A tutorial introduction to space-time coding: mathematical models, information theoretical aspects, and coding for MIMO channels

Joseph J. Boutros

Texas A&M University at Qatar

Lecture at the Conference on Coding and Cryptography
Carcans, France

March 2008

Where can I download these slides?

Slides on the web page

http://www.josephboutros.org/C2

Questions can be sent by email to

boutros@tamu.edu

Presentation outline

1. Part I - Introduction
   - Diversity to combat erasures and fadings.
   - The multiple antenna channel model.
   - Coding gain and diversity in MIMO channels.
2. Part II - Information Theory
   - Capacity when channel is unknown at transmitter.
   - Outage probability for non-ergodic channels.
3. Part III - Coding
   - Quick introduction to STBC.
   - Code design criteria for block fading channels.
   - Example of an LDPC code for MIMO channels.

Coding for erasure channels (1)

The erasure channel is an extremal case of the Rayleigh fading channel (see Proakis 2000, Tse & Viswanath 2005). The erasure channel can also model an application layer where packets are lost due to a failure at the physical layer.

Let us consider the binary erasure channel (BEC). A codeword $c = (c_1, c_2, \dots, c_N)$ belonging to $\mathcal{C}[N, K, d_{min}]_2$ is transmitted on the BEC, where $\mathcal{C}$ is a linear binary code of length N, dimension K, and minimum Hamming distance d_{min}.

The iid BEC

- The channel is memoryless. If y denotes the channel output then

$$p(y|c) = \prod_{i=1}^{N} p(y_i|c_i), \qquad p(y_i|c_i) = \begin{cases} 1-\epsilon, & y_i = c_i, \\ \epsilon, & y_i = X, \\ 0, & y_i = \overline{c_i}, \end{cases}$$

  where $c_i \in F_2$, X represents an erasure, and $\epsilon \in [0,1]$.
- Binary elements are erased independently from each other. The output y_i is equal to the input c_i with probability $1-\epsilon$. No errors are encountered.

Coding for erasure channels (1)

The erasure channel is an extremal case of the Rayleigh fading channel (see Proakis 2000, Tse & Viswanath 2005). The erasure channel can also model an application layer where packets are lost due to a failure at the physical layer.

Let us consider the binary erasure channel (BEC). A codeword $c = (c_1, c_2, \ldots, c_N)$ belonging to $\mathcal{C}[N, K, d_{min}]_2$ is transmitted on the BEC, where $\mathcal{C}$ is a linear binary code of length N, dimension K, and minimum Hamming distance d_{min}.

The iid BEC

- The channel is memoryless. If y denotes the channel output then

$$p(y|c) = \prod_{i=1}^{N} p(y_i|c_i), \qquad p(y_i|c_i) = \begin{cases} 1-\epsilon, & y_i = c_i, \\ \epsilon, & y_i = X, \\ 0, & y_i = \overline{c_i}, \end{cases}$$

  where $c_i \in F_2$, X represents an erasure, and $\epsilon \in [0, 1]$.
- Binary elements are erased independently from each other. The output y_i is equal to the input c_i with probability $1-\epsilon$. No errors are encountered.

Coding for erasure channels (2)

The non-ergodic BEC

- The channel has memory. Let us divide the codeword c into L blocks C_ℓ, $\ell = 1 \dots L$, each block has length N/L bits.
- Blocks are erased independently from each other, an erasure occurs with probability ϵ. After writing $y = (Y_1, \dots, Y_L)$ and $c = (C_1, \dots, C_L)$, we get

$$p(y|c) = \prod_{\ell=1}^{L} p(Y_\ell|C_\ell), \qquad p(Y_\ell|C_\ell) = \begin{cases} 1-\epsilon, & Y_\ell = C_\ell, \\ \epsilon, & Y_\ell = X_1^L, \\ 0, & otherwise, \end{cases}$$

where $Y_\ell \in F_2^{N/L} \bigcup \{X_1^L\}$, and X_1^L represents L erased bits.

Degrees of Freedom

The iid BEC has N degrees of freedom whereas the non-ergodic BEC has only L degrees of freedom. Exempli gratia, $N = 1000$ and $L = 3$.

Coding for erasure channels (2)

The non-ergodic BEC

- The channel has memory. Let us divide the codeword c into L blocks C_ℓ, $\ell = 1 \ldots L$, each block has length N/L bits.
- Blocks are erased independently from each other, an erasure occurs with probability ϵ. After writing $y = (Y_1, \ldots, Y_L)$ and $c = (C_1, \ldots, C_L)$, we get

$$p(y|c) = \prod_{\ell=1}^{L} p(Y_\ell|C_\ell), \qquad p(Y_\ell|C_\ell) = \begin{cases} 1-\epsilon, & Y_\ell = C_\ell, \\ \epsilon, & Y_\ell = X_1^L, \\ 0, & otherwise, \end{cases}$$

where $Y_\ell \in F_2^{N/L} \bigcup \{X_1^L\}$, and X_1^L represents L erased bits.

Degrees of Freedom

The iid BEC has N degrees of freedom whereas the non-ergodic BEC has only L degrees of freedom. Exempli gratia, $N = 1000$ and $L = 3$.

Coding for erasure channels (3)

- Consider an iid BEC and a repetition code $\mathcal{C}(N, 1, N)_2$. The word error probability after decoding is $P_e = \epsilon^N$.
- Consider an iid BEC and a non-trivial code $\mathcal{C}(N, K, d_{min})_2$. A maximum likelihood decoder fails to decode an erasure pattern iff this pattern contains the support of a nonzero codeword (e.g. see Schwartz & Vardy 2005). Let $\Psi_{ML}(\omega)$ denote the number of such erasure patterns with weight ω. Then,

$$P_e(ML) = \sum_{\omega=d_{min}}^{N} \Psi_{ML}(\omega)\epsilon^{\omega}(1-\epsilon)^{N-\omega}.$$

  A similar expression is obtained under iterative decoding using the notion of stopping sets (e.g. see Di, Proietti, Teletar, Richardson, & Urbanke 2002).
- For small ϵ, the asymptotic behavior is

$$P_e(ML) \propto \epsilon^{d_{min}}.$$

Coding for erasure channels (3)

- Consider an iid BEC and a repetition code $\mathcal{C}(N,1,N)_2$. The word error probability after decoding is $P_e = \epsilon^N$.
- Consider an iid BEC and a non-trivial code $\mathcal{C}(N,K,d_{min})_2$. A maximum likelihood decoder fails to decode an erasure pattern iff this pattern contains the support of a nonzero codeword (e.g. see Schwartz & Vardy 2005). Let $\Psi_{ML}(\omega)$ denote the number of such erasure patterns with weight ω. Then,

$$P_e(ML) = \sum_{\omega=d_{min}}^{N} \Psi_{ML}(\omega)\epsilon^{\omega}(1-\epsilon)^{N-\omega}.$$

  A similar expression is obtained under iterative decoding using the notion of stopping sets (e.g. see Di, Proietti, Teletar, Richardson, & Urbanke 2002).
- For small ϵ, the asymptotic behavior is

$$P_e(ML) \propto \epsilon^{d_{min}}.$$

Coding for erasure channels (4)

- Consider a non-ergodic BEC with parameter L and a repetition code $\mathcal{C}(L,1,L)_2$ or a direct sum of N/L versions of this code, i.e., $\mathcal{C} = (L,1,L) \oplus (L,1,L) \ldots \oplus (L,1,L)$. The error probability after decoding the repetition code is $P_e = \epsilon^L$.
- Consider a non-ergodic BEC and a non-trivial code $\mathcal{C}(N,K,d_{min})_2$. Then, the word error probability after decoding satisfies $P_e \geq \epsilon^L$.

Definition 1: Diversity on erasure channels

The diversity order d attained by a code $\mathcal{C}$ is defined as

$$d = \lim_{\epsilon \longrightarrow 0} \frac{\log P_e}{\log \epsilon}.$$

Notice that diversity on iid BEC is upper bounded by d_{min} ($d_{min} \leq N$) whereas diversity on non-ergodic BEC is upperbounded by L ($L \leq N$). Non-ergodic channels are also referred to as limited diversity channels.

Coding for erasure channels (4)

- Consider a non-ergodic BEC with parameter L and a repetition code $\mathcal{C}(L,1,L)_2$ or a direct sum of N/L versions of this code, i.e., $\mathcal{C} = (L,1,L) \oplus (L,1,L) \ldots \oplus (L,1,L)$. The error probability after decoding the repetition code is $P_e = \epsilon^L$.
- Consider a non-ergodic BEC and a non-trivial code $\mathcal{C}(N,K,d_{min})_2$. Then, the word error probability after decoding satisfies $P_e \geq \epsilon^L$.

Definition 1: Diversity on erasure channels

The diversity order d attained by a code $\mathcal{C}$ is defined as

$$d = \lim_{\epsilon \to 0} \frac{\log P_e}{\log \epsilon}.$$

Notice that diversity on iid BEC is upper bounded by d_{min} ($d_{min} \leq N$) whereas diversity on non-ergodic BEC is upperbounded by L ($L \leq N$). Non-ergodic channels are also referred to as limited diversity channels.

Coding for erasure channels (5)

Let ω_ℓ be the Hamming weigh of the block C_ℓ. The codeword weight $\omega(c)$ is the sum of partial weights, i.e., $\omega(c) = \sum_{\ell=1}^{L} \omega_\ell$.

Theorem 2: Design criterion for non-ergodic BEC

$\mathcal{C}$ is full diversity $(d = L)$ under ML decoding on a non-ergodic BEC if and only if, $\forall c \in \mathcal{C} \setminus \{0\}$, all partial Hamming weights are non zero, i.e., $\omega_\ell \neq 0, \forall \ell$.

Example: (Boutros, Guillén i Fàbregas, & Calvanese Strinati 2005)
The code is $\mathcal{C} = [8, 4, 4]_2$ and $L = 2$. There exist $8!/(4!)^2 = 70$ possibilities to define blocks C_1 and C_2. The 70 multiplexers are grouped into 2 different classes:

- 14 multiplexers with diversity 1 (no diversity) and weight enumerator

$$A(x, y) = \sum_{c \in \mathcal{C}} x^{\omega_1} y^{\omega_2} = 1 + x^4 + y^4 + 12x^2y^2 + x^4y^4.$$

- 56 multiplexers with full diversity $(d = L = 2)$ and weight enumerator

$$A(x, y) = 1 + 6x^2y^2 + 4x^3y + 4xy^3 + x^4y^4.$$

Exercice: Define $C_1 = (c_1, \ldots, c_{12})$ and $C_2 = (c_{13}, \ldots, c_{24})$. Find a full-diversity version of the [24, 12, 8] Golay code on $L = 2$ non-ergodic erasure channel.

Coding for erasure channels (5)

Let ω_ℓ be the Hamming weigh of the block C_ℓ. The codeword weight $\omega(c)$ is the sum of partial weights, i.e., $\omega(c) = \sum_{\ell=1}^{L} \omega_\ell$.

Theorem 2: Design criterion for non-ergodic BEC

$\mathcal{C}$ is full diversity ($d = L$) under ML decoding on a non-ergodic BEC if and only if, $\forall c \in \mathcal{C} \setminus \{0\}$, all partial Hamming weights are non zero, i.e., $\omega_\ell \neq 0, \forall \ell$.

Example: (Boutros, Guillén i Fàbregas, & Calvanese Strinati 2005)
The code is $\mathcal{C} = [8, 4, 4]_2$ and $L = 2$. There exist $8!/(4!)^2 = 70$ possibilities to define blocks C_1 and C_2. The 70 multiplexers are grouped into 2 different classes:

- 14 multiplexers with diversity 1 (no diversity) and weight enumerator

$$A(x, y) = \sum_{c \in \mathcal{C}} x^{\omega_1} y^{\omega_2} = 1 + x^4 + y^4 + 12x^2y^2 + x^4y^4.$$

- 56 multiplexers with full diversity ($d = L = 2$) and weight enumerator

$$A(x, y) = 1 + 6x^2y^2 + 4x^3y + 4xy^3 + x^4y^4.$$

Exercice: Define $C_1 = (c_1, \ldots, c_{12})$ and $C_2 = (c_{13}, \ldots, c_{24})$. Find a full-diversity version of the [24, 12, 8] Golay code on $L = 2$ non-ergodic erasure channel.

Coding for fading channels (1)

The fading channel is defined by the input-output relation

$$y_i = h_i x_i + \eta_i,$$

where the fading coefficient h_i is $\mathcal{CN}(0,1)$ (known at the receiver side) and the additive white noise η_i is $\mathcal{CN}(0, 2\sigma^2)$. The channel likelihood is

$$p(y_i | x_i, h_i) = \frac{1}{2\pi\sigma^2} \exp(-\frac{|y_i - h_i x_i|^2}{2\sigma^2}).$$

Erasure and Fading

By restricting $\alpha_i = |h_i|$ to $\{0, +\infty\}$, the fading channel becomes an erasure channel. $P(\alpha_i = 0) = \epsilon$ and $P(\alpha_i = +\infty) = 1 - \epsilon$.

Usually $x_i = f(c_i)$, where $f : F_q \rightarrow \mathbb{Z}^2$ is a mapping that converts the finite field elements into complex symbols. This mapping is known as a QAM modulation.

Coding for fading channels (1)

The fading channel is defined by the input-output relation

$$y_i = h_i x_i + \eta_i,$$

where the fading coefficient h_i is $\mathcal{CN}(0,1)$ (known at the receiver side) and the additive white noise η_i is $\mathcal{CN}(0, 2\sigma^2)$. The channel likelihood is

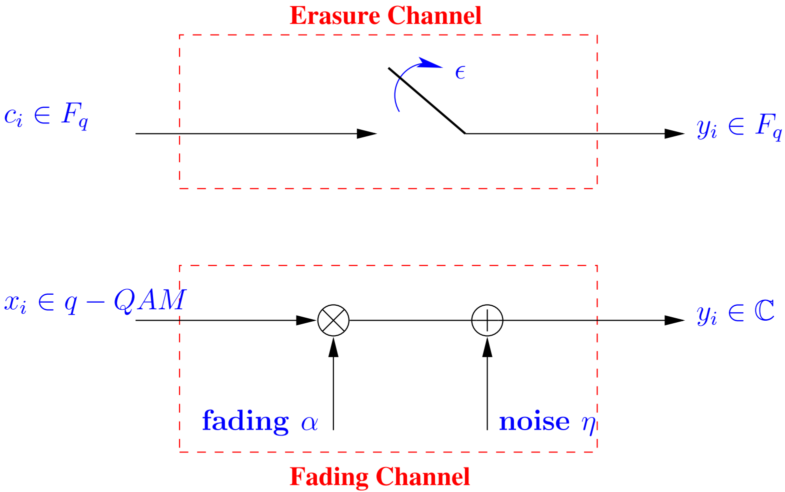
$$p(y_i|x_i, h_i) = \frac{1}{2\pi\sigma^2} \exp(-\frac{|y_i - h_i x_i|^2}{2\sigma^2}).$$

Erasure and Fading

By restricting $\alpha_i = |h_i|$ to $\{0, +\infty\}$, the fading channel becomes an erasure channel. $P(\alpha_i = 0) = \epsilon$ and $P(\alpha_i = +\infty) = 1 - \epsilon$.

Usually $x_i = f(c_i)$, where $f : F_q \to \mathbb{Z}^2$ is a mapping that converts the finite field elements into complex symbols. This mapping is known as a QAM modulation.

Coding for fading channels (2)



Coding for fading channels (3)

- The notions of iid and non-ergodic fading channels are directly derived from those defining the BEC given on slides 1 and 2.
- Definition 1 of diversity on BEC and Theorem 2 on the code design criterion for BEC are still valid on a Rayleigh fading channel. The erasure probability ϵ is replaced by the signal-to-noise ratio

$$\gamma = \frac{\mathcal{E}[|x_i|^2]}{\mathcal{E}[|\eta_i|^2]}.$$

- The word error probability at the decoder output is denoted by P_e. The full-diversity behavior $P_e \propto \epsilon^L$ becomes $P_e \propto 1/\gamma^L$.

Definition 3: Diversity on fading channels

The diversity order d attained by a code $\mathcal{C}$ is defined as

$$d = -\lim_{\gamma \to +\infty} \frac{\log P_e}{\log \gamma}.$$

Coding for fading channels (3)

- The notions of iid and non-ergodic fading channels are directly derived from those defining the BEC given on slides 1 and 2.
- Definition 1 of diversity on BEC and Theorem 2 on the code design criterion for BEC are still valid on a Rayleigh fading channel. The erasure probability ϵ is replaced by the signal-to-noise ratio

$$\gamma = \frac{\mathcal{E}[|x_i|^2]}{\mathcal{E}[|\eta_i|^2]}.$$

- The word error probability at the decoder output is denoted by P_e. The full-diversity behavior $P_e \propto \epsilon^L$ becomes $P_e \propto 1/\gamma^L$.

Definition 3: Diversity on fading channels

The diversity order d attained by a code $\mathcal{C}$ is defined as

$$d = -\lim_{\gamma \to +\infty} \frac{\log P_e}{\log \gamma}.$$

Coding for fading channels (4)

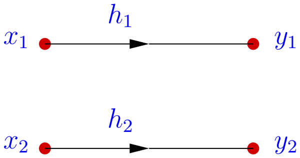
- Let $x = f(c)$ and $\hat{x} = f(\hat{c})$ be two codewords. The partial Hamming weights ω_ℓ defined for the non-ergodic BEC can now be extended to the non-ergodic fading channel as follows:
  - Divide a codeword into L blocks, each block containing N/L components.
  - The quantity ω_ℓ is the weight of the ℓth block in $x - \hat{x}$, i.e., the number of non-zero components in the difference.
- Example: $N = 4$, $q = 4$, and $L = 2$. Take $x = (+1, +1, -3, -3)$, if $\hat{x} = (-1, -1, -3, -3)$ then $\omega_1 = 2$ and $\omega_2 = 0$. If $\hat{x} = (-1, +1, +3, +3)$ then $\omega_1 = 1$ and $\omega_2 = 2$.

Theorem 4: ML design criterion for non-ergodic fading channels

$\mathcal{C}$ is full diversity $(d = L)$ on a non-ergodic fading channel iff, $\forall x, \hat{x} \in f(\mathcal{C})$, $x \neq \hat{x}$, all partial Hamming weights are non zero, i.e., $\omega_\ell \neq 0, \forall \ell$.

The MIMO channel (1)

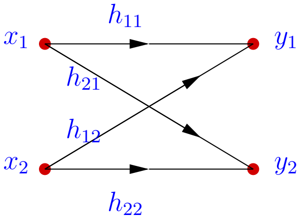
Two parallel single antenna fading channels



$$y = Hx + \eta$$

$$H = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}$$

A multiple antenna (MIMO) fading channel



$$y = Hx + \eta$$

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

The MIMO channel (3)

Mathematical (analytical) models

Physical modeling of a MIMO channel cannot lead to space-time coding design criteria. Mathematical modeling is necessary.

The simplest mathematical model for a $n_t \times n_r$ MIMO channel is

$$y = Hx + \eta,$$

where

- $H = [h_{ij}]$ is a $n_r \times n_t$ matrix with complex circularly symmetric iid gaussian entries of zero mean and unit variance, $h_{ij} \sim \mathcal{CN}(0, 1)$.
- x is a column vector including the n_t transmitted symbols, $x_i \in q - QAM \subset \mathbb{Z}^2$.
- η is a noise vector whose components are complex gaussian and iid, $\eta_i \sim \mathcal{CN}(0, 2\sigma^2)$.

The MIMO channel (5)

- The diversity on a MIMO channel is also given by Definition 3, i.e.,

$$P_e \approx \frac{1}{(g\gamma)^d} \qquad \gamma \gg 1,$$

where g is referred to as the coding gain.

- The MIMO channel as defined in its simplest model on the previous slide has $n_t \times n_r$ degrees of freedom. For a static channel H, i.e., H is constant within a codeword, we have

$$n_r \leq d \leq n_t \times n_r.$$

The lower bound is attained in absence of coding. The ratio d/n_r is known as the transmit diversity.

- The (receive) space dimension for one channel use is n_r. Hence, achieving the maximal diversity $n_t \times n_r$ must require n_t channel transmissions at least. The expression "space-time coding" describes the spreading in both space and time of codes designed for MIMO channels.

The MIMO channel (5)

- The diversity on a MIMO channel is also given by Definition 3, i.e.,

$$P_e \approx \frac{1}{(g\gamma)^d} \qquad \gamma \gg 1,$$

  where g is referred to as the coding gain.

- The MIMO channel as defined in its simplest model on the previous slide has $n_t \times n_r$ degrees of freedom. For a static channel H, i.e., H is constant within a codeword, we have

$$n_r \leq d \leq n_t \times n_r.$$

  The lower bound is attained in absence of coding. The ratio d/n_r is known as the transmit diversity.

- The (receive) space dimension for one channel use is n_r. Hence, achieving the maximal diversity $n_t \times n_r$ must require n_t channel transmissions at least. The expression "space-time coding" describes the spreading in both space and time of codes designed for MIMO channels.

The MIMO channel (6)

- The main objective of space-time coding is to build easily encodable and decodable codes that maximize both coding gain g and diversity d for a given information rate R.
- Let $R_c = K/N$ denotes the coding rate of $\mathcal{C}$. Then, the information rate expressed in bits per channel use (bpcu) is given by

$$R = n_t \times R_c \times \log_2(q) \qquad bpcu$$

- Distributing the components of a space-time code over the n_t transmit antennas is referred to as spatial multiplexing.
- When compared to an uncoded single antenna system, the MIMO information rate is multiplied by a factor $\mu = n_t R_c$. This increase in information rate is called multiplexing gain.
- Another (asymptotic) information theoretical definition of μ is given by

$$\mu = \lim_{\gamma \to +\infty} \frac{R(\gamma)}{\log_2 \gamma},$$

which is equivalent to the assumption $R = \mu \log_2 \gamma + O(1)$.

The MIMO channel (6)

- The main objective of space-time coding is to build easily encodable and decodable codes that maximize both coding gain g and diversity d for a given information rate R.
- Let $R_c = K/N$ denotes the coding rate of $\mathcal{C}$. Then, the information rate expressed in bits per channel use (bpcu) is given by

$$R = n_t \times R_c \times \log_2(q) \qquad bpcu$$

- Distributing the components of a space-time code over the n_t transmit antennas is referred to as spatial multiplexing.
- When compared to an uncoded single antenna system, the MIMO information rate is multiplied by a factor $\mu = n_t R_c$. This increase in information rate is called multiplexing gain.
- Another (asymptotic) information theoretical definition of μ is given by

$$\mu = \lim_{\gamma \to +\infty} \frac{R(\gamma)}{\log_2 \gamma},$$

which is equivalent to the assumption $R = \mu \log_2 \gamma + O(1)$.

Capacity of MIMO channels (1)

A single antenna (also single user) ideal channel without fading, known as AWGN channel, is described by

$$y = x + \eta \qquad x, y, \eta \in \mathbb{C}$$

Capacity is given by the famous formula (Shannon 1948)

$$C_{AWGN} = \log_2(1+\gamma).$$

- Recall that it is possible to find a code $\mathcal{C}$ such that $P_e \rightarrow 0$ when $N \rightarrow +\infty$ iff $R < C$ (e.g., see Cover & Thomas 1993, Gallager 1968).

At high SNR, on a single antenna AWGN channel, doubling the transmitted energy increases the capacity by one bit only

$$C_{AWGN}(2\gamma) \approx \log_2(2\gamma) \approx 1 + C_{AWGN}(\gamma) \qquad bpcu.$$

Capacity of MIMO channels (1)

A single antenna (also single user) ideal channel without fading, known as AWGN channel, is described by

$$y = x + \eta \qquad x, y, \eta \in \mathbb{C}$$

Capacity is given by the famous formula (Shannon 1948)

$$C_{AWGN} = \log_2(1+\gamma).$$

- Recall that it is possible to find a code $\mathcal{C}$ such that $P_e \rightarrow 0$ when $N \rightarrow +\infty$ iff $R < C$ (e.g., see Cover & Thomas 1993, Gallager 1968).

At high SNR, on a single antenna AWGN channel, doubling the transmitted energy increases the capacity by one bit only

$$C_{AWGN}(2\gamma) \approx \log_2(2\gamma) \approx 1 + C_{AWGN}(\gamma) \qquad bpcu.$$

Capacity of MIMO channels (2)

On a fading channel with L parallel branches (diagonal MIMO), where the L fading coefficients $H = \text{Diag}(h_1, \ldots, h_L)$ are only known at the receiver side and the SNR per branch is $\gamma_i = \gamma/L$, we have

$$C(H) = \sum_{i=1}^{L} \log_2(1+|h_i|^2\gamma_i) = \log_2\left(\prod_{i=1}^{L}(1+|h_i|^2\frac{\gamma}{L})\right) = \log_2 \det\left(I_L + \frac{\gamma}{L} H H^\dagger\right).$$

The above result is still valid for any $n_t \times n_r$ MIMO channel (Telatar 1995) where the exact proof is based on the fact that a circularly symmetric complex gaussian vector with covariance matrix Γ yields a maximal differential entropy equal to $\log_2 \det(\pi e \Gamma)$.

For the MIMO channel $y = Hx + \eta$, $x \in \mathbb{C}^{n_t}, y \in \mathbb{C}^{n_r}$, the conditional capacity $C(H)$ is defined as the average mutual information $I(x;y)$ between x and y for a given channel matrix H. The capacity $C(H)$ is obtained by assuming that the input is gaussian with covariance matrix $Q = \frac{\gamma}{n_t} I_{n_t}$ (uniform gaussian input).

Capacity of MIMO channels (2)

On a fading channel with L parallel branches (diagonal MIMO), where the L fading coefficients $H = \mathrm{Diag}(h_1, \ldots, h_L)$ are only known at the receiver side and the SNR per branch is $\gamma_i = \gamma/L$, we have

$$C(H) = \sum_{i=1}^{L} \log_2(1+|h_i|^2\gamma_i) = \log_2\left(\prod_{i=1}^{L}(1+|h_i|^2\frac{\gamma}{L})\right) = \log_2 \det\left(I_L + \frac{\gamma}{L}HH^\dagger\right).$$

The above result is still valid for any $n_t \times n_r$ MIMO channel (Telatar 1995) where the exact proof is based on the fact that a circularly symmetric complex gaussian vector with covariance matrix Γ yields a maximal differential entropy equal to $\log_2 \det(\pi e \Gamma)$.
For the MIMO channel $y = Hx + \eta$, $x \in \mathbb{C}^{n_t}, y \in \mathbb{C}^{n_r}$, the conditional capacity $C(H)$ is defined as the average mutual information $I(x;y)$ between x and y for a given channel matrix H. The capacity $C(H)$ is obtained by assuming that the input is gaussian with covariance matrix $Q = \frac{\gamma}{n_t}I_{n_t}$ (uniform gaussian input).

Capacity of MIMO channels (3)

Assume a uniform gaussian input and H only known at the receiver side.

Definition 5: Capacity of an ergodic iid Rayleigh MIMO channel

The ergodic capacity of the $n_t \times n_r$ MIMO channel is

$$C(\gamma, n_t, n_r) = \mathcal{E}_H \left[\log_2 \det \left(I_{n_r} + \frac{\gamma}{n_t} H H^\dagger \right) \right].$$

At high SNR ($\gamma \gg 1$), It can be shown that (Foschini 1996)

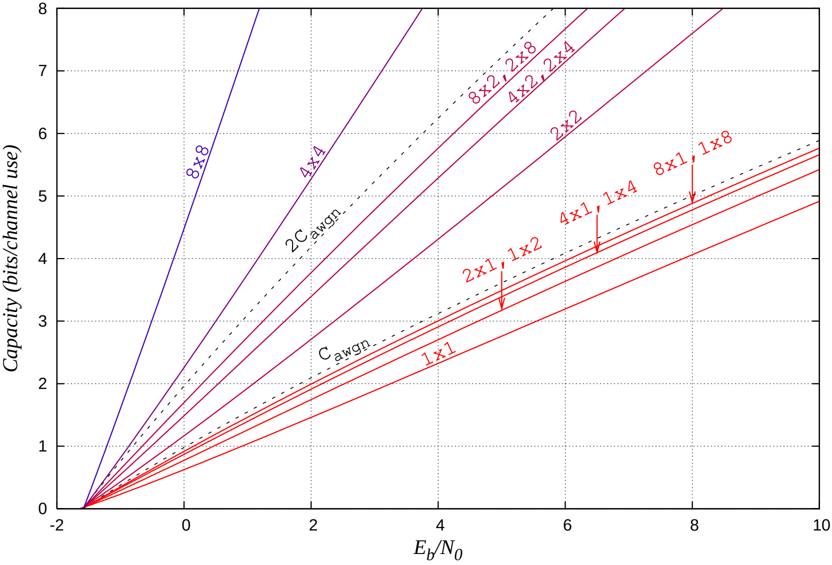
$$C(\gamma, n_t, n_r) = \min(n_t, n_r) \log_2(\gamma) + O(1).$$

Then, for a symmetric channel $n_t = n_r = n$, we have

$$C(2\gamma, n) \approx n \log_2(2\gamma) \approx n + C(\gamma, n)$$

Doubling the transmitted energy increases the capacity by n bits. In the next slide, the ergodic capacity is plotted versus $E_b/N_0 = n_r\gamma / C(\gamma, n_t, n_r)$, recall that $Q = \mathcal{E}[xx^\dagger] = \frac{\gamma}{n_t} I_{n_t}$.

Capacity of MIMO channels (4)



Outage probability (2)

- For general non-ergodic fading channels, a key idea is to consider the mutual information $I(x;y|H)$ betweeen the channel input and output as a random variable. Each time we pick up a random instance of the channel H, it renders a new instantaneous value of $I(x;y|H)$. For a given information rate R to be transmitted, an information theoretical limit on the word error probability is given by (Ozarow, Shamai, Wyner 1994, see also Biglieri, Proakis, Shamai 1998)

$$P_{out} = P\left(\ I(x;y|H)\ <\ R\ \right)$$

- A similar approach is used for non-ergodic MIMO channels (Telatar 1999, Foschini & Gans 1998). There is an outage each time $C(H,Q)$ is less than the targeted information rate. Here, the average mutual information between x and $y = Hx + \eta$ is indexed by the MIMO channel matrix H and the covariance $Q = \mathcal{E}[xx^\dagger]$. The outage probability is

$$P_{out} = P\left(\ \log\det\left(I_{n_r} + HQH^\dagger\right)\ <\ R\ \right)$$

It is conjectured that P_{out} is minimized by using a uniform power allocation over a subset of the transmit antennas.

Outage probability (2)

- For general non-ergodic fading channels, a key idea is to consider the mutual information $I(x;y|H)$ betweeen the channel input and output as a random variable. Each time we pick up a random instance of the channel H, it renders a new instantaneous value of $I(x;y|H)$. For a given information rate R to be transmitted, an information theoretical limit on the word error probability is given by (Ozarow, Shamai, Wyner 1994, see also Biglieri, Proakis, Shamai 1998)

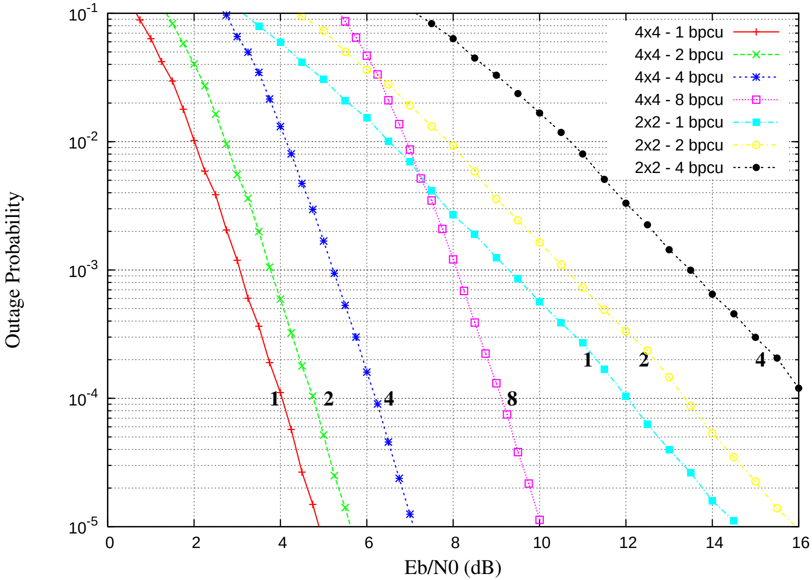
$$P_{out} = P\left(\ I(x;y|H)\ <\ R\ \right)$$

- A similar approach is used for non-ergodic MIMO channels (Telatar 1999, Foschini & Gans 1998). There is an outage each time $C(H,Q)$ is less than the targeted information rate. Here, the average mutual information between x and $y = Hx + \eta$ is indexed by the MIMO channel matrix H and the covariance $Q = \mathcal{E}[xx^\dagger]$. The outage probability is

$$P_{out} = P\left(\ \log\det\left(I_{n_r} + HQH^\dagger\right)\ \ <\ \ R\ \right)$$

It is conjectured that P_{out} is minimized by using a uniform power allocation over a subset of the transmit antennas.

Outage probability (3)



Coding for MIMO channels (1)

Coding for ergodic channels

• The problem of designing error-correcting codes for ergodic MIMO channels (fast fading) is not an issue.
• Any capacity-achieving code designed for the AWGN channel will do the job when transmitted on a channel with an infinite number of degrees of freedom.

Coding for non-ergodic channels

• The problem of designing error-correcting codes for non-ergodic MIMO channels (slow fading) is a difficult issue.
• For example, an outage-approaching code should mix both randomness and determinism in its structure.

Coding for MIMO channels (1)

Coding for ergodic channels

• The problem of designing error-correcting codes for ergodic MIMO channels (fast fading) is not an issue.
• Any capacity-achieving code designed for the AWGN channel will do the job when transmitted on a channel with an infinite number of degrees of freedom.

Coding for non-ergodic channels

• The problem of designing error-correcting codes for non-ergodic MIMO channels (slow fading) is a difficult issue.
• For example, an outage-approaching code should mix both randomness and determinism in its structure.

Coding for MIMO channels (2)

Unless otherwise stated, we restrict the rest of this lecture to static channels (non-ergodic) where a codeword undergoes a unique channel instance ($n_c = 1$). When $n_c > 1$, the diversity is multiplied accordingly and the code design is similar.

- As shown in the first part of this lecture, a rate $1/2$ repetition code $\mathcal{C}[2,1]_q$ can achieve diversity 2 on a block erasure channel with two independent blocks per codeword, $P_e = \epsilon^2$.
- Let us start with a simple example on a 2×1 MIMO channel. What is the equivalent of $\mathcal{C}[2,1]_q$ on a MIMO channel?
  One simple solution: the Alamouti code.

Coding for MIMO channels (3)

Consider the following codeword (Alamouti 1998) written in matrix format

$$x = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix}$$

where $x_i \in q - QAM \subset \mathbb{Z}^2$.

- Row 1 is transmitted on antenna 1. Row 2 is transmitted on antenna 2.
- Two time periods are needed to transmit x on a 2×1 MIMO channel. The rate is $R = n_t \times R_c \times \log_2(q) = \log_2(q)\ \ bpcu$, where $n_t = 2$ and $R_c = 1/2$.
- The channel output is

$$y = Hx + \eta$$

where $H = [h_1\ h_2]$ and $y, \eta \in \mathbb{C}^{1\times 2}$.

Coding for MIMO channels (4)

- Develop the expression of the channel output when Alamouti code is transmitted.

$$
\begin{aligned}
y_1 &= h_1 x_1 + h_2 x_2 + \eta_1 \\
y_2 &= -h_1 x_2^* + h_2 x_1^* + \eta_2
\end{aligned}
$$

- To decode, let us compute

$$
\begin{aligned}
h_1^* y_1 + h_2 y_2^* &= \left(|h_1|^2 + |h_2|^2\right) x_1 + \eta_1' \\
h_2^* y_1 - h_1 y_2^* &= \left(|h_1|^2 + |h_2|^2\right) x_2 + \eta_2'
\end{aligned}
$$

- Transmit diversity 2 is achieved since (see Tse & Viswanath 2005)

$$
P\left((|h_1|^2 + |h_2|^2)\gamma < 1\right) \propto \frac{1}{\gamma^2}
$$

Coding for MIMO channels (4)

- Develop the expression of the channel output when Alamouti code is transmitted.

$$\begin{aligned} y_1 &= h_1 x_1 + h_2 x_2 + \eta_1 \\ y_2 &= -h_1 x_2^* + h_2 x_1^* + \eta_2 \end{aligned}$$

- To decode, let us compute

$$\begin{aligned} h_1^* y_1 + h_2 y_2^* &= \left(|h_1|^2 + |h_2|^2\right) x_1 + \eta_1^{'} \\ h_2^* y_1 - h_1 y_2^* &= \left(|h_1|^2 + |h_2|^2\right) x_2 + \eta_2^{'} \end{aligned}$$

- Transmit diversity 2 is achieved since (see Tse & Viswanath 2005)

$$P\left((|h_1|^2 + |h_2|^2)\gamma \ < \ 1\right) \ \propto \ \frac{1}{\gamma^2}$$

Coding for MIMO channels (5)

- Alamouti code belongs to the family of Orthogonal Space-Time Block codes (OSTBC). The codewords satisfy

$$xx^{\dagger} \propto I_{n_t}$$

- Two examples of OSTBC for $n_t = 3$ and $n_t = 4$ antennas both with rate $R = \frac{3}{4}\log_2(q)$. Four time periods are needed to transmit a codeword.

$$\begin{bmatrix} x_1 & -x_2^* & x_3^* & 0 \\ x_2 & x_1^* & 0 & -x_3^* \\ x_3 & 0 & -x_1^* & x_2^* \end{bmatrix} \qquad \begin{bmatrix} x_1 & 0 & x_2 & -x_3 \\ 0 & x_1 & x_3^* & x_2^* \\ -x_2^* & -x_3 & x_1^* & 0 \\ x_3^* & -x_2 & 0 & x_1^* \end{bmatrix}$$

- OSTBC is an important subclass of linear STBC. For more information See the book by Larsson & Stoica 2003, or the book by Oestges & Clercks 2007.
- Main drawback: they suffer from a weak information rate R (the equivalent embedded R_c is too small).

Coding for MIMO channels (5)

- Alamouti code belongs to the family of Orthogonal Space-Time Block codes (OSTBC). The codewords satisfy

$$xx^\dagger \propto I_{n_t}$$

- Two examples of OSTBC for $n_t = 3$ and $n_t = 4$ antennas both with rate $R = \frac{3}{4}\log_2(q)$. Four time periods are needed to transmit a codeword.

$$\begin{bmatrix} x_1 & -x_2^* & x_3^* & 0 \\ x_2 & x_1^* & 0 & -x_3^* \\ x_3 & 0 & -x_1^* & x_2^* \end{bmatrix} \qquad \begin{bmatrix} x_1 & 0 & x_2 & -x_3 \\ 0 & x_1 & x_3^* & x_2^* \\ -x_2^* & -x_3 & x_1^* & 0 \\ x_3^* & -x_2 & 0 & x_1^* \end{bmatrix}$$

- OSTBC is an important subclass of linear STBC. For more information See the book by Larsson & Stoica 2003, or the book by Oestges & Clercks 2007.
- Main drawback: they suffer from a weak information rate R (the equivalent embedded R_c is too small).

Coding for MIMO channels (6)

- Consider a linear code $\mathcal{C}[Nn_t, K]_q$ of rate $R_c = Nn_t/K$. Write a codeword as

$$c = \begin{bmatrix} c_1^1 & c_2^1 & \dots & c_N^1 \\ \vdots & \vdots & & \vdots \\ c_1^{n_t} & c_2^{n_t} & \dots & c_N^{n_t} \end{bmatrix}$$

- Using the mapping $f : F_q \rightarrow \mathbb{Z}^2$, transmit $x = f(c)$ on $n_t \times n_r$ channel.
- It can be shown (e.g., see El Gamal & Hammons 2003) that the pairwise error probability is upper bounded as (ML decoder assumed)

$$P(c \rightarrow c') \quad \leq \quad \left(\frac{1}{\prod_{i=1}^{t}(1 + \lambda_i \gamma / 4n_t)} \right)^{n_r} \quad \leq \quad \left(\frac{g\gamma}{4n_t} \right)^{-tn_r}$$

where $t = rank(f(c) - f(c'))$, the coding gain is $g = (\lambda_1 \lambda_2 \cdots \lambda_t)^{1/t}$, and $\{\lambda_i\}$ are the eigen values of $[f(c) - f(c')][f(c) - f(c')]^\dagger$.

Coding for MIMO channels (6)

- Consider a linear code $\mathcal{C}[Nn_t, K]_q$ of rate $R_c = Nn_t/K$. Write a codeword as

$$c = \begin{bmatrix} c_1^1 & c_2^1 & \dots & c_N^1 \\ \vdots & \vdots & & \vdots \\ c_1^{n_t} & c_2^{n_t} & \dots & c_N^{n_t} \end{bmatrix}$$

- Using the mapping $f : F_q \rightarrow \mathbb{Z}^2$, transmit $x = f(c)$ on $n_t \times n_r$ channel.
- It can be shown (e.g., see El Gamal & Hammons 2003) that the pairwise error probability is upper bounded as (ML decoder assumed)

$$P(c \rightarrow c') \quad \leq \quad \left(\frac{1}{\prod_{i=1}^{t}(1 + \lambda_i \gamma / 4n_t)} \right)^{n_r} \quad \leq \quad \left(\frac{g\gamma}{4n_t} \right)^{-tn_r}$$

where $t = rank(f(c) - f(c'))$, the coding gain is $g = (\lambda_1 \lambda_2 \cdots \lambda_t)^{1/t}$, and $\{\lambda_i\}$ are the eigen values of $[f(c) - f(c')][f(c) - f(c')]^\dagger$.

Coding for MIMO channels (7)

From the expression of the pairwise error probability, we can state (Guey, Fitz, Bell, & Kuo 1996, and Tarokh, Seshadri, & Calderbank 1998)

Design criterion for static MIMO

Under ML decoding, a space-time code should satisfy (over all pairs of distinct codewords c and c')

- Rank: Maximize the transmit diversity $t = rank(f(c) - f(c'))$.
- Product distance: Maximize the coding gain $g = (\lambda_1 \lambda_2 \cdots \lambda_t)^{1/t}$.

- The above design criterion cannot guarantee the construction of outage-achieving codes.
- The above design criterion cannot be used to build iteratively decodable graph codes (e.g., LDPC codes) for MIMO channels.
- Nevertheless, it has been used to successfully build space-time block codes (not including an error-correcting code $\mathcal{C}$) that guarantee excellent performance for uncoded q-QAM modulations.

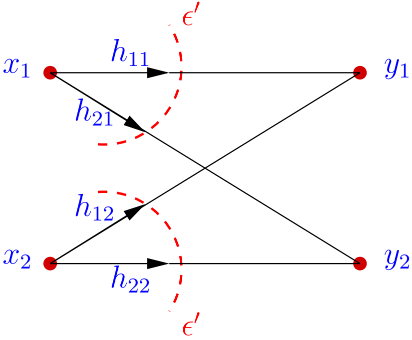
Coding for MIMO channels (8)

Let us study an example of full-diversity LDPC coding for a 2×2 MIMO channel.

- Each transmit antenna behaves like a channel state. The state generates an erasure with a probability ϵ'

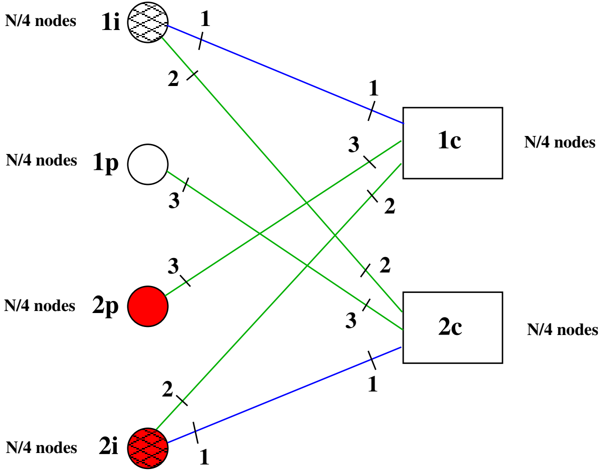
$$\epsilon' = P\left((|h_{11}|^2 + |h_{21}|^2)\gamma \ < \ 1\right) \quad \propto \quad \frac{1}{\gamma^2}$$

- Our aim is to achieve $(\epsilon')^2$, i.e., $P_e \propto \frac{1}{\gamma^4}$



Coding for MIMO channels (9)

Rate-1/2 Full-diversity root-LDPC code for a 2-state block fading channel (Boutros, Guillén i Fàbregas, Biglieri, & Zémor 2007)



White-colored bits: antenna 1 Red-colored bits: antenna 2

Coding for MIMO channels (10)

The parity-check matrix of the root-LDPC code has the following structure (mixture of randomness and determinism).

$$H = \begin{array}{cc} \begin{array}{cccc} 1i & 1p & 2i & 2p \end{array} & \\ \left[\begin{array}{c:c:c:c} \begin{matrix}1&&&\\&1&&\\&&1&\\&&&1\end{matrix} & \mathbf{0} & H_{2i} & H_{2p} \\ \hdashline H_{1i} & H_{1p} & \begin{matrix}1&&&\\&1&&\\&&1&\\&&&1\end{matrix} & \mathbf{0} \end{array}\right] & \begin{array}{c} 1c \\ 2c \end{array} \end{array}$$

Theorem 6: threshold in absence of fading

On a gaussian channel, under iterative decoding, a $(\lambda(x), \rho(x))$ root-LDPC code has the same decoding threshold as a random $(\lambda(x), \rho(x))$ LDPC code.

Theorem 7: full-diversity for a rate-$1/2$ root-LDPC

Consider a static $2 \times n_r$ MIMO channel. Under iterative decoding, a $(\lambda(x), \rho(x))$ root-LDPC code achieves full state diversity $(\epsilon')^2$, i.e., $P_e \propto 1/(\gamma)^{2n_r}$.

Summary

- The erasure channel can be a starting point for the study of more complex channels such as the MIMO channel.
- The MIMO channel offers a higher capacity (higher data rates) than a single antenna medium.
- Several coding techniques are well established for space-time coding, practical applications are slowed down by the decoding complexity.