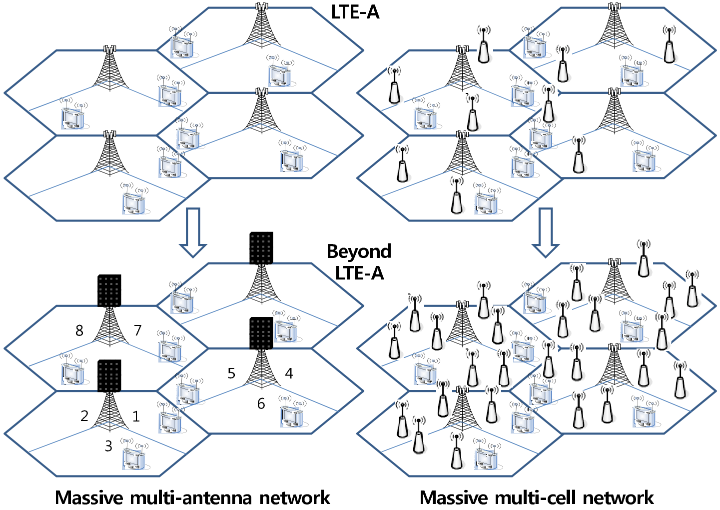
Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



  - Chapter 5,12,13
    - Section: 5.4, 12.2, 12.6, 12.8, 13.7

Introduction



Point-to-Point i.i.d. Channels

Proposition

In i.i.d. Rayleigh fading channels, the ergodic capacity with CDIT is achieved under an equal power allocation scheme $\mathbf{Q} = \mathbf{I}_{n_t}/n_t$*, i.e.,*

$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = \mathcal{E}\left\{\log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}_w\mathbf{H}_w^H\right]\right\},$$

or equivalently,

$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = \mathcal{E}\left\{\sum_{k=1}^{n}\log_2\left[1 + \frac{\rho}{n_t}\lambda_k\right]\right\},$$

where $n = r(\mathbf{H}_w)$ *is the rank of* $\mathbf{H}_w$ *and* $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ *are the non-zero eigenvalues of* $\mathbf{H}_w\mathbf{H}_w^H$.

Point-to-Point i.i.d. Channels

Proposition

The ergodic capacity of i.i.d. Rayleigh fast fading channels with CDIT is given by

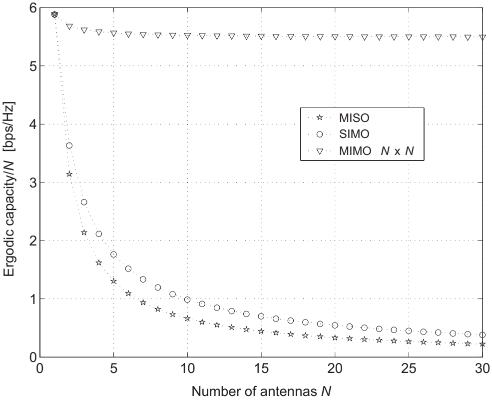
$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = n \int_0^\infty \log_2(1 + \rho\lambda/n_t) p_\lambda(\lambda) d\lambda,$$

where $p_\lambda(\lambda)$ is the distribution of a randomly selected (non-ordered) eigenvalue of $\mathbf{T}_w$ ($\mathbf{T}_w = \mathbf{H}_w \mathbf{H}_w^H$ for $n_t > n_r$ and $\mathbf{T}_w = \mathbf{H}_w^H \mathbf{H}_w$ for $n_r > n_t$).

Point-to-Point i.i.d. Channels

- Three particular cases of MIMO systems:
  1. $n_t = N$ and $n_r = 1$ (MISO)
  $$\bar{C}_{CDIT} = e^{N/\rho}\log_2(e)\sum_{p=1}^{N} E_p\left(\frac{N}{\rho}\right),$$
  2. $n_t = 1$ and $n_r = N$ (SIMO)
  $$\bar{C}_{CDIT} = e^{1/\rho}\log_2(e)\sum_{p=1}^{N} E_p\left(\frac{1}{\rho}\right),$$
  3. $n_t = n_r = n = N$
  $$\begin{aligned}\bar{C}_{CDIT} &= e^{N/\rho}\log_2(e)\sum_{k=0}^{N-1}\sum_{l=0}^{k}\sum_{m=0}^{2l}\left\{\frac{(-1)^m}{2^{2k-m}}\binom{2l}{l}\right.\\ &\quad\left.\times\binom{2k-2l}{k-l}\binom{2l}{m}\sum_{p=1}^{m+1} E_p\left(\frac{N}{\rho}\right)\right\}\\ &\approx e^{1/\rho}\log_2(e)E_1\left(\frac{1}{\rho}\right)\\ &\quad +(N-1)\left\{2\log_2\left(1+\sqrt{4\rho+1}\right)-\frac{\log_2(e)}{4\rho}\left(\sqrt{4\rho+1}-1\right)^2-2\right\}.\end{aligned}$$

Point-to-Point i.i.d. Channels



- Observations: Asymptotically,
  - $\bar{C}_{CDIT}$ scales linearly with N for squared MIMO systems at any SNR,
  - For MISO and SIMO systems, $\bar{C}_{CDIT}/N$ decreases toward zero with increasing N (and the decrease is faster for MISO than for SIMO).
  - If the number of antennas is increased only at one side, the asymptotic rate of the capacity growth is equal to zero.

Large Antenna Array Regime in Point-to-Point i.i.d. Channels

- For $n_t = n_r = N$, we may take the limit for $N \to \infty$ of $\bar{C}_{CDIT}/N$, and obtain

$$\lim_{N\to\infty} \frac{\bar{C}_{CDIT}}{N} = 2\log_2\left(1+\sqrt{4\rho+1}\right) - \frac{\log_2(e)}{4\rho}\left(\sqrt{4\rho+1}-1\right)^2 - 2$$

- Observations:
  - $\lim_{N\to\infty} \bar{C}_{CDIT}/N$ is a constant, which only depends on the SNR.
  - The capacity scales with $N = n$ at any SNR in the large antenna array regime!
- More general scenario of $N > n$? Consider
  - $N = n_r \to \infty$, with $n = n_t$ fixed,
  - $N = n_t \to \infty$, with $n = n_r$ fixed,
  - $N = n_t \to \infty$, $n = n_r \to \infty$, in a constant ratio $N/n > 1$.

Large Antenna Array Regime in Point-to-Point i.i.d. Channels

- $N = n_r \to \infty$, with $n = n_t$ fixed:
  - $\mathbf{W}/N = \mathbf{H}^H\mathbf{H}/N$ converges to $\mathbf{I}_n$ as $N \to \infty$.
  - This implies that, for a fixed value of n, the n eigenvalues of $\mathbf{W}/N$ approach one, i.e. the empirical distribution $p_{\lambda'}(\lambda')$ (where $\lambda' \triangleq \lambda/N$) approaches $\delta(\lambda' - 1)$.
  - Hence,

$$\lim_{N=n_r\to\infty} \frac{\bar{C}_{CDIT}}{n} = \log_2\left(1 + \rho\frac{N}{n}\right).$$

- $N = n_t \to \infty$, with $n = n_r$ fixed:
  - The empirical distribution of the eigenvalues of $\mathbf{W}/N = \mathbf{H}\mathbf{H}^H/N$ also converges almost surely to $\delta(\lambda' - 1)$,
  - Hence

$$\lim_{N=n_t\to\infty} \frac{\bar{C}_{CDIT}}{n} = \log_2(1+\rho),$$

  which is equal to the capacity of a SISO AWGN channel!

- $N = n_t \to \infty$, $n = n_r \to \infty$, in a constant ratio $N/n > 1$:
  - $p_{\lambda'}(\lambda')$ (where $\lambda' \triangleq \lambda/N$) can be computed
  - Hence

$$\lim_{n\to\infty} \frac{\bar{C}_{CDIT}}{n} = \log_2\left(1 + \rho + \rho\frac{n}{N} - \rho\beta\right) + \left(1 - \frac{N}{n}\right)\log_2(1-\beta) - \log_2(e)\frac{N}{n}\beta$$

  with β a function of $\frac{N}{n}$ and ρ.

Large Antenna Array Regime in Point-to-Point i.i.d. Channels

- The capacity in the large antenna regime scales linearly with n at any SNR
  - Recall that in the non-asymptotic case, the linear increase in n is observed only at high SNR!
  - The growth rate is only function of the SNR and the ratio N/n.
- The convergence is very fast, so that the large antenna array regime is reached already for values of n as small as 3.
- C_{CDIT} is Gaussian distributed, with a mean value given by in previous slide, and a variance decreasing as $1/N$ in the first two scenarios. In the third scenario, the variance dependence towards ρ and N/n is more complex.
- Under the large antenna array regime, the channel becomes much more deterministic and the channel matrices better conditioned (as opposed to random),
  - see e.g. the distribution of the eigenvalues of $\mathbf{W}/N$ in scenarios 1 and 2.

Large Antenna Array Regime in Multi-User Channels

- Recall

Proposition

For channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$, SNR η_q, number of receive antennas n_r, the gain of DPC over TDMA is upper-bounded by the minimum between the number of transmit antennas n_t and the number of users K

$$\frac{C_{BC}(\mathbf{H})}{C_{TDMA}(\mathbf{H})} \leq \min\{n_t, K\}.$$

- In the limit of large n_t and with a fixed K and n_r , the DPC gain over TDMA equals K:

Proposition

For any fixed n_r, K and effective SNR η_q $\forall q$, the gain of DPC over TDMA for a large number of transmit antennas ($n_t \rightarrow \infty$) in fading channels that are independent and identically Rayleigh distributed across antennas and independent across users writes as

$$\frac{\bar{C}_{BC}}{\bar{C}_{TDMA}} \overset{n_t \nearrow}{\sim} K.$$

Intuition

- Fading channels are independent and identically Rayleigh distributed across antennas and independent across users.
- Hence, by the law of large numbers, the $n_r K$ rows of $\mathbf{H}$ become mutually orthogonal as n_t becomes large,

$$\lim_{n_t \to \infty} \frac{1}{n_t} \mathbf{H}_l \mathbf{H}_p^H = \mathbf{I}_{n_r} \delta_{lp}, \qquad \forall l, p = 1, \ldots, K,$$

- By transmitting to the best user, TDMA exploits only n_r orthogonal dimensions while DPC can use up to $n_r K$ dimensions, i.e. K times as many signaling dimensions as TDMA.
- For all SNR! Yes, this factor K is translated into a factor K increase in rate in the limit of a large n_t because the received SNR linearly increases with n_t and effectively reaches the high SNR regime for large n_t.

Sum-Rate Capacity of Massive MISO BC

- Using the BC-MAC duality, sum-rate capacity of the MISO BC

$$C_{BC}(\mathbf{H}, E_s) = \max_{\{s_{ul,q}\}} \log_2 \det\left[\mathbf{I}_{n_t} + \sum_{q=1}^{K} \mathbf{h}_q^H s_{ul,q} \mathbf{h}_q\right],$$
$$= \max_{\{s_{ul,q}\}} \log_2 \det\left[\mathbf{I}_{n_t} + \mathbf{H}^H \mathbf{S}_d \mathbf{H}\right],$$

  where $\mathbf{S}_d = \text{diag}\{s_{ul,1}, \ldots, s_{ul,K}\}$ and the maximization is performed over $s_{ul,q} \geq 0$ $\forall q$ with $\sum_{q=1}^{K} s_{ul,q} \leq E_s$.
- Assuming a large number of transmit antennas n_t, decorrelation leads to $1/n_t \mathbf{H}\mathbf{H}^H \approx \mathbf{\Lambda}_d$ where $\mathbf{\Lambda}_d = \text{diag}\{\Lambda_1^{-1}/\sigma_{n,1}^2, \ldots, \Lambda_K^{-1}/\sigma_{n,K}^2\}$.
- C_{BC} with a large number of transmit antennas approximates as

$$\bar{C}_{BC} \approx C_{BC}(\mathbf{H}, E_s) \approx \max_{\{s_{ul,q}\}} \log_2 \det\left[\mathbf{I}_K + n_t \mathbf{\Lambda}_d \mathbf{S}_d\right],$$
$$= \max_{\{s_{ul,q}\}} \sum_{q=1}^{K} \log_2\left(1 + n_t \Lambda_q^{-1}/\sigma_q^2 s_{ul,q}\right).$$

- Assuming $\Lambda_q^{-1}/\sigma_q^2$ is the same for all users, the optimal power allocation boils down to the uniform power allocation $s_{ul,q} = E_s/K$ $\forall q$.

Sum-Rate of Linear-Precoded Massive MISO BC

- At high and low SNR, linear beamforming (BF) techniques (based on ZFBF) can achieve the same scaling rate as DPC.

Proposition

At high SNR and low SNR, DPC and BF have the same scaling rate

$$C_{BC}(\mathbf{H}) \overset{E_s \nearrow}{\sim} C_{BF}(\mathbf{H})$$

$$C_{BC}(\mathbf{H}) \overset{E_s \searrow}{\sim} C_{BF}(\mathbf{H})$$

Sum-Rate of Linear-Precoded Massive MISO BC

- In the limit large n_t, thanks to the decorrelation, matched beamforming achieves, similarly to DPC or ZFBF, a factor K increase in rate compared to TDMA.
  - matched beamformer $\mathbf{w}_s = \bar{\mathbf{h}}_s^H = \mathbf{h}_s^H / \|\mathbf{h}_s\| \; \forall s = 1, \ldots, K$
  - the SINR ρ_q of user q

$$\rho_q = \frac{\Lambda_q^{-1} \left|\mathbf{h}_q \mathbf{w}_q\right|^2 s_q}{\sum_{\substack{p \in \mathbf{K} \\ p \neq q}} \Lambda_q^{-1} \left|\mathbf{h}_q \mathbf{w}_p\right|^2 s_p + \sigma_{n,q}^2} \overset{n_t \nearrow}{\approx} \frac{\Lambda_q^{-1} n_t s_q}{\sigma_{n,q}^2}$$

  - achievable sum-rate

$$\bar{C}_{BF} \approx C_{BF} \approx \sum_{q=1}^{K} \log_2\left(1 + n_t \Lambda_q^{-1} / \sigma_{n,q}^2 s_q\right).$$

  - Assuming $\Lambda_q^{-1}/\sigma_q^2$ is the same for all users, the uniform power allocation $s_q = E_s/K$ $\forall q$ maximizes $\bar{C}_{BF}$.
  - Strong similarity with the sum-rate capacity: Matched beamforming achieves the sum-rate capacity in the very large transmit antenna regime if s_q are chosen equal to $s_{ul,q}$.
  - Rate approximations are only valid in the large antenna regime for $K/n_t \to 0$. If $K, n_t \to \infty$ with the ratio $n_t/K = \alpha$, C_{BF} with matched beamforming exhibits an error floor in the limit of large SNR.

Massive SIMO MAC

- Similar observations also hold in the multiple access channels (i.e. in the uplink) with a large number of receive antennas. Assuming a SIMO MAC (with single antenna transmitters) with large n_r,

$$\bar{C}_{MAC} \approx C_{MAC} \approx \sum_{q=1}^{K} \log_2 \left(1 + n_r \eta_q\right).$$

- Under the large receive antenna regime, this sum-rate is achievable with a simple receive matched filter.

Large Antenna Array Regime in Multi-User Channels

- The transmit (and also receive) beamforming gain approximates as n_t for large n_t. As n_t increases, the value of $\|\mathbf{h}_q\|^2$, being a $\chi^2_{2n_t}$ distributed random variable, concentrates indeed more and more around its mean.
- The SINR and the sum-rates become exclusively a function of Λ_q and not of the fading (that is so useful to benefit from MU diversity)!
- For large n_t, $\bar{C}_{BF}/\bar{C}_{TDMA} = K$!

Proposition

For any fixed n_r, K and effective SNR η_q $\forall q$, the gain of DPC/BF over TDMA for a large number of transmit antennas ($n_t \to \infty$) in fading channels that are independent and identically Rayleigh distributed across antennas and independent across users writes as

$$\frac{\bar{C}_{BC/BF}}{\bar{C}_{TDMA}} \overset{n_t \nearrow}{\sim} K.$$

  - For large n_t, $\bar{C}_{TDMA} \approx \log_2\left(1 + n_t \max_{q=1,\ldots,K}\{\eta_q\}\right)$ and $\bar{C}_{BF} \approx \sum_{q=1}^{K} \log_2\left(1 + \frac{n_t \eta_q}{K}\right)$ (assuming uniform power allocation).
- Transmit/Receive beamforming and MU diversity are somehow not complementary. A large n_t benefits the array gain and multiplexing gain but restricts the MU diversity gain.

Channel Hardening and Scheduling

- This behavior is called channel hardening and originates from the fact that, with transmit and/or receive beamforming, the variance of rate σ_R^2 decreases with n_t and n_r as n_t and n_r tend to infinity, respectively.
- In general, any transmission scheme that exploits spatial diversity reduces the multi-user diversity gain because of the channel hardening effect.
- Scheduling in Massive MIMO becomes much simpler!

Linear Precoding: Matched Beamforming

- As n_t increases, with perfect CSIT, the matched beamformer of user q becomes orthogonal to co-scheduled users ($s \neq q$) channels. Hence, the multi-user interference is naturally eliminated.
- Massive MIMO is spectrally and energy efficient:
  - Assume a single receive antenna for simplicity, and transmit with a precoder $\mathbf{w}_s = \bar{\mathbf{h}}_s^H$ and a transmit power $s_s = E_s/n_t \; \forall s = 1, \ldots, K$
  - For large n_t, the SINR ρ_q of user q simplifies as

$$\rho_q = \frac{\Lambda_q^{-1} \left|\mathbf{h}_q \mathbf{w}_q\right|^2 E_s/n_t}{\sum_{\substack{p \in \mathbf{K} \\ p \neq q}} \Lambda_q^{-1} \left|\mathbf{h}_q \mathbf{w}_p\right|^2 E_s/n_t + \sigma_{n,q}^2} \overset{n_t \nearrow}{\approx} \frac{\Lambda_q^{-1} \left\|\mathbf{h}_q\right\|^2 E_s/n_t}{\sigma_{n,q}^2} \approx \frac{\Lambda_q^{-1} E_s}{\sigma_{n,q}^2} = \eta_q$$

    and the sum-rate is equal to

$$C_{BF}\left(\mathbf{H}\right) \approx \bar{C}_{BF} \approx \bar{C}_{BF} \approx \sum_{q=1}^{K} \log_2\left(1 + \eta_q\right)$$

    for a total transmit power $K E_s/n_t$.
  - By matched beamforming with a power E_s/n_t per user in a large MISO system (i.e. the transmit power is scaled down proportionally to $1/n_t$), each of the K users gets the same rate as if it were scheduled on a SISO AWGN channel with a transmit power E_s (and received SNR η_q) without any intra-cell interference and without any fading!
  - Assuming $\eta_q = \eta \; \forall q$, the total achievable sum-rate writes as K times the SISO AWGN rate.
  - The transmit power is scaled down proportionally to $1/n_t$ and the multiplexing gain increased proportional to K!

Linear Precoding: Zero-Forcing Beamforming

- ZFBF precoding: $\mathbf{w}_q$ for user $q \in \mathbf{K}$ writes as

$$\mathbf{w}_q = \mathbf{F}(:,q)/\|\mathbf{F}(:,q)\| = \bar{\mathbf{F}}(:,q)/\|\bar{\mathbf{F}}(:,q)\|$$

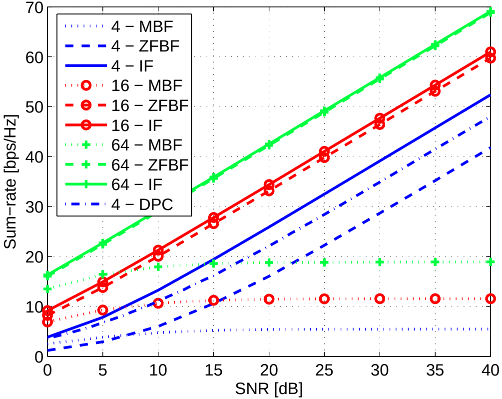
  with

$$\begin{aligned}\mathbf{F} &= \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H\right)^{-1}, \\ &= \underbrace{\bar{\mathbf{H}}^H\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H\right)^{-1}}_{\bar{\mathbf{F}}}\mathbf{D}^{-1}.\end{aligned}$$

- Massive MIMO effect also benefits ZFBF:
  - As n_t grows, $\mathbf{H}\mathbf{H}^H$ and $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ become better conditioned, thereby simplifying the computation of the matrix inverse.
  - In the limit where user channels are orthogonal, $\mathbf{H}\mathbf{H}^H$ and $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ are diagonal and ZFBF boils down to MBF.

Sum-Rate Evaluations

- Performance of MBF, ZFBF, DPC, IF in $n_t = 4, 16, 64$ and $K = 4$ i.i.d. Rayleigh fading channels.
  - IF stands for interference free and is the upper bound on the performance obtained assuming perfect matched beamforming, no intra-cell interference and an uniform power allocation across the four users, leading to a sum-rate of $\sum_{q=1}^{K} \log_2 \left(1 + \eta_q / K \left\|\mathbf{h}_q\right\|^2\right)$.



Sum-Rate Evaluations

- As n_t grows, the gap between IF and ZFBF shrinks significantly:
  - Severe gap exists in the four transmit antenna case between IF, DPC and ZFBF,
  - The gap completely vanishes with 64 transmit antennas with ZFBF performing as well as an IF system.
  - Hence the performance gain of advanced precoding techniques does not justify the complexity increase.
- MBF on the other hand performs relatively poorly (except at low SNR)
  - sum-rate performance fundamentally limited by intra-cell interference and his SINR is limited at high SNR by the ratio $\alpha = n_t/K$.
  - MBF requires a much larger number of antennas to reach the same performance as ZFBF.
- In summary, for $K >> n_t$ and $n_t >> K$, simple linear precoding schemes provide very competitive alternatives to more complex (non-linear) strategies.

Inter-Cell Interference

- As the number of transmit antennas n_t increases, assuming perfect CSIT, the matched beamformer of user q in cell i becomes orthogonal to co-scheduled users' channels $(s \neq q)$ in cell i but also to victim users' channels in adjacent cells. Hence, the intra-cell and inter-cell interference is naturally eliminated as evidenced by (assuming for simplicity $n_{t,i} = n_t \ \forall i$ and $n_{r,q} = n_r \ \forall q$)

$$\lim_{n_t \to \infty} \frac{1}{n_t} \mathbf{H}_{l,i} \mathbf{H}_{p,i}^H = \mathbf{I}_{n_r} \delta_{lp}, \qquad \forall l, p \in \mathcal{K}_i,$$
$$\lim_{n_t \to \infty} \frac{1}{n_t} \mathbf{H}_{l,i} \mathbf{H}_{p,j}^H = \mathbf{I}_{n_r} \delta_{lp}, \qquad \forall l \in \mathcal{K}_i, p \in \mathcal{K}_j.$$

- Inter-cell interference is naturally mitigated and the need for multi-cell coordination or cooperation therefore vanishes as the number of transmit antennas increases!
- Same for Uplink

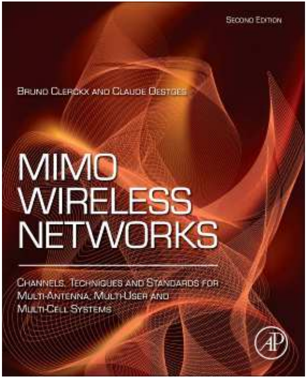
Practical issues

- Cost of a BS may increase.
- Cost of the network infrastructure (backhaul and coordination) may decrease.
- Transmit power can be decreased proportionally to n_t.
  - RF electronics behind every antenna is therefore able to operate at a significantly lower power operating point.
- More antennas packed in a limited volume, increased spatial correlation and antenna coupling.
  - But recall that spatial correlation/LoS can be beneficial to MU-MIMO.
- Accurate CSIT
  - FDD?
  - Reciprocity in TDD (if perfect calibration is performed), but the presence of pilot contamination originating from the reuse of the same pilots by users in different cells degrades uplink channel estimates and therefore limits the performance of Massive MIMO

Real-World MIMO Wireless Networks

Reference Book

- Bruno Clerckx and Claude Oestges, "MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems," Academic Press (Elsevier), Oxford, UK, Jan 2013.



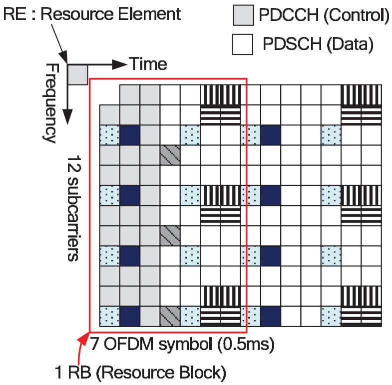
  - Chapter 14

System Requirements

- peak rate
  - highest theoretical throughput achievable with SU-MIMO spatial multiplexing but are typically not achieved in practical deployments.
  - e.g. 8x8 Spatial multiplexing with 8 streams transmission
- cell average spectral efficiency
  - average spectral efficiency of a cell (with K users).
  - much more representative of throughput encountered in practice
- cell edge user spectral efficiency
  - spectral efficiency achieved by at least 95% of the users in the network.
  - much more representative of throughput encountered in practice

Frame Structure

- Multiplexing/Access
  - DL: OFDM
  - UL: DFT-Spread OFDM (SC-FDM)
- Frame structure
  - OFDMA/SC-FDMA create a time-frequency grid composed of time-frequency resources
  - A resource block (RB) is formed by 12 consecutive REs in the frequency domain for a duration of 7 OFDM symbols in the time domain.
  - A subframe is formed of 14 consecutive OFDM/SC-FDM symbols.
  - Scheduling and data transmission is performed at the RB-level with the minimum scheduling unit consisting of two RBs within one subframe.
  - First 3 symbols used to carry control information.



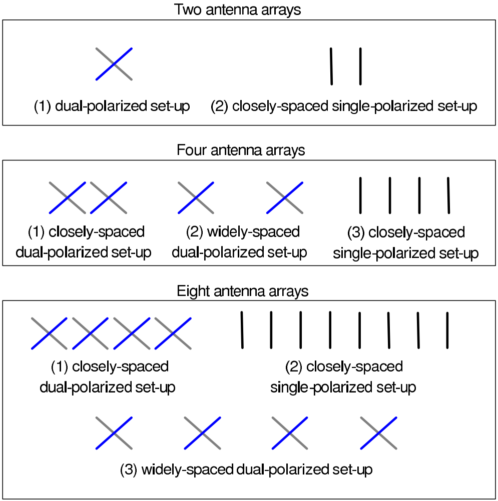
Key Downlink Technologies

- Antenna configurations: 2, 4 or 8 transmit antennas and a minimum of 2 receive antennas
- LTE Rel. 8 (finalized in Dec 2008):
  - Up to 4x4 (up to 4 stream transmission)
  - Transmit diversity (to protect against fading) using Orthogonal Space-Frequency Block Coding (O-SFBC) for 2Tx, non-orthogonal SFBC for 4TX
  - Open-loop (for high speed) Spatial Multiplexing with rank adaptation based on predefined precoders
  - Closed-loop (for low speed) Spatial Multiplexing based on codebook precoding
  - Stone-age MU-MIMO based on common reference signals (CRS)
- LTE Rel. 9 (finalized in Dec 2009):
  - Up to 4x4 (up to 4 stream transmission)
  - Introduction of demodulation reference signals (DM-RS)
  - Enhancement of MU-MIMO to support ZFBF-like precoding
- LTE-A Rel. 10 (finalized mid 2011):
  - Up to 8x8 (up to 8 stream transmission)
  - New channel measurement reference signals (CSI-RS)
  - New feedback mechanisms for 8Tx (dual codebook $\mathbf{W}_1\mathbf{W}_2$ structure)
  - HetNet - eICIC
- LTE-A Rel. 11 (finalized in Dec 2012):
  - Coordinated Multi-Point Transmission/Reception (CoMP) for Homogeneous (Macro) and heterogeneous (pico, DAS) networks
    - Dynamic cell/point selection combined with dynamic ON/OFF blanking

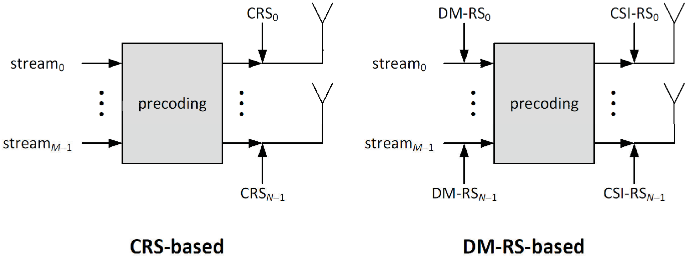
Key Uplink Technologies

- Antenna configurations: 1, 2 or 4 transmit antennas in the uplink with a minimum of 2 receive antennas
- LTE Rel. 8 (finalized in Dec 2008):
  - single antenna transmission and transmit antenna selection
  - MU-MIMO
- LTE-A Rel. 10 (finalized mid 2011):
  - Spatial Multiplexing with codebook
  - Transmit diversity (for control channels)
- LTE-A Rel. 11 (finalized in Dec 2012):
  - Coordinated Multi-Point Transmission/Reception (CoMP)

Antenna Deployments

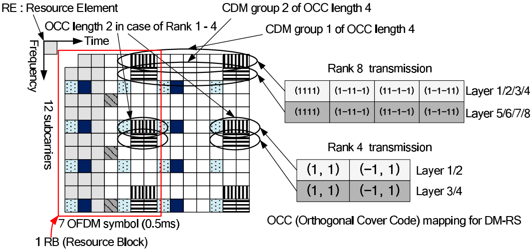


Reference Signals

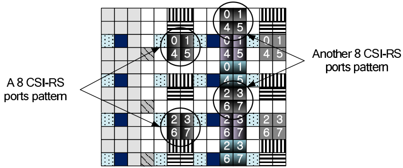


Dedicated RS (DRS)	Common RS (CRS)
For demodulation	For demodulation and measurement
Targets a specific terminal	Shared among a group of terminals
Terminal specifically precoded	Commonly non-precoded
Overhead proportional to the number of transmitted streams	Overhead proportional to the number of transmit antennas
Sent in RBs where data is present	Sent in all RBs
Channel estimation less flexible	Channel estimation more flexible

Reference Signals



(a) DM-RS ports

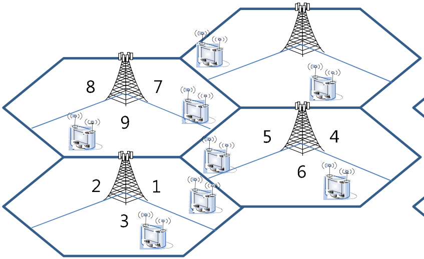


(b) 8 CSI-RS ports

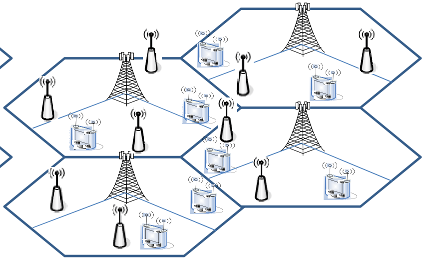
Channel State Information (CSI) feedback

- Three main feedback information:
  - Rank Indicator (RI): the preferred number of streams (denoted as layers in LTE) a user would like to receive
  - Precoding Matrix Indicator (PMI): the preferred precoder in the codebook
  - Channel Quality Indicator (CQI): the rate achievable with each stream (used to perform link adaptation)
- Open-Loop relies only on RI and CQI
  - High mobility or limited CSI feedback prevent the use of PMI
- Closed-Loop (Spatial Multiplexing and MU-MIMO) rely on RI, CQI and PMI
  - If Spatial Multiplexing, the actual precoder is the same as the one selected by the user (PMI)
  - If MU-MIMO based on CRS, the actual precoder is the same as the one selected by the user (PMI)
  - If MU-MIMO based on DM-RS, the actual precoder (e.g. ZFBF) is computed based on the one selected by the user (PMI).

Network Deployments



(a) Homogeneous network



(b) Heterogeneous network

Network Deployment Scenarios

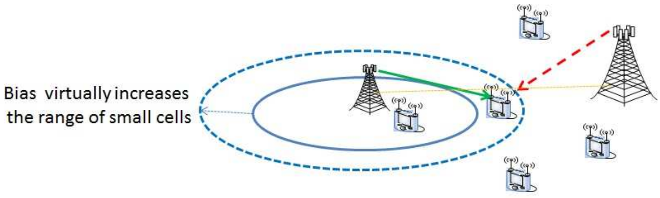
0. *uncoordinated macrocell deployment*: each macrocell is controlled independently of its neighbors,
1. *macro intra-site homogeneous deployment*: coordination between the cells (sectors), e.g. cells 1,2,3, controlled by the same macro base station (where no standardized backhaul interface is needed),
2. *macro inter-site homogeneous deployment*: coordination between cells, e.g. cells 1,5,9, belonging to different radio sites from a macro network,
3. *macro-pico heterogeneous deployment*: macrocell overlaid with low power open access points with possible coordination between the macrocell and low power transmission/reception points within its coverage, each point controlling its own cell (with its own cell identity),
4. *distributed antennas*: the same deployment as (3), except that the low-power transmit/receive points constitute distributed antennas of the macrocell, and are thus all associated with the macrocell identity,
5. *macro-femto heterogeneous deployment*: macrocell overlaid with low power closed access points with no standardized interface between the macro and femtocells.

Macro-Pico Heterogeneous Deployment

- Cell association
  - Connect a terminal (for both DL and UL links) to the cell with the strongest received DL power.
  - In HetNet, owing to the transmit power difference between the high power and low power node, it is preferable to connect to the cell with the lowest path-loss in the UL and connect to the cell with the strongest received DL power in the DL.
  - The UL coverage area therefore becomes larger than the DL coverage area, leading to different cell associations in the DL and UL and particularly complexifying the system.
- Cell loading
  - Balance the load among macro- and picocells in highly loaded cells in order to maximize the resource reuse between cells.

Macro-Pico Heterogeneous Deployment

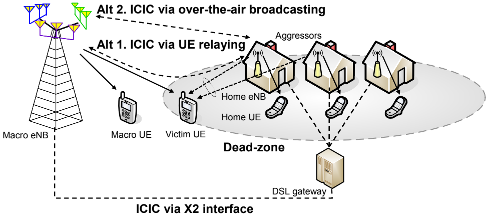
- To allow flexible load balancing and keep the same cell association for DL and UL, LTE-A supports *cell range expansion* (CRE)
  - The range of the low-power node is controlled by a cell association bias.
  - Expansion so that a UE may be associated with a cell which does not provide the strongest received signal power.
  - Unfortunately, range expansion brings inter-cell interference whose severity increases as the cell association bias increases.



- Enhanced ICIC: Almost Blank Subframes (ABSF) introduced in LTE-A
  - Semi-static form of the time-domain ON-OFF (binary) power control
- CoMP: Dynamic cell/point selection combined with dynamic ON/OFF blanking

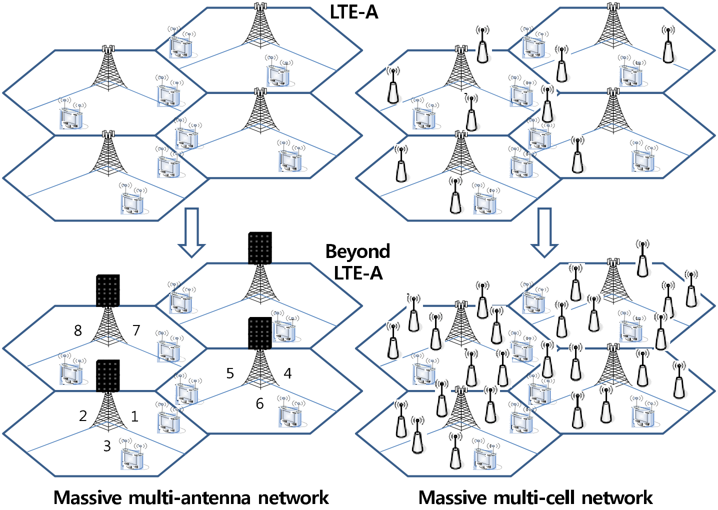
Macro-Femto Heterogeneous Deployment

- In heterogeneous femto networks, femtocells are randomly deployed by consumers without coordination.
- Since they cannot rely on the X2 interface in LTE-A, femtocells cause severe downlink and uplink interference to adjacent cells.
- Downlink dead-zone problem: When a non-CSG UE is located in the vicinity of a Femto BS (HeNB), the harsh interference from HeNB will create an outage area.



- When the probability of being located in the vicinity of HeNB and the deployment ratio of HeNB are both low, *static ICIC* is sufficient.
  - silencing resources by e.g. ON-OFF, blank subframes
- When the deployment ratio of HeNB is high and the outage probability becomes noticeable, *dynamic ICIC* is useful.

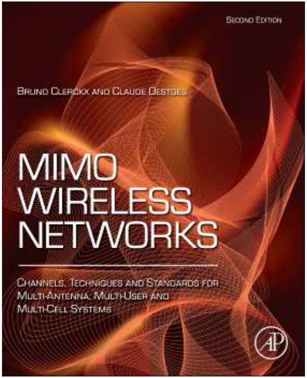
Beyond LTE-A: Massive Multi-Cell and Massive Multi-Antenna Networks



System-Level Performance Evaluations

Reference Book

- Bruno Clerckx and Claude Oestges, “MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems,” Academic Press (Elsevier), Oxford, UK, Jan 2013.



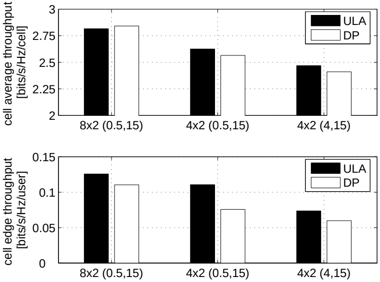
  - Chapter 15

System Level Assumptions

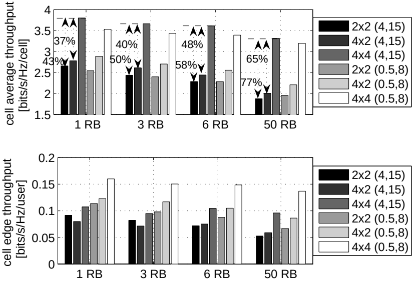
- System Level Simulations compliant with LTE-A evaluation methodology
- Assumptions:
  - DL synchronized LTE-Advanced network based on FDD and 10 MHz bandwidth made of 50 resource blocks (RB).
  - Homogeneous network
  - 57 (=19x3) cell sites and 10 users dropped per cell.
  - Full-buffer traffic
  - Link adaptation: Adaptive coding, modulation and transmission rank combined with HARQ based on Chase Combining (target BLER 10%)
  - Proportional Fair scheduling
- Careful about notations: $n_t \times n_r$ as in LTE terminology!
- Performance metric
  - cell average throughput [bits/s/Hz/cell]: sum of the average throughput of each user in a cell.
  - cell edge troughput [bits/s/Hz/user]: 5^{th} percentile of the CDF of the user average throughput.

Single-User MIMO

Antenna deployment (with 6RB)



Antenna configuration (for ULA)



Observations:

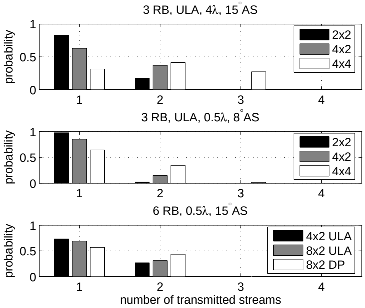
- $8 \times 2 > 4 \times 2$ (larger transmit array gain and SM gain).
- DP $>$ ULA for cell average if n_t large (SM gain $\&$ large array gain on each pol).
- ULA$>$DP for cell edge (large array gain due to spatial correlation).
- Large AS $<$ small AS (large array gain $>$ large SM gain).

Observations:

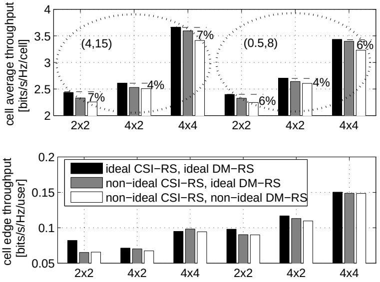
- $4 \times 2 \geq 2 \times 2$, $4 \times 4 >> 4 \times 2$ (symmetric conf. better).
- gain of large n_r higher for less correlated scenarios and large subband size.

Single-User MIMO

Statistics of the transmission rank



Channel estimation errors (ULA,3RB)



Observations:

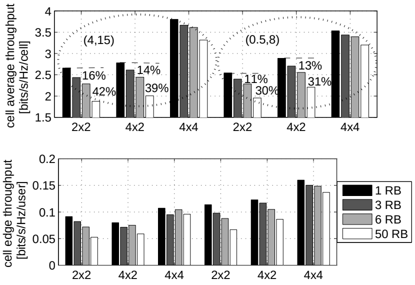
- higher ranks as n_t, n_r, spacing and AS increase.
- higher ranks for DP than ULA.
- rank-1 transmission most encountered.
- symmetric antenna set-up: full rank transmission negligible.

Observations:

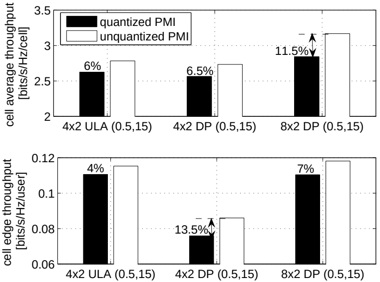
- loss of 6 to 7% in the cell average and cell edge throughputs.

Single-User MIMO

Feedback subband size



Quantized and unquantized PMI



Observations:

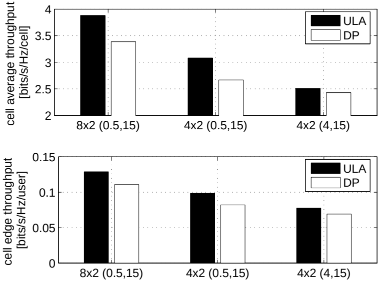
- as the subband size increases, the performance decreases (due to channel frequency selectivity)
- less pronounced in spatially correlated environment $(0.5, 8)$.

Observations:

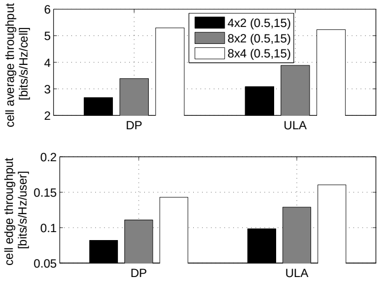
- loss incurred by codebook quantization larger in DP compared to ULA deployments and in 8×2 compared to 4×2.

Multi-User MIMO

Antenna deployment



Antenna configuration



Observations:

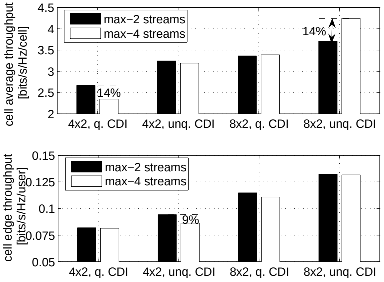
- ULA $>$ DP for cell average for any n_t
- ULA $>$ DP at the cell edge
- $0.5\lambda > 4\lambda$

Observations:

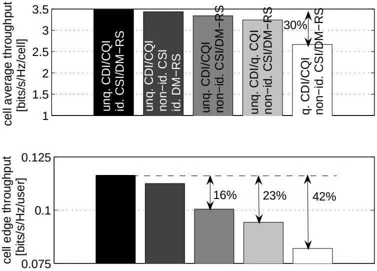
- 8x4 provides significant gain over 8x2
  - 8Tx ZFBF is far from nulling out MU interference
  - more pronounced in DP

Multi-User MIMO

MU-MIMO dimensioning without overhead



CSI measurement and feedback



Observations:

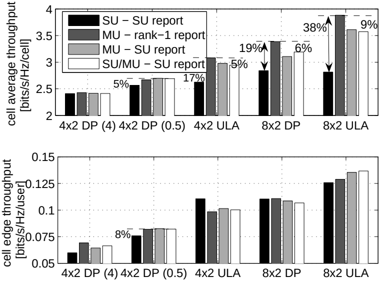
- 4×2: 4 streams $>$ 2 streams with accurate feedback, 2 streams $>$ 4 streams with quantized feedback
- 8×2: 4 streams $>$ 2 streams if accurate and quantized feedback

Observations:

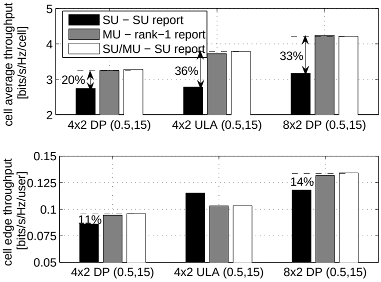
- ranking of losses in increasing order of severity: CSI-RS $<$ q. CQI $<$ DM-RS $<<$ q. CDI (assuming 6RB subband size and feedback delay)

Multi-User MIMO

Dynamic switching based on quantized feedback



Dynamic switching based on unquantized feedback



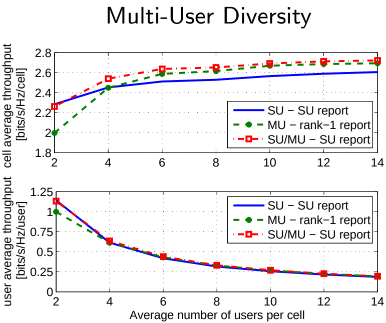
Observations:

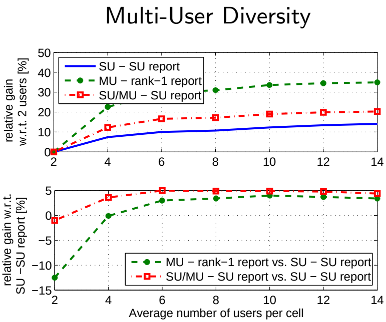
- MU-MIMO and SU/MU-MIMO dynamic switching bring negligible gain over SU in 4×2 DP (4,15)
- MU-MIMO and SU/MU-MIMO dynamic switching bring only 5% gain over SU in 4×2 DP (0.5,15)
- MU with rank-1 report $>$ MU w/ SU-MIMO report and SU/MU w/ SU

Observations:

- dynamic SU/MU-MIMO $\geq$ MU-MIMO based on rank-1 report

Multi-User MIMO



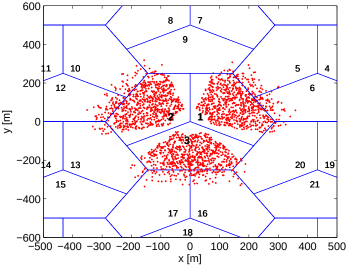


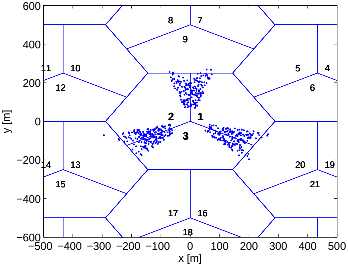
Observations:

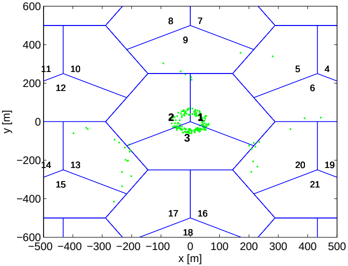
- SU-MIMO $>$ MU-MIMO for small K (less than 4) while MU-MIMO $>$ SU-MIMO for large K.
- MU-MIMO relies more heavily on multi-user diversity than SU-MIMO
- SU/MU-MIMO dynamic switching outperforms both SU-MIMO and MU-MIMO for any number of users.

Multi-Cell MIMO

Which users benefit from intra-site cooperation?

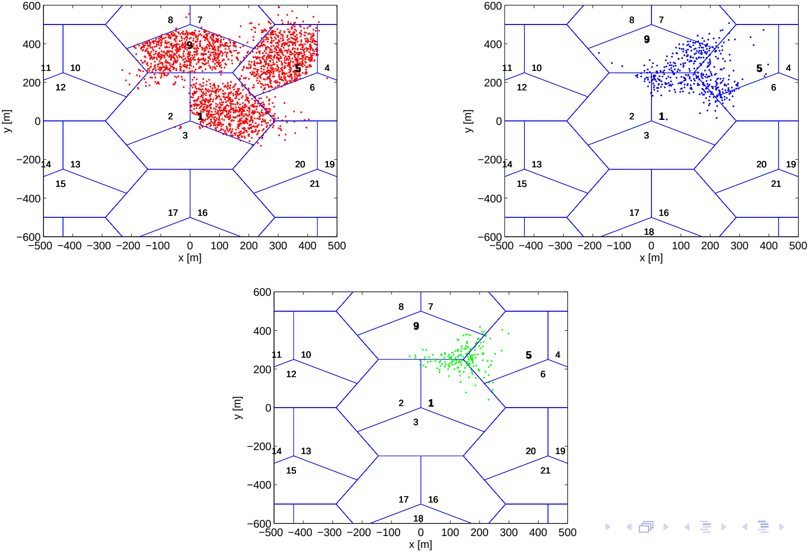






Multi-Cell MIMO

Which users benefit from inter-site cooperation?



Multi-Cell MIMO: Intra-site vs. Inter-site Clustering

- Percentage of users whose CoMP (MC) measurement set size is 1 to 6 for inter-site and intra-site (10dB triggering threshold).

CoMP measurement set size	1	2	3	4	5	6
Inter-site CoMP	53%	23%	18%	3%	2%	1%
Intra-site CoMP	75%	19%	6%	0%	0%	0%

Observations:

- By constraining the cooperation to intra-site deployments, the percentage of CoMP users is significantly decreased.
- No benefit to consider CoMP measurement set sizes larger than 3 for such triggering threshold.

Multi-Cell MIMO: User-centric vs. Network Predefined Clustering

- Percentage of users whose CoMP (MC) measurement set size is 1 to 6 for user-centric and network predefined cooperating (or clustering) set (10dB triggering threshold).

CoMP measurement set size	1	2	3	4	5	6
user-centric cooperating set	53%	23%	18%	3%	2%	1%
inter-site network predef. cooperating set	80%	14%	6%	0%	0%	0%

- Percentage of users whose CoMP (MC) measurement set size is 1 to 6 for two different network predefined cooperating (or clustering) sets (10dB triggering threshold).

CoMP measurement set size	1	2	3
intra-site network predefined cooperating set	75%	19%	6%
inter-site network predefined cooperating set	80%	14%	6%

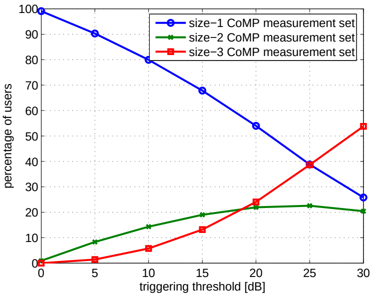
Observations:

- Network predefined clustering reduces the occurence of CoMP (MC) users
- CoMP measurement set size depends on the network predefined cooperating set.
  - intra-site network predefined cooperating set provides larger CoMP measurement sets than an inter-site network predefined cooperating set

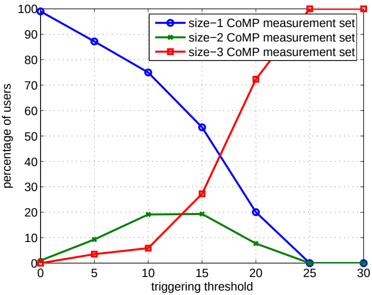
Multi-Cell MIMO: Feedback Overhead

- Percentage of users whose CoMP measurement set size is 1,2 or 3 as a function of the triggering threshold with a network predefined clustering.

Inter-site



Intra-site



Observations:

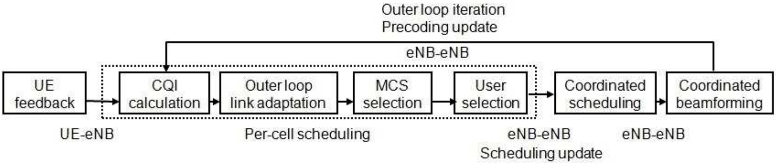
- With a 10dB triggering threshold

Deployment	Absolute overhead	Overhead increase
intra-site	$K * B * (0.75 * B + 0.19 * 2 + 0.06 * 3) = 1.31KB$	31%
inter-site	$K * B * (0.53 * B + 0.23 * 2 + 0.24 * 3) = 1.71KB$	71%

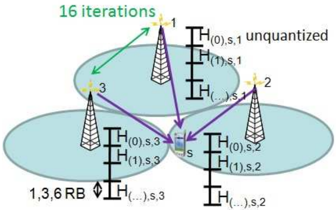
- User centric clustering requires higher overhead than network predefined clustering.
- With network predefined clustering, inter-site requires less overhead than intra-site.

Coordinated Scheduling and Beamforming

- Assumptions:
  - Coordinated SU-MIMO in homogeneous network: one user scheduled at a time in each cell on a given time/frequency resource
  - Network level iterative coordinated scheduling and beamforming based on interference pricing, Signal-to-Leakage-and-Noise-Ratio (SLNR) filter design and user-centric clustering

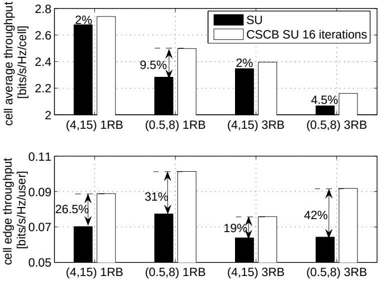


  - Unquantized but average (at the subband level) CSIT

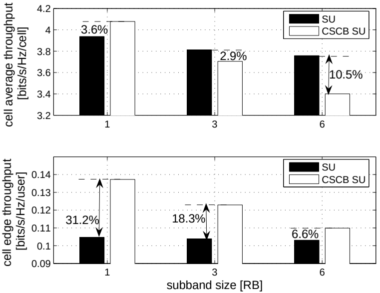


Coordinated Scheduling and Beamforming

Iterative CSCB SU-MIMO vs. SU-MIMO
$n_t \times n_r = 4 \times 2$ ULA (4,15) and (0.5,8)



Iterative CSCB SU-MIMO vs. SU-MIMO
$n_t \times n_r = 4 \times 4$ ULA (4,15)

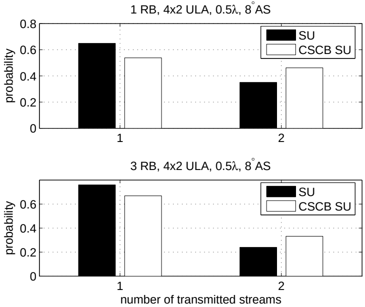


Observations:

- Gain of about 30% coordination at the cell edge for a CSI overhead increase of 71%
- Big loss as the CSIT accuracy decreases
  - ... and this assumed unquantized CSI, user receiver implementation assumed known at the Tx, perfect CSI measurement, no delay

Coordinated Scheduling and Beamforming

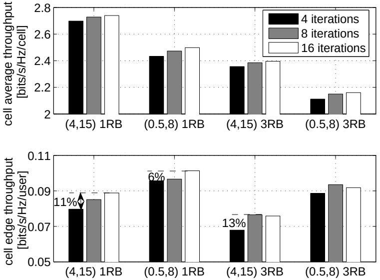
Statistics of the transmission rank



Observations:

- Coordination allows cell edge users to benefit from spatial multiplexing gains

Number of iterations $n_t \times n_r = 4 \times 2$ ULA

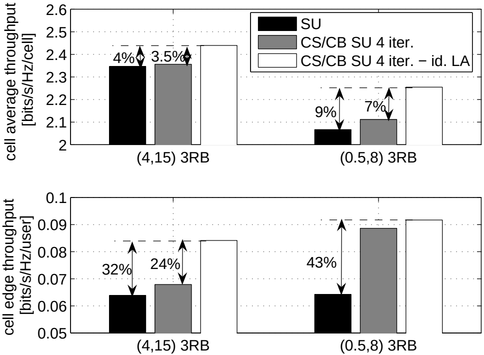


Observations:

- Convergence rate depends on deployment

Coordinated Scheduling and Beamforming

Ideal and non-ideal link adaptation

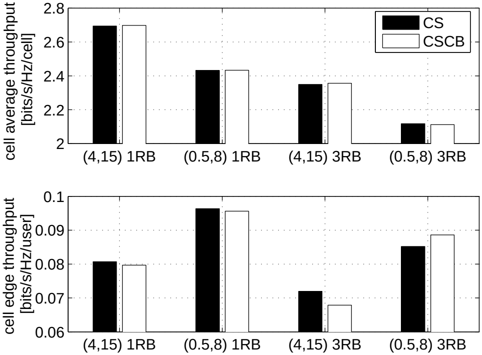


Observations:

- Most of the potential gain lost due to inaccurate LA.
- Inaccurate CQI prediction hampers the appropriate selection of the users, the transmission ranks and the beamformers at every iteration of scheduler and ultimately the whole link adaptation and the convergence of the scheduler

Coordinated Scheduling and Beamforming

CS vs. CS/CB



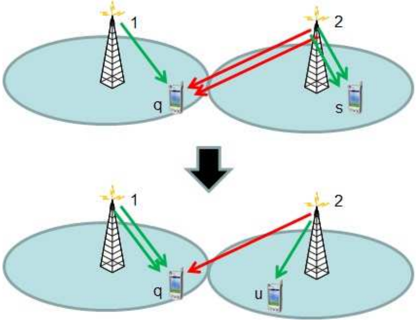
Observations:

- CS only brings all the gains

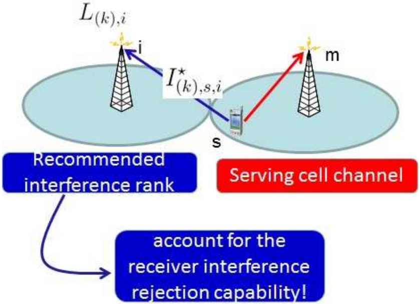
Robust Joint Scheduling and Rank Coordination

B. Clerckx et al., A Practical Cooperative Multicell MIMO-OFDMA Network based on Rank Coordination, IEEE Trans. on Wireless Comm. vol. 12, no. 4, pp. 1481-1491, April 2013.

Improve cell edge user experience by enabling robust multi-streams transmission to cell edge users

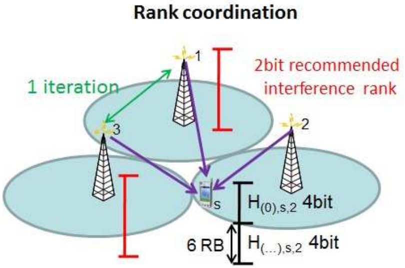


Rank coordination: Each cell edge UE recommends the interfering cells to use a transmission rank that is the most beneficial to its performance

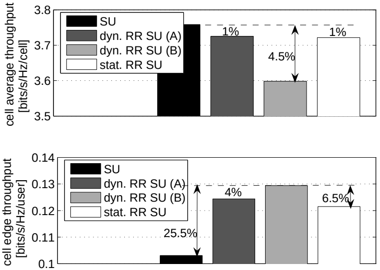


Robust Joint Scheduling and Rank Coordination

Simulation assumptions



SU-MIMO with and without coordination

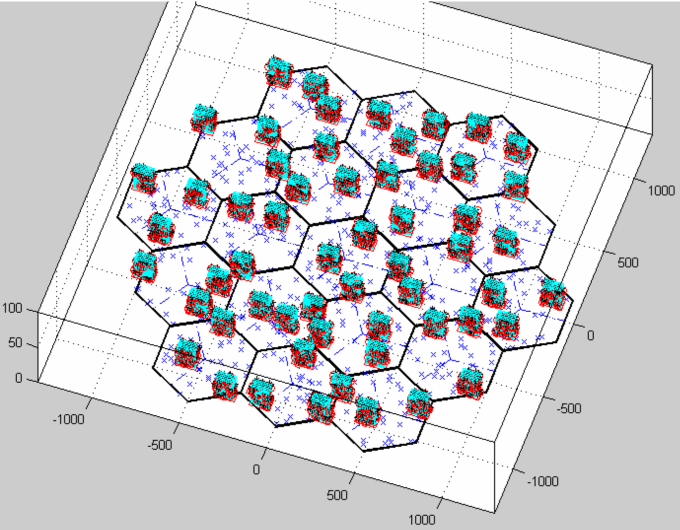


Observations:

- Better performance gain with a significantly lower feedback overhead and scheduler complexity
- About 20% gain at the cell edge with only 2-bit additional feedback compared to uncoordinated SU-MIMO

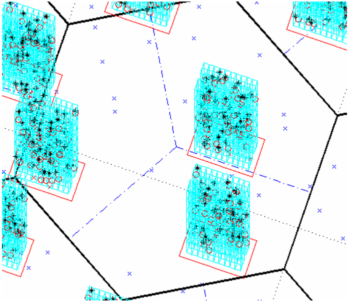
Heterogeneous Networks

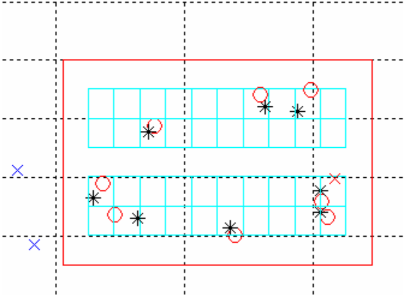
Two-tier Macro - Femto



Heterogeneous Networks: Femto

Two-tier Macro - Femto

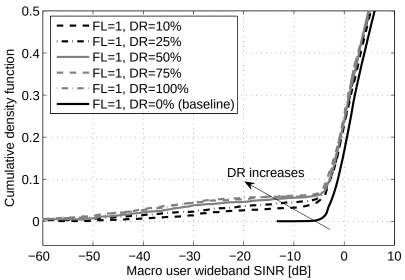


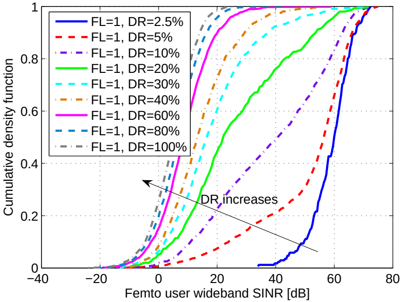


Macro MT ($\times$) inside and outside the apartment complex, femto MT (o), femto BS ($*$).

Heterogeneous Networks: Femto

Downlink Dead-Zone Problem: Wideband SINR distribution



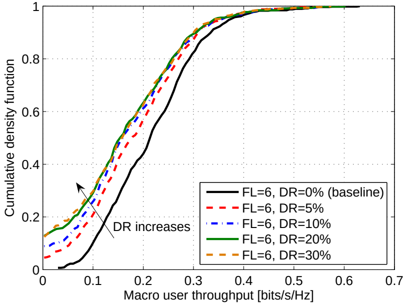


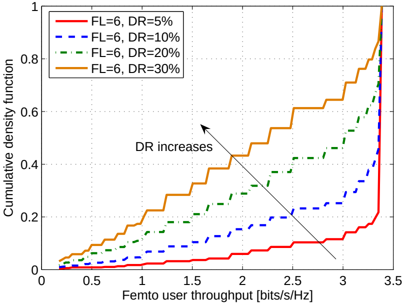
Observations:

- Significant degradation as the deployment ratio (DR) increases
- Interference to Macro MT originates from a small number of dominant Femto BS
- Femto user subject to the increase of the interference level from Femto BS as DR increases

Heterogeneous Networks: Femto

Downlink Dead-Zone Problem: Throughput Analysis



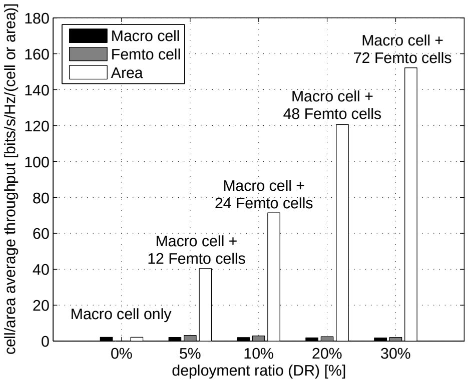


Observations:

- A non-negligible cell average throughput loss for macro
- The macro outage probability (about 13.5%) matches the ratio between the HeNB cluster area and the cell sector area.
- The femto cell throughput is affected by the increase of inter-femto cells interference.

Heterogeneous Networks: Femto

Downlink Dead-Zone Problem: Throughput Analysis

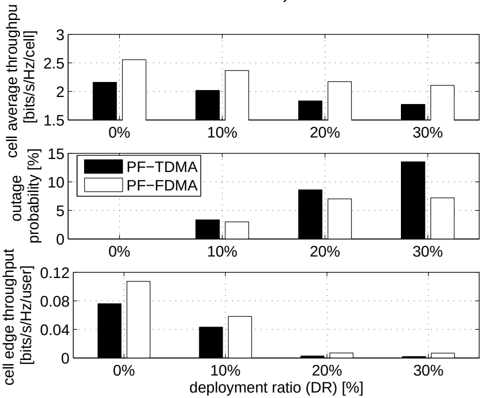


Observations:

- The cell area throughput is boosted and benefits from the increase in the cell-splitting gain of femto cells.
- As the DR (i.e., the number of femtocells) is doubled, the area throughput is however not doubled due to the inter-cell interference increase.

Heterogeneous Networks: Femto

- PF-TDMA (all frequency resources allocated to a single user on a subframe basis) was assumed.
- PF-FDMA (resources allocated to users on RB-level and multiple users supported in different allocations in the same subframe)?

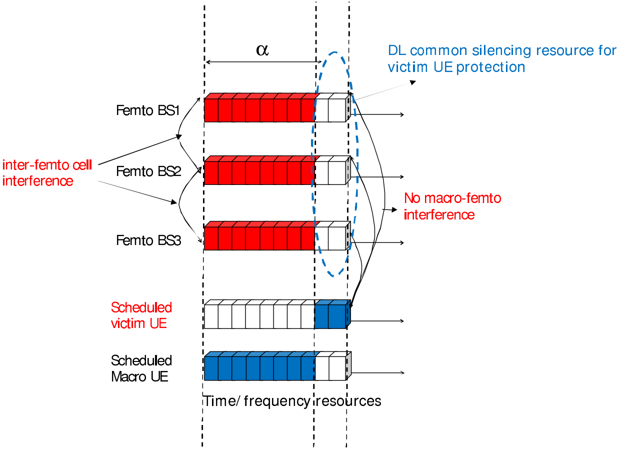


Observations:

- PF-FDMA helps the macro user performance (Outage still large but decreased).
- Exploit frequency selective scheduling in interference-limited environments.

Static Binary ON/OFF Power Control in Heterogeneous Networks

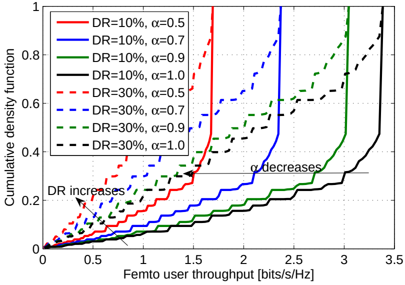
Common Resource Silencing for non reliable backhaul with PF-FDMA



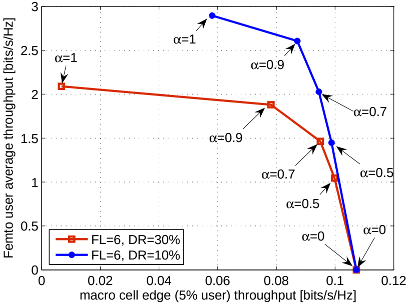
- Femto resource utilization factor α
- Common silencing resource is allocated exclusively for macro MT protection.
- No femto MT can be allocated on that common resource.

Static Binary ON/OFF Power Control in Heterogeneous Networks

Femtocell throughput



Macro and femto throughput

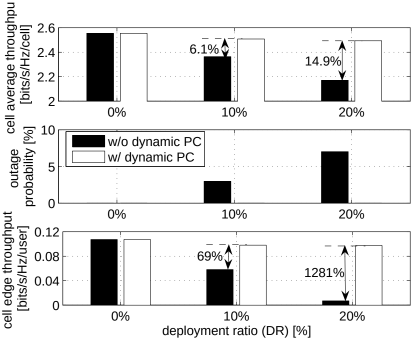


Observations:

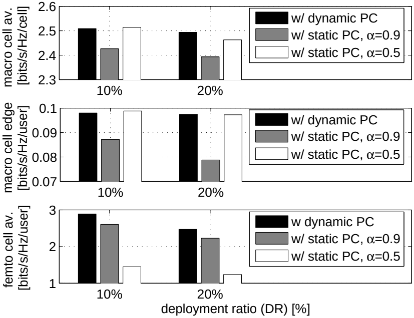
- As α decreases, a decrease of the femto MT throughput is observed because of the inability of femto cells to use some of the resources.
- The frequency-domain common resource silencing boosts the macro cell edge (macro MT 5%) throughput but decreases the femto MT throughput.
- Any enhancement of the victim macroMT throughput goes with a femto MT throughput degradation.
- the inter-femto cell interference is not negligible and significantly affects the performance: need to protect victim macro MT and victim femto MT

Dynamic Binary ON/OFF Power Control in Heterogeneous Networks

Macro performance - Dynamic PF-FDMA



Dynamic vs. static PF-FDMA coordination

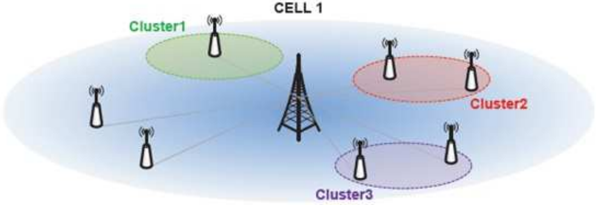


Observations:

- PF-TDMA/FDMA eliminate outage
- PF-FDMA > PF-TDMA but gain of coordination smaller in PF-FDMA than PF-TDMA
- PF-FDMA increases the overhead of inter-cell message passing
- Dynamic coordination enables a better tradeoff between macro throughput and femto throughput compared to static coordination

Heterogeneous Networks - DAS

- Distributed Antenna System (DAS)



  - Numerous nodes create more cell boundaries
  - Overlay of macro and small cells enlarges the interference zone.
  - More UEs become eligible to benefit from CoMP in HetNet

- Dynamic point selection with dynamic blanking (ON/OFF power control) in clustered heterogeneous deployments

	N	Av. area thrpt	cell edge thrpt
Rel. 10 (0dB RE,no ABSF)	4	16.41	0.0574
Rel. 10 (20dB RE,60% ABSF)	4	16.50 (1%)	0.0668 (16%)
DAS with DS	4	15.55 (-5.2%)	0.0698 (21.6%)
DAS with DS/DB	4	16.68 (1.6%)	0.0840 (46.3%)
Rel. 10 (0dB RE,no ABSF)	10	22.33	0.0708
Rel. 10 (20dB RE,60% ABSF)	10	23.76 (6%)	0.0937 (32%)
DAS with DS	10	22.66 (1.5%)	0.0820 (15.8%)
DAS with DS/DB	10	23.27 (4.2%)	0.1067 (50.7%)

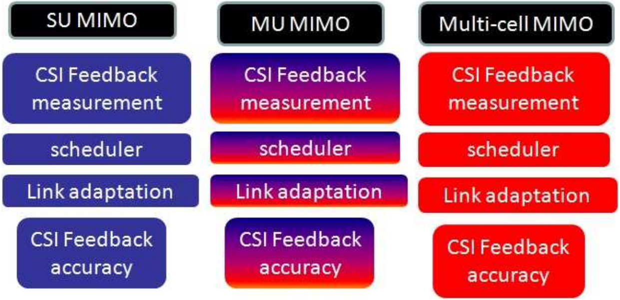
- cooperation/coordination gain larger in heterogeneous than homogeneous

Some Conclusions

- Potential gain of single-cell and multi-cell MIMO in theory but benefits may vanish in practical scenarios
  - Sensitivity to CSI measurement: channel estimation errors particularly large for cell edge users
  - CSI feedback inaccuracy: Limited feedback, Subband feedback with strong frequency selectivity within subband, Particularly problematic in dual-polarized antenna deployments
  - Latency of the feedback and the backhaul
  - Feedback and message exchange overhead: Target cell edge users
  - Inaccurate link adaptation: due to feedback inaccuracy, BS does not know the receiver at the mobile terminal, Traffic model, Fast variation of the inter-cell interference
  - Scheduler convergence and complexity
  - Many other issues left: time/frequency synchronization, antenna calibration, ...
- Sensitivity different depending on SU-MIMO, MU-MIMO, Muti-Cell MIMO

Some Conclusions

- Current wireless system design is at the network level
  - Lots of aspects interact with each other
- Network designs become more and more sensitive to impairments



- Gap between theory and practice gets much bigger as we move from single-cell to (cooperative/coordinated) multi-cell designs
- Account for impairments