# Lecture Note on

# Wireless Communications

Heung-No Lee

# Wireless Communications

Module-1



## Agenda

- Course Schedule
- Signal Space Representation & Optimal Receiver

## E-mail

- My e-mail is
  heungno@gist.ac.kr
- I will use e-mail for sending lecture notes and special announcements.



## Course Information

- Class hours: 9:00-10:30 am Monday, Wednesday
- Lecture room: B201
- Office hours:
  - 10:30am ~ 11:30am Monday,
  - 10:00am ~ 11:00am Tuesday.
  - Or make an appointment via e-mail.

## Grade Distribution

- Two exams (Midterm#1: 20%, Final: 30%)
- Homework + Homework Grading (20%)
- Term Project (30%)
  - Binary modulation over AWGN channel simulation
  - Add LDPC encode/decode simulation
  - Add variations such as
    - Wireless network codes
    - Slepian Wolf distributed source coding
    - Compressive sensing



## Tentative Schedule

| | | |
|---|---|---|
| 1st week | General overview (Shannon's 1948 paper) | |
| 2nd week | Optimal Transceiver | |
| 3rd week | Gallager's Channel Coding Theorem | |
| 4th week | Gallager's Channel Coding Theorem | |
| 5th week | LDPC codes and probabilistic decoders | |
| 6th week | Multipath fading channels/Diversity systems | |
| 7th week | MIMO capacity theorems | |
| 8th week | MIMO transceivers | Midterm 1 |
| 9th week | Design of LDPC and space-time codes and receivers | |
| 10th week | Performance evaluation of MIMO transceivers | |
| 11th week | Multi-user capacity/multi-user MIMO receivers | |
| 12th week | Design of pre-coding MIMO signals | |
| 13th week | Network codes | |
| 14th week | Wireless network codes | |
| 15th week | Overview | |
| 16th week | Final project + Final Exam | Final project due |

## Homework, Class-Project Policies

- Discussion and exchange of ideas are strongly encouraged.
- You may submit your homework and project reports as a team of two persons.
- On each homework and class project set, a reviewer team will be assigned (will take turns).
- The job of each reviewer team is to
  - grade homework/project sets,
  - type up the best homework solution(rec. WORD with Mathtype),
  - get an approval of the solution manual from me, and
  - distribute the graded homework and solution to the students within a week.



## Scope of this course

- In this course, we will learn wireless and MIMO networks with help of
  - Information Theory
  - Digital Communications Theory
  - Channel Coding Theory
- What's relevant are
  - Complexity of the system (Is the system implementable?)
  - Performance of the system (Probability of decision errors)
  - How far is it from the theoretical limit?

## Scope of this course (2)

- Learn to apply the estimation/detection theory to communications problems.
- Learn to simulate communications systems for the purpose of evaluating a communications system.
- Be able to analyze the obtained simulation data and predict the performance of a given system, and provide a better design.
- Once we know how to predict/evaluate the performance of a communications system, we will use these knowledge and tool sets to design a better performing communications system.
- I say this is the way how the communications theory has been evolved.



## Text Books

- Textbook: Proakis/Salehi, Digital Communications, 5th Edition, McGraw-Hill.

- Reference-1: Robert Gallager, Information Theory and Reliable Communication, John Wiley & Sons, Inc. New York, NY, USA, 1968. ISBN:0471290483
- Reference-2: David Tse and P. Viswanath, Fundamentals of Wireless Communication, Cambridge Press, 2005.

  ISBN: 0521845270
- Reference-3: J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, Prospect Heights , Illinois, Waveland Press, 1990.

## About Proakis/Salehi

- The book is like an encyclopedia of communications theory.
  - Covers a lot of topics
  - I certainly will not aim to cover all of them
  - I'll go over certain topics quickly to help your reading.
- Some homework problems will be taken from the book.



## Interaction with other related areas

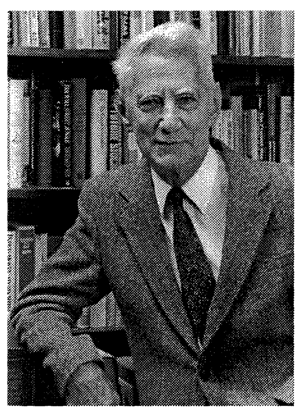
- Optimization
- Signal and Image Processing
- Estimation/Detection Theory
- Pattern Recognition
- Neural Network
- Artificial Intelligence
- Bio-informatics

- Now, let's begin…

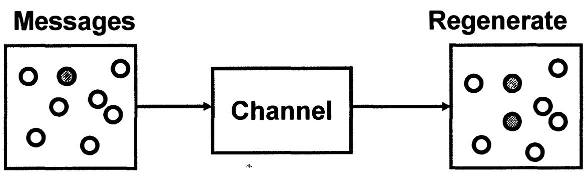


## Claude E. Shannon (1916 -- 2001)

- Math/EE Bachelor from UMich (1936)
- MSEE and Math Ph.D. from MIT (1940)
- A landmark paper "Mathematical Theory of Communications" (1948)
  - Founder of Information Theory
  - Fundamental limits on communications
  - Information quantified as a logarithmic measure
- For more info on him, make a visit to http://www.bell-labs.com/news/2001/february/26/1.html



## Shannon's Perspective on Communications



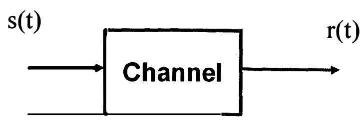

- Communications: Transfer of information from a source to a receiver
- Messages (information) can have *meaning*; but they are irrelevant for the design of communications system.
- What's important then?
  - *A message is selected from a set of all possible messages and transmitted, and regenerated at the receiver*
  - The size of the message set is the amount of information
- ***The capacity of a channel is the maximum size of message set that can be transferred over the channel and can be regenerated almost error-free at the receiver.***



## Digital Communication

- It is to send a message index $m$ (out of $M$ total) over *the channel*
  - for the duration of time $T$, and
  - have an expectation that the same index $m$ can be recovered almost error-free at the receiver.
- Transmission rate $R = \log_2(M)/T$ [bits/sec]
- *If* $R < C$, *then* almost error-free recovery can be achieved.
- We need to find a set of $M$ waveforms to interface the channel.
  - An analog (physical) waveform shall be chosen to carry the messages. Why?
  - We may choose orthogonal waveforms
    - Pulse-position, frequency-position (OFDM), sin(x)/x, …, or any other orthogonal signal set
  - We need to find out how we can choose them.

## Additive Noise Channel




- Let's consider a simple channel

  $$r(t) = s(t) + n(t)$$

  - received signal = signal + noise



## Digital Representation

- Note that the simple ANC is based on continuous signals and noise
- We aim to replace the ANC with digits and vectors. Why?
  - Easier to deal with (digits rather than continuous waveforms)
    - Computer simulations without loss of information
    - Easier to do analysis
    - Easier to design
- How?
  - Via the use of *vector space* idea.

## Hilbert Space

- Vectors of finite length can be added together and multiplied by scalars. So they form a *vector space* (Hilbert space).
  - Norm exists ( $\sum_k |v_k|^2 < \infty$ or $\int_t |v(t)|^2\, dt < \infty$)
  - Schwartz inequality holds (
    - $v_1 \cdot v_2| \le ||v_1||\ ||v_2||$
    - $|\int_t f(t)g^*(t)\, dt| \le (\int_t |f(x)|^2\, dt)^{1/2} (\int_t |g(t)|^2\, dt)^{1/2})$
- Extension of the ideas of *length* and *inner product* from vector space to signal space.
- Approximation of a function with a series of (finite number of) orthonormal functions.
  - the Karhunen-Loeve Expansion (pg. 76, Proakis/Salehi)



## Hilbert Space

- Can we extend the idea of the Euclidean vector space to functions and signals?
- Yes. But how?
  - Use a set of orthogonal basis (e.g. Fourier, Harr, …)
  - Infinite dimensional the space becomes.
  - Let's consider only finite length vectors
    - (1, ½, 1/3, …) included; but (1, 1, 1, 1, …) excluded.
- Vectors of finite length is closed under addition and multiplication by a scalar:
  - Triangular inequality: $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$
    - But TI can be proved by Sch. Ineq. (Use $\mathbf{z}$ = x+y and consider $\mathbf{z}^T\mathbf{z}$).

## Transformation

- A set of *linearly independent* vectors which *spans* a vector space is called a *basis.*
- A coordinate system can be represented by a basis.
- There can be infinitely many coordinate systems in a *vector space*.
- Any vector with a finite norm can be represented by a linear combination of basis vectors.
- Any vector (a signal) with a finite norm can be represented by any coordinate system.
- Change of basis to represent a vector can be performed. This is called *transformation.*



## Transformation (2)

- Example of transformation of a vector in different coordinate system
  - Representation of vectors in Cartesian coordinate systems
- Usually, we choose an orthonormal set of vectors as a basis
  - Norm of each basis vector is 1.
  - Inner product between any pair of basis vectors is 0.

# Transformation (3)

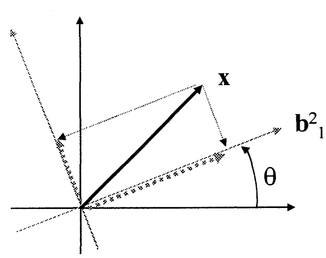
- A vector (a signal) represented by $\{\mathbf{b}^1{}_k\}$ can be transformed into other coordinate system, or be represented by $\{\mathbf{b}^2{}_k\}$.
  - Let $\{\mathbf{b}^1{}_k\}$ and $\{\mathbf{b}^2{}_k\}$ be two different bases (*orthonormal*)
    - Let $\{\mathbf{b}^1{}_k\} = \{(1, 0, 0, \ldots), (0, 1, 0, \ldots), (0, 0, 1, \ldots), \ldots\}$ be Cartesian
  - Representation of a vector $\mathbf{x} = (x_1, x_2, \ldots)$ with the 1st basis is
    $\mathbf{x} = \sum_k (\mathbf{x} \cdot \mathbf{b}^1{}_k)\, \mathbf{b}^1{}_k = \sum_k v_k\, \mathbf{b}^1{}_k$
  - Representation of the same vector with the 2nd basis is
    $\mathbf{x} = \sum_k (\mathbf{x} \cdot \mathbf{b}^2{}_k)\, \mathbf{b}^2{}_k$

projection k-th basis vector



# Cartesian Coordinate

- $\mathbf{x} = (x_1, x_2) = (1, 1)$
- **Basis-1 =** $\{\mathbf{b}^1{}_1, \mathbf{b}^1{}_2\}$
  - $\mathbf{b}^1{}_1 = (1, 0)$
  - $\mathbf{b}^1{}_2 = (0, 1)$
  - $\mathbf{x} = x_1\, \mathbf{b}^1{}_1 + x_2\, \mathbf{b}^1{}_2$
- **Basis-2 =** $\{\mathbf{b}^2{}_1, \mathbf{b}^2{}_2\}$
  - $\mathbf{b}^2{}_1 = (\cos(\theta), \sin(\theta))$
  - $\mathbf{b}^2{}_2 = (-\sin(\theta), \cos(\theta))$
  - $\mathbf{x} = (\mathbf{x} \cdot \mathbf{b}^2{}_1)\, \mathbf{b}^2{}_1 + (\mathbf{x} \cdot \mathbf{b}^2{}_2)\, \mathbf{b}^2{}_2$
- Transform (changing the basis from 1 to 2)
- Inverse-Transform (change back to 1 from 2)




**Note that x stays the same, throughout!**

## Orthogonal Signals

- Functions $\psi_n(t)$ and $\psi_m(t)$, a < $t$ < b, are orthogonal when
  $\int_a^b \psi_n(t)\, \psi^*_m(t)\, dt = 0$, for $m \neq n$
  - Inner-product is zero
- Self inner-product is the energy $K_n$ of the function
  $\int_a^b \psi_n(t)\, \psi^*_n(t)\, dt = K_n$
- A collection of orthogonal functions $\{\psi_n(t)\}$ is said to be an orthogonal set when the collection satisfies the following:
  $\int_a^b \psi_n(t)\, \psi^*_m(t)\, dt = K_n\, \delta_{n,m}$

Kronecker delta func
cf) Dirac delta function



## Geometric View of Signals and Noise

- We aim to represent signals and noise with orthonormal signals.
- Suppose we have a collection of signals,
  $\{\psi_j(t), j \in 1,2, \ldots\}$, $0 \leq t \leq T$, *orthonormal* to each other.
  - *Orthonormality*: $\int_0^T \psi_j(t)\, \psi^*_k(t)\, dt = \delta(k - j)$
    where δ(k-j) is the Kronecker's delta function.
- The set of *orthonormal* signals can form a vector space.
- We can use the first $N$ signals $\{\psi_j(t),\ j \in 1,2, \ldots, N\}$ as a basis for the signal space.

## Representation of signals using $\{\psi_j(t),\ j \in 1,2, \ldots, N\}$

- A basis can be used to represent ANY signal in the space
  $s(t) = \sum_{j=1}^{N} s_j\ \psi_j(t),\ 0 \le t \le T$
  where $s_j = \int_0^T s(t)\ \psi^*_j(t)\ dt$, for $1 \le j \le N$.
- Finally, a continuous signal $s(t)$ can be written as an $N$-tuple vector, i.e., $\mathbf{s} = (s_1, s_2, \ldots, s_N)$.



## Energy Conservation

- The signal energy E
  $E = \int_0^T |s(t)|^2\ dt = \int_0^T [\sum_j s_j\ \psi_j(t)]\ [\sum_q s_q \psi_q(t)]^*\ dt$
  $= \sum_j \sum_q s_j\ s^*_q \int_0^T \psi_j(t)\ \psi^*_q(t)\ dt$
  $= \sum_{j=1}^{N} |s_j|^2$

## Representation of noise using $\{\psi_j(t)\}$

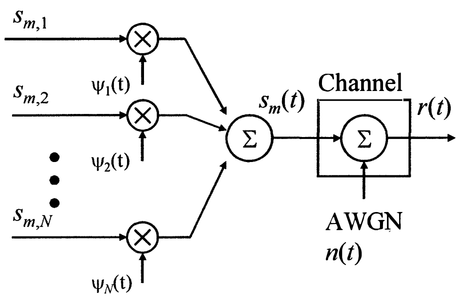
- Compute the projection onto each basis signal
  $n_j = \int_0^T n(t)\, \psi^*_j(t)\, dt$
- Then, we note the noise have two parts
  $n(t) = n_i(t) + n_o(t)$
  where $n_i(t) = \sum_{j=1}^{N} n_j\, \psi_j(t)$
  (the noise which resides inside the signal-space)
  $n_o(\mathrm{t}) = \sum_{j=N+1}^{\infty} n_j\, \psi_j(t)$
  (the noise orthogonal to, and thus resides outside, the signal space)
- Then, we can say, $n(t) \Rightarrow \mathbf{n} = (n_1, n_2, \ldots, n_N)$.
- The continuous noise can be represented by an $N$-tuple random noise vector $\mathbf{n}$.



## Additive White Gaussian Noise $n(t)$

- Suppose $n(t)$ is Gaussian white noise with zero mean and double sided PSD $N_0/2$.
- Then, each projection coefficient $n_j$ is a random variable
  $n_j = \int_0^T n(t)\, \psi^*_j(t)\, dt$
- What kind of r.v. is $n_j$?
  - Linear combination of Gaussian random variables is Gaussian.
  - Mean and covariance?
    - Mean = 0
    - Consider $E\{n_j n^*_k\} = E\{\int_0^T n(t_1)\, \psi^*_j(t_1)\, dt_1 \int_0^T n^*(t_2)\, \psi_k(t_2)\, dt_2\}$
      $= \int_0^T \int_0^T E\{n(t_1)\, n^*(t_2)\}\, \psi^*_j(t_1)\, \psi_k(t_2)\, dt_2\, dt_1$
      $= \int_0^T \int_0^T N_0/2\, \delta(t_2 - t_1)\, \psi^*_j(t_1)\, \psi_k(t_2)\, dt_2\, dt_1$
      $= N_0/2 \int_0^T \psi^*_j(t_2)\, \psi_k(t_2)\, dt_2 = N_0/2\, \delta_{k,j}$
    - Variance is $N_0/2$
  - Mutually uncorrelated Gaussian r.v.s

## Representation of signals and noise




- From the development so far, we can say that
  $r(t) = s(t) + n(t)$
  $\Rightarrow \mathbf{r} = \mathbf{s} + \mathbf{n}$
- Thus, the figure on the right makes sense!!!
  - Let's check



## How to choose the basis and the signal set

- It depends on your resource and what you want
- Time division
- Frequency division
- Code-division
- Let's see some examples

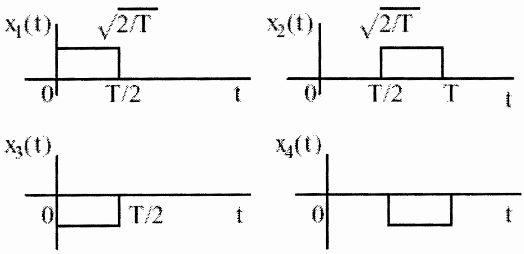
# Examples

(Signal Space Representation)

❖ Time Division Examples

- It's simply because it's easier to illustrate the point
- But, certainly you should not be limited by these simple cases

❖ Example 1:




$\phi_1(t) = x_1(t)$
$\phi_2(t) = x_2(t)$
Dimension = 2
$x_1 = (1\ 0)^t$
$x_2 = (0\ 1)^t$
$x_3 = (-1\ 0)^t$
$x_4 = (0\ -1)^t$



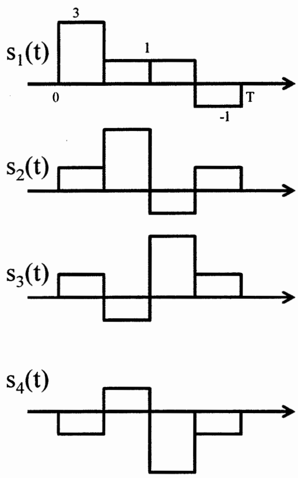
# Signal Space Representation

❖ We may attempt to draw signals in the signal space (it's doable up to three dimension)

- Signal Space Plot, we call it.
- When you draw, treat each basis vector as a coordinate.

❖ Can you compare *M* and *N*?

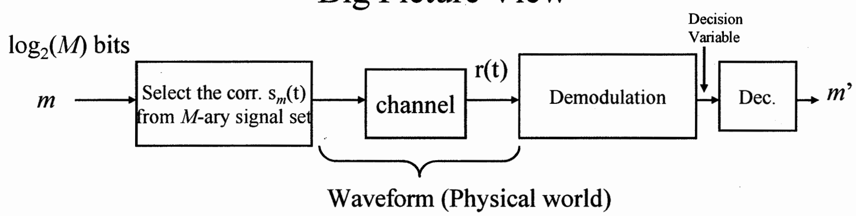
## Example 2




- Find the dimensionality of the signal set
- Find a basis set
- Obtain a vector representation
- Draw signal space plot
- Compute distances between signals



## Big Picture View




- Communication over AWGN channel

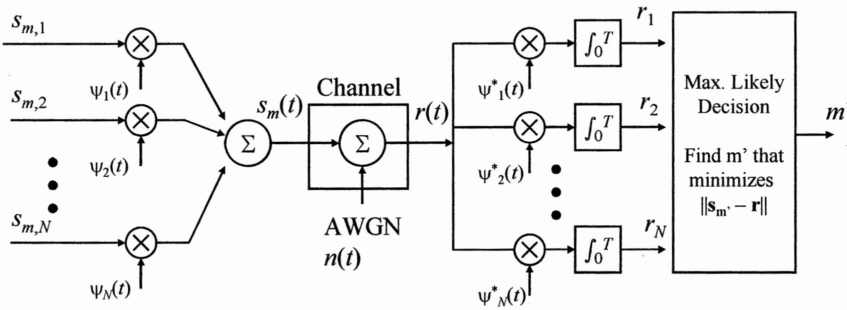
  $r(t) = s_m(t) + n(t),\ 0 \leq t \leq T$
- Choose a set of $M$ distinct signal waveforms.
- From previous discussion (Sig. Space rep), we know we can represent the set of $M$ messages using a set of basis functions.
- The size of the basis set $\{\psi_j(t), j = 1,2,\ldots,N\}$ required to span the $M$-ary signal set is the dimension $N$ of the signal set.

# Big Picture View (2)

- *Modulation* (wide sense) is to imply the mapping rule which takes a string of $k=\log_2(M)$ bits as input and transmits the corresponding signal $s_m(t)$ over the channel.
  - Transmission rate is $k$ bit/$T$ sec
- *Demodulation* (wide sense) is to imply the process of converting the analog waveform into a string of digits—a set of test statistics or decision variables
  - Use the matched filter (or correlator) for optimum performance.
  - This results in the receiver which is a *mirror image* of the transmitter.
- The decision device makes the final decision $m'$ on the message.
- A goal is to minimize the probability of error, $P(e) = P(m \neq m')$.



# Optimal Transceiver




- Use a basis $\{\psi_j, j=1, 2, \ldots, N\}$ to modulate and demodulate a message $m$.
- $s_m(t) \leftrightarrow \mathbf{s}_m=(s_{m,1}, s_{m,2}, \ldots, s_{m,N})$
- $n(t) \leftrightarrow \mathbf{n} = (n_1, n_2, \ldots, n_N)$ — $n_1, n_2, \ldots$ i.i.d. Gaussian~ $\mathcal{N}(0, \sigma^2)$
- $r(t) \leftrightarrow \mathbf{r} =(r_1, r_2, \ldots, r_N)$ — These are called sufficient statistics (test variables, decision variables)

# Inner Product Receivers

- Note that at each receiver branch, the receiver calculates the inner-product.
- After obtaining **r**, it's easy to calculate the distance $||\mathbf{s}_{m'} - \mathbf{r}||$ for each candidate $m'$.
- ML decision is the Minimum Distance decision.
- Thus, the inner product receiver is the ML sense optimal receiver.

- We refer to the inner product operator as the *correlator*.
  - The received signal gets *correlated* with each basis vector.



# Optimum Decision Rules

- Let's consider the optimal decision receiver.
- The receiver has the received vector
  - $\mathbf{r} = \mathbf{s}_m + \mathbf{n}$,
  - $\mathbf{n}$ is a multivariate Gaussian with $p(\mathbf{n}) = (2\pi\sigma^2)^{-N/2} \exp[-||\mathbf{n}||^2/2\sigma^2]$ .
  - Each of the marginal distr. $p(n_i)$ is Gaussian with zero mean and variance $\sigma^2 = N_o/2$ . We use notation $\mathcal{N}(0, N_o/2)$.
- Maximum *a-posteriori* (MAP) criterion:
  Find the message index $m$, among 1, 2,…, $M$, that maximizes the posterior probability

$$P(\mathbf{s}_m \mid \mathbf{r}) = P(\mathbf{s}_m, \mathbf{r})/P(\mathbf{r}) \propto P(\mathbf{r} \mid \mathbf{s}_m)\, P(\mathbf{s}_m)$$

$$\text{Posteriors} \propto \text{Likelihoods} \times \text{Priors}$$

- When $P(\mathbf{s}_m) = 1/M$ (equally likely), maximization on the posterior probability is equivalent to maximization on the likelihood function.
  - Implies that Maximum A Posteriori decision = Maximum Likelihood decision

## MAP = ML Detection, with equally likely input = Min. E. D. Rule, with Gaussian noise

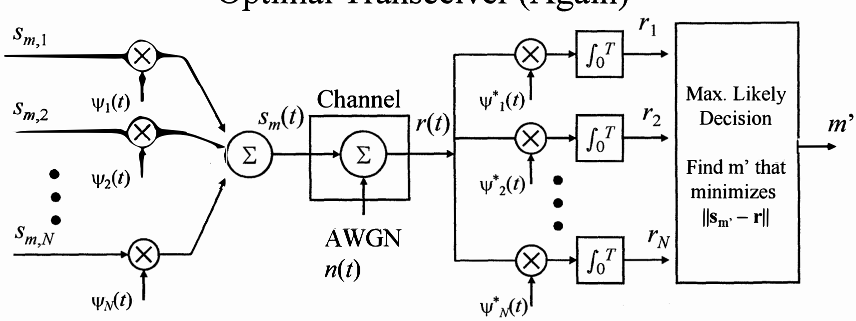
- The MLD rule for AWGN
- $P(\mathbf{r} \mid \mathbf{s}_m) = P(\mathbf{n} = \mathbf{s}_m - \mathbf{r}) = (2\pi\sigma^2)^{-N/2} \exp[- \|\mathbf{r} - \mathbf{s}_m\|^2 / 2\sigma^2]$

Norm/Euclidean distance Between **r** and hypothesis $\mathbf{s}_m$

- Log is monotone increasing function
- Use $\log[\Pr(\mathbf{r} \mid \mathbf{s}_m)] \propto -\| \mathbf{r} - \mathbf{s}_m\|^2 / 2\sigma^2$
- Finally, we note that
  Maximum Likelihood Detection rule = Minimum Euclidean Distance Rule



## Optimal Transceiver (Again)




- Use a basis $\{\psi_j, j=1, 2, \ldots, N\}$ to modulate and demodulate a message $m$.
- $s_m(t) \leftrightarrow \mathbf{s}_m=(s_{m,1}, s_{m,2}, \ldots, s_{m,N})$
- $n(t) \leftrightarrow \mathbf{n} = (n_1, n_2, \ldots, n_N)$ $\quad n_1, n_2, \ldots$ i.i.d. Gaussian~ $\mathcal{N}(0, N_o/2)$
- $r(t) \leftrightarrow \mathbf{r} =(r_1, r_2, \ldots, r_N)$ $\quad$ These are called sufficient statistics (test variables, decision variables)

## Decision Cells and Decision Boundary

- Example) a binary signal set

- Example) 4-ary signal set

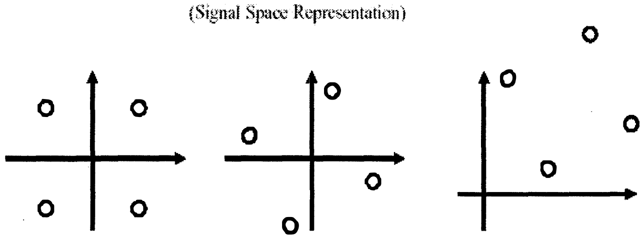


## Signal Design Criteria

- Bandwidth efficient design
- Power efficient design
- A signal requires resources
  - Signals occupy time and frequency
    - Use as little *frequency* bandwidth as possible
    - Use as little *time* as possible
    - Fundamental limit—Time-frequency uncertainty
    - Thus, need a balance
- Another resource is power
  - Use as little power as possible

## Good/Bad Constellations

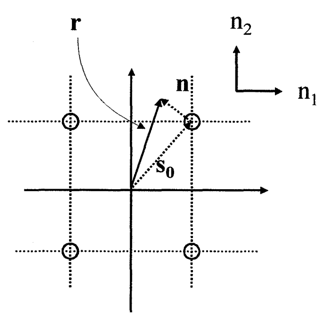
(Signal Space Representation)





## Example with 4-QAM

- Let's consider a 4-ary signal space plot again.
- 4-QAM constellation is a two dimensional signal set.
- Basis vectors of this space are
  $\psi_1(t)$ = sqrt(2/$T$) cos($w_c t$),
  $\psi_2(t)$ = - sqrt(2/$T$) sin($w_c t$),
  for $0 \leq t \leq T$
- The input/output signals can be represented with 2-tuples
  **r = s + n.**



## Example with $M$-ary FSK signals

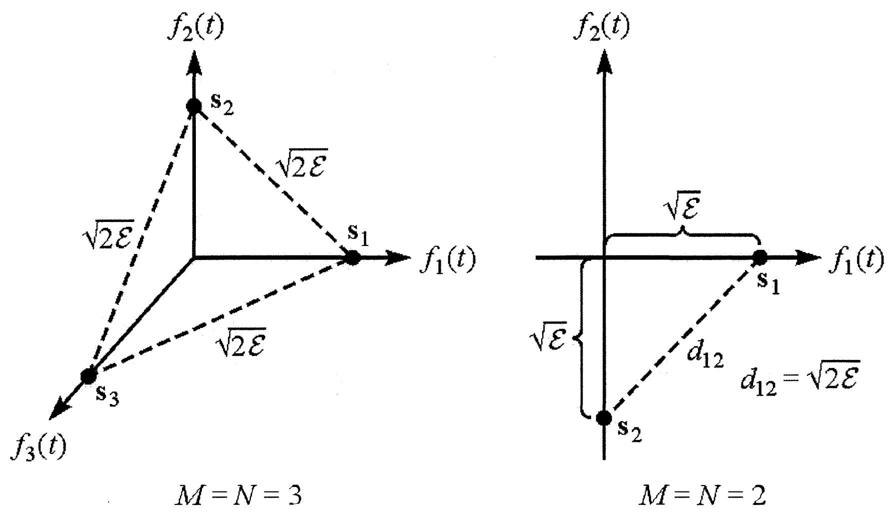
- $M$-ary FSK signals, for $0 \le m \le M-1$

  $$s_m(t) = \mathrm{Re}\{g_m(t)\exp(j\,2\pi f_c\,t)\}$$
  $$= \sqrt{\tfrac{2E_s}{T}}\cos[2\pi f_c\,t + 2\,\pi\,m\,\Delta f t], \qquad 0 \le t \le T,$$
  $$g_m(t) = \sqrt{\tfrac{2E_s}{T}}\exp(j\,2\pi\,m\,\Delta f\,t), \qquad 0 \le t \le T,$$

  where $E_s$ is the energy of the signal.

- The minimum frequency separation between adjacent signals, for orthogonality, is $1/2T$.





## Matched filters, why?

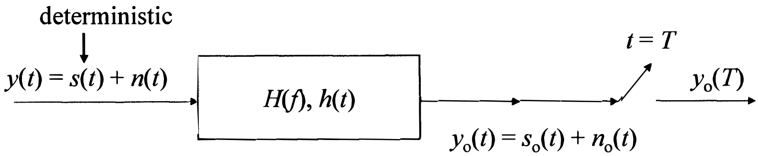
- We use the inner-product receiver (or the correlator) for the optimal receiver.
- The correlator output can also be obtained from using matched filters.
- Thus, the optimal receiver can also be realized with a matched filter bank.
- Matched filter is sometimes easier to implement than the correlator.
- The following discussion shows that matched filter maximizes the SNR when it is sampled at the right moment.



## Schwarz's Inequality

- Schwartz inequality
- $|\mathbf{v}_1 \cdot \mathbf{v}_2| \leq ||\mathbf{v}_1||\ ||\mathbf{v}_2||$ with equality iff $\mathbf{v}_2 = c\ \mathbf{v}_1$
- $|\int_t f(t)\, g^*(t)\, dt| \leq (\int_t |f(t)|^2\, dt)^{1/2} (\int_t |g(t)|^2\, dt)^{1/2}$ with equality if and only if $f(t) = K\, g(t)$

## Matched Filter Maximizes $(S/N)_{out}$




- $y(t)$ is the received signal. We know that the signal $s(t)$ has a duration $0 \le t \le T$, and the PSD of the noise $n(t)$ is $P_n(f)$.
- Find the best filter $H(f)$ so that $(S/N)_{out} := \|s_o(t)\|^2/E\{|n_o(t)|^2\}$ is maximum at $t=T$.
- The signal part: $s_o(t=T) = \int H(f)S(f) \exp(j2\pi f T)\, df$.
- The noise part: $n_o(t=T) = \int H(f)\, N(f) \exp(j2\pi f T)\, df$.



## Matched Filter maximizes $(S/N)_{out}$

- $E\{n^*_o(T)\, n_o(T)\} = \int |H(f)|^2\, P_n(f)\, df$
- Thus, the output SNR is

$$\begin{aligned}(S/N)_{out} &= \frac{|\int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi fT}df|^2}{\int_{-\infty}^{\infty}|H(f)|^2 P_n(f)df} \\ &\le \frac{\int_{-\infty}^{\infty}|H(f)|^2 P_n(f)df \int_{-\infty}^{\infty}\frac{|S(f)|^2}{P_n(f)}df}{\int_{-\infty}^{\infty}|H(f)|^2 P_n(f)df}\end{aligned}$$

  - where the second line is due to Schwartz' inequality
  - Equality is achieved iff $A(f) = K\, B(f)$
    - $A(f) = H(f)\, P_n(f)^{1/2}$
    - $B(f) = S^*(f) \exp(-j2\pi fT)/P_n(f)^{1/2}$

## Matched Filter Maximizes $(S/N)_{out}$

- $(S/N)_{out} \le \int |S(f)|^2/P_n(f)\, df$
- The equality is attained when

  $H(f) = K\, S^*(f)\, e^{-j2\pi fT}/P_n(f)$

- When the noise is white, then $P_n(f) = N_o/2$, we have $H(f) = S^*(f)\, e^{-j2\pi fT}$ (From IF ➔ $h(t) = s^*(T-t)$)
  - $(S/N)_{out} \le \frac{\int |S(f)|^2 df}{N_o/2} = \frac{E_s}{N_o/2}$



## Closer Look at $SNR_o$

- When the noise is white, $H(f) = S^*(f)\, e^{-j2\pi fT}$
- The variance of the noise sample $n_o(T)$ (note that the mean is zero) is

  $E\{n_o^*(T)n_o(T)\}$

  $= E\{\int\int H^*(f_1)H(f_2)\, N(f_2)\, N(f_1)\, \exp(j2\pi f_1 T)\, \exp(-j2\pi f_2 T)\, df_1\, df_2\}$

  $= \int\int H^*(f_1)H(f_2)\, E\{N^*(f_2)\, N(f_1)\}\, \exp(j2\pi f_1 T)\, \exp(j\text{-}2\pi f_2 T)\, df_1\, df_2$

  $= \int\int H^*(f_1)H(f_2)\, (N_o/2)\, \delta(f_2-f_1)\, \exp(j2\pi(f_2-f_1)T)\, df_1\, df_2$

  $= (N_o/2) \int H^*(f)H(f)\, df$

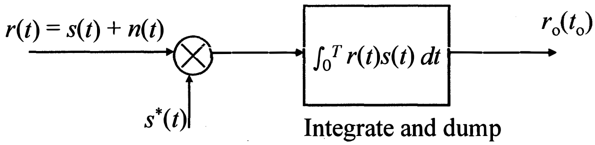
  $= (N_o/2) \int |H(f)|^2\, df$

  $= (N_o/2) \int |S(f)|^2\, df$

  $= (N_o/2)\, E_s$

- Or simply we note that the PSD of the noise after the filter is $(N_o/2)\, |S(f)|^2$. Thus, the noise power at the output of the matched filter is $(N_o/2) \int |S(f)^2|\, df = E_s\, (N_o/2)$.
- The energy (power) of the signal $s_o(T)$ is $E_s^2$.
- Thus $SNR_o = E_s^2/[E_s(N_o/2)] = E_s/(N_o/2)$.

## Correlator Realization of Matched Filtering (White noise only)

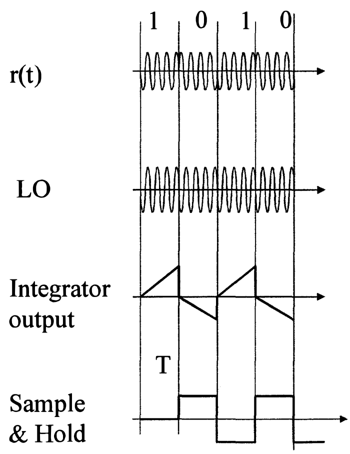



- For white noise, the matched filter result is the same as the correlator output—integrated and dumped.



## Correlator for Switched Sinusoids

- Integrator resets every $t = T$
- The sampler takes sample every $t = T$

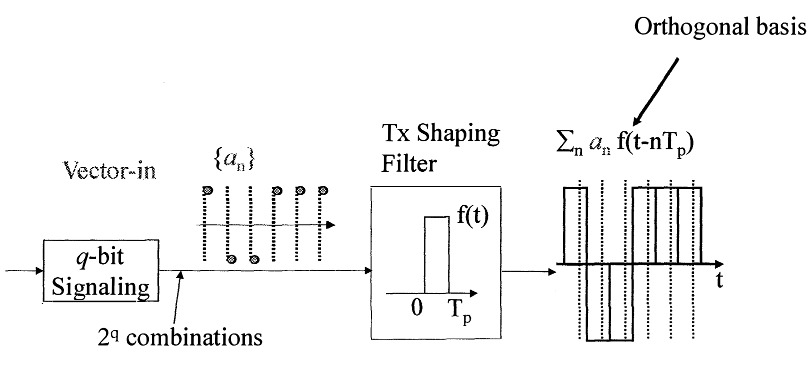


- Now let's consider a very simple transmitter and receiver pair.
- The purpose is to illustrate the overall idea of what we have learned so far.
  - What are they???



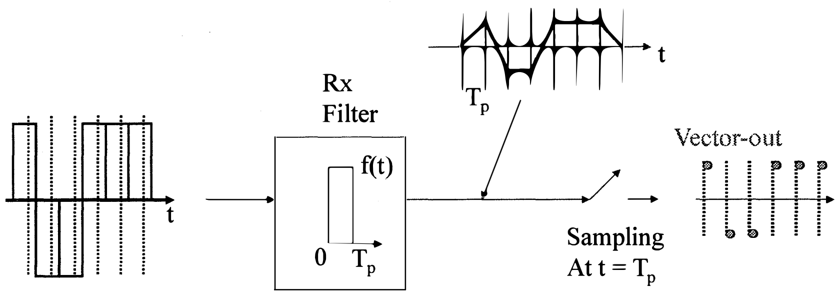
## Transmitter

- Consider the following example



# Receiver

- The Receiver is a mirrored system of the transmitter






# Binary Signals: Pass Band

- On-Off Keying:
  - $A_c \cos(w_c t)$ for "1"
  - Nothing for "0"
- Binary Phase-Shift Keying (BPSK)
  - $s(t) = A_c \cos[w_c t + \pi/2 \ m(t)]$
  - m(t) is the polar baseband signal.
  - $A_c \sin(w_c t)$ for "1" and $-A_c \sin(w_c t)$ for "0."
- Binary Frequency-Shift Keying (FSK)
  - Two FM signals for "1" and "0."
  - $s(t) = A_c \cos(w_1 t + \theta_1)$ or $A_c \cos(w_2 t + \theta_2)$

## Binary Bandpass Signals

- Shown left are two kinds of signals.
  - Baseband signals
  - Bandpass signals
- Baseband signals
  - Unipolar
  - Polar
- Bandpass signals
  - On-Off keying
  - Phase Shift Keying
  - Frequency Shift Keying

The pictures scanned from Crouch, 6th Edition.

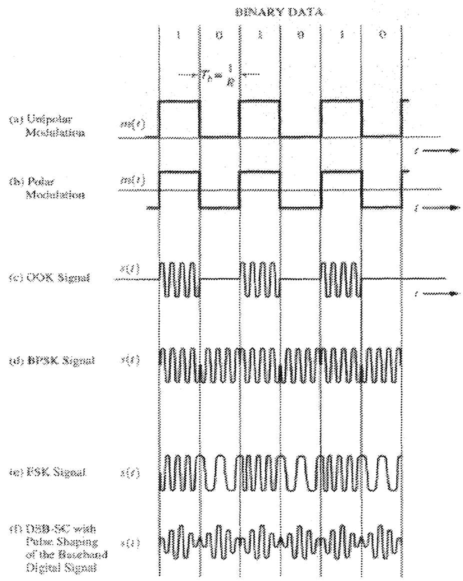


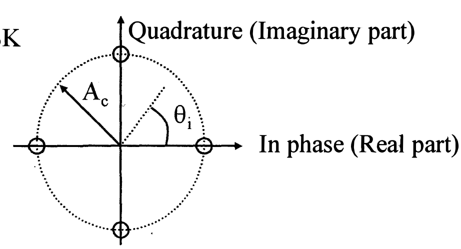

Figure 5–19 Bandpass digitally modulated signals.



## Multi-level Modulated Bandpass Signals

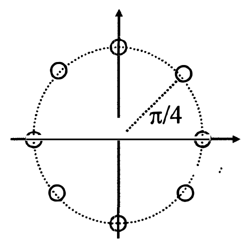
- *M*-ary Phase Shift Keying
  - $s(t) = A_c \cos(w_c t + \theta_i)$, for $i = 0, 1, 2, \ldots, M-1$.
  - The phase carries information
  - We use two-dimensional signal space plot in order to represent the signal.

4-PSK



## 8 PSK

- 8 PSK example
  i=0, 1, 2, …, 7
- π/4 is the angle between any two adjacent points in the signal space plot.
- We call the plot shown right *digital signal constellation.*



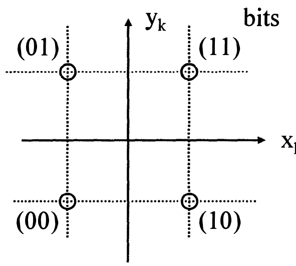



## Baseband vs. Pass band

- $s(t) = A_c \cos(w_c t + \theta_i)$, for $i = 0, 1, 2, \ldots, M - 1$.
- Get rid of the carrier
- $s(t) = \mathrm{Re}\{A_c \exp(\mathrm{j}\,(w_c t + \theta_i))\}$
  $= \mathrm{Re}\{A_c \exp(\mathrm{j}\ \theta_i) \exp(\mathrm{j}\ w_c t\,)\}$
  $= \mathrm{Re}\{g(t) \exp(\mathrm{j} w_c t)\}$
- Baseband signal: $g(t) = A_c \exp(\mathrm{j}\ \theta_i))$
- Complex number → real and imaginary part

# 4-QAM signals

- Baseband signal: $g(t) = x(t) + \mathrm{j}\, y(t) = R(t) \exp(\mathrm{j}\, \theta(t))$
  - $R(t)$ is called the envelop signal (the magnitude part). It's $A_c$ here.
- Let $f(t) = \mathrm{rect}((t - T/2)/T)$ for now.
- Suppose $g(t)$ in the $k$-th signaling interval takes a value from the four predefined constellation points.
  - $A_c(1 + \mathrm{j}), A_c(-1 + \mathrm{j}), A_c(-1 - \mathrm{j}), A_c(+1 - \mathrm{j})$
  - $A_c(x_k + \mathrm{j}\, y_k)$, where $x_k, y_k = +1$ for bit "1" and -1 for bit "0"

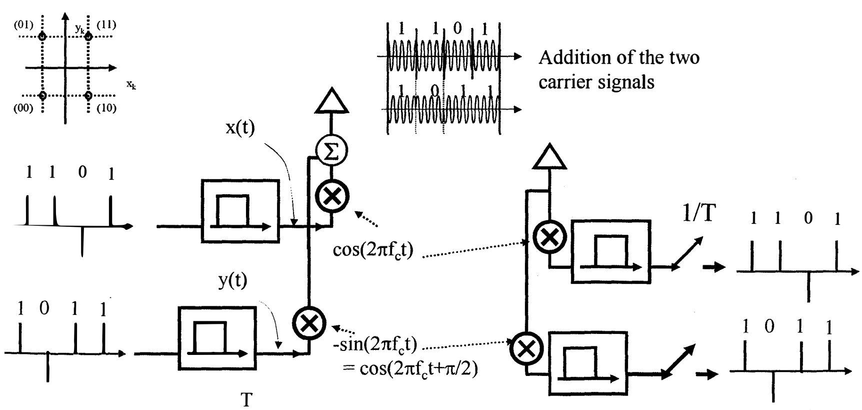


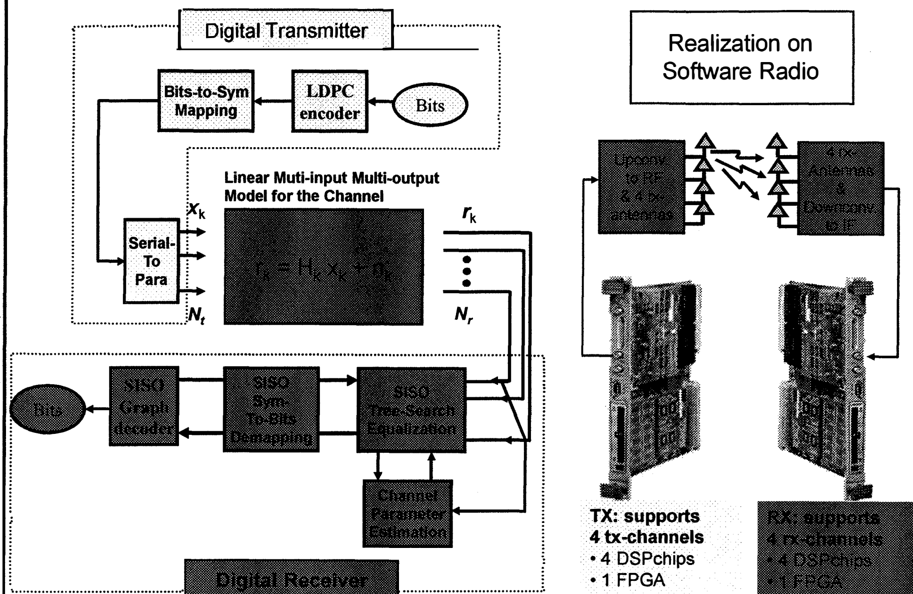

$$
\begin{aligned}
g(t) &= \textstyle\sum_k A_c(x_k + \mathrm{j}\, y_k) f(t - kT) \\
&= A_c \textstyle\sum_k x_k f(t - kT) + \mathrm{j}\, A_c \textstyle\sum_k y_k f(t - kT) \\
&= x(t) + \mathrm{j}\, y(t)
\end{aligned}
$$



# Quadrature Amplitude Modulation

- Two independent channels can be obtained over the same RF spectrum (Orthogornality: $\int_0^T \cos(2\pi f_c t) \sin(2\pi f_c t)\, dt = 0$)
- $s(t) = x(t) \cos(w_c t) - y(t) \sin(w_c t)$








## GNU Software Radio

- "GNU Radio is a free software development toolkit that provides the signal processing runtime and processing blocks to implement software radios using readily-available, low-cost external RF hardware and commodity processors." http://gnuradio.org/redmine/wiki/gnuradio.

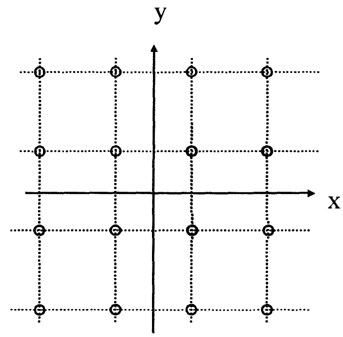
- INFONET's software radio
  - Three wireless nodes
  - Aim to do network coding, multi-user detection, distributed source coding

## 16 QAM Constellation

- Each symbol carries four information bits.
- In general the in-phase $x(t)$ and quadrature part $y(t)$ can be written as

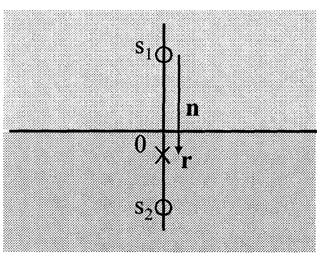
  $x(t) = \sum_k x_k f(t - kT_p)$

  $y(t) = \sum_k y_k f(t - kT_p)$
- For example of 16-QAM, $y_k$, $x_k \in \{A_c, 3A_c, -A_c, -3A_c\}$






## Decision Regions for BPSK




Minimum distance rule gives a decision boundary which divides the entire space into two mutually-exclusive decision regions.

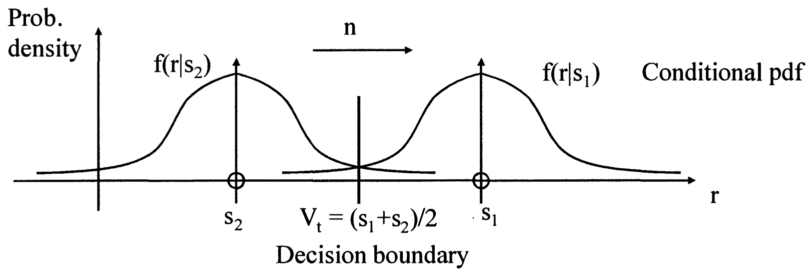
Let's assume $s_1$ was sent, we note that error occurs when $n \leq |s_1 - s_2|/2$

- $s_1(t) = A_c \sin(w_c t)$, $0 \leq t \leq T$, for binary digit "1 ."
- $s_2(t) = -A_c \sin(w_c t)$, $0 \leq t \leq T$, for binary digit "0."
- The dimensionality of the signal set is 1 (use $\psi(t) = \text{sqrt}(2/T) \sin(w_c t)$).

  $s_1(t) \leftrightarrow s_1 = A_c \text{ sqrt}(T/2)$

  $s_2(t) \leftrightarrow s_2 = -A_c \text{ sqrt}(T/2)$
- Thus, the difference energy $E_d \leq d_E^2 = 2A_c^2 T$.
  - "=" achieved when the matched filter is used.

## Error Probability for Binary Signaling




- $\Pr(\text{error}) = P(\text{error}, s_1 \text{ sent}) + P(\text{error}, s_2 \text{ sent})$
  $= P(\text{error} \mid s_1 \text{ sent})\, P(s_1 \text{ sent}) + P(\text{error} \mid s_2 \text{ sent})\, P(s_2 \text{ sent})$
  $= P(r \le V_t \mid s_1 \text{ sent})\, P(s_1 \text{ sent}) + P(r \ge V_t \mid s_2 \text{ sent})\, P(s_2 \text{ sent})$
  $= P(n \le V_t - s_1)\, P(s_1 \text{ sent}) + P(n \ge V_t - s_2)\, P(s_2 \text{ sent})$
  --- Let pdf of n is $P(n) = f(n)$
  $= P(s_1 \text{ sent}) \int_{-\infty}^{V_t} f(r|s_1)\, dr + P(s_2 \text{ sent}) \int_{V_t}^{\infty} f(r|s_2)\, dr$ --- (1)

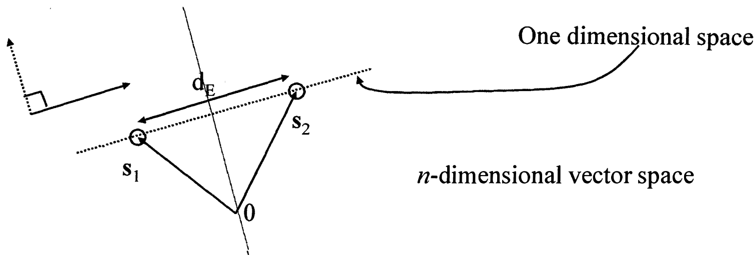


## $P_e$ for binary signaling

- Assume $s_1$ and $s_2$ equally likely $P(s_1 \text{ sent})=P(s_2 \text{ sent})=0.5$
- $f(n)$ is Gaussian with zero mean and variance $\sigma^2$
  - $f(r|s_1) = (\pi N_o)^{-1/2} \exp(-(r - s_1)^2/ N_o)$
  - $f(r|s_2) = (\pi N_o)^{-1/2} \exp(-(r - s_2)^2/ N_o)$
- We note that because of symmetry, the optimal choice of $V_t$ is the half point $(s_1 + s_2)/2$.
- Making use of symmetry of the Gaussian pdf's, we can write
  - $P_e = \int_{(s_1-s_2)/2}^{\infty} \text{sqrt}(\pi N_o)^{-1} \exp(-n^2/ N_o)\, dn$

$$= Q\left(\frac{|s_1 - s_2|}{\sqrt{2N_o}}\right) = Q\left(\sqrt{\frac{(s_1-s_2)^2}{2N_o}}\right) = Q\left(\sqrt{\frac{E_d}{2N_o}}\right)$$

## Probability of Making Errors for Binary Signaling




- $d_E = \| \mathbf{s}_1 - \mathbf{s}_2 \|$ is the Euclidean distance between the two signals.
- $d_E^2$ is the difference in signal energy.

- Decision error depends only upon the one dimensional noise acting along the line shown above (the noise components perpendicular to the line do not contribute toward generating the decision error event.).

- Assume that any one signal is sent. Then the error event is the set of all outcomes in which the received signal lands on the decision regions of the other signals. This event occurs when the one dimensional noise is directed toward the other codeword with its magnitude large enough to move across the decision boundary.



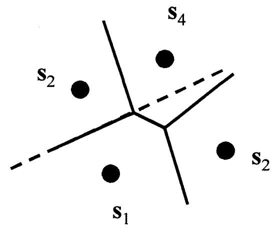
## P(e) for General Cases

- How to find P(e) for general high dimensional $M$-ary constellations?

- Difficult to obtain exact P(e)
- Let's use
  - Union upper bound
  - Lower bound
  - Approximation
- Note the in/out relation
  - $\mathbf{r} = \mathbf{s}_m + \mathbf{w}$
  - $\mathbf{w} \sim$ each element of $\mathbf{w}$ is i.i.d. with $\mathcal{N}(0, N_o/2)$

# Union Upper Bounds

- Equally likely symbols




$$P(e) = \sum_{i=1}^{M} \sum_{j \neq i}^{M} P\{\mathbf{r} \in \mathcal{D}_j | \mathbf{s}_i\} p(\mathbf{s}_i)$$

$$= \sum_{i=1}^{M} \frac{1}{M} \sum_{j \neq i}^{M} P\{\mathbf{r} \in \mathcal{D}_j | \mathbf{s}_i\}$$

- Let's denote the pairwise error probability $P(\mathbf{s}_i \to \mathbf{s}_j)$
- It's an error probability assuming there are only two symbols in the signal space.
- Thus, we have $P\{\mathbf{r} \in \mathcal{D}_1 | \mathbf{s}_i\} \leq P(\mathbf{s}_i \to \mathbf{s}_j)$. Why?

$$\leq \sum_{i=1}^{M} \frac{1}{M} \sum_{j \neq i}^{M} P(\mathbf{s}_i \to \mathbf{s}_j)$$

Let's denote $d_{i,j}$ the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$.

Then, we note $P(\mathbf{s}_i \to \mathbf{s}_j) = Q\left(\frac{d_{i,j}}{\sqrt{2N_o}}\right)$



# Union Upper Bounds(2)

$$P(e) \leq \sum_{i=1}^{M} \frac{1}{M} \sum_{j \neq i}^{M} Q\left(\frac{d_{i,j}}{\sqrt{2N_o}}\right)$$

- Let's denote $d_i := \min\{d_{i,j}, i \neq j, \; j = 1, 2, ..., M\}$

$$\leq \sum_{i=1}^{M} \frac{1}{M}(M-1) Q\left(\frac{d_i}{\sqrt{2N_o}}\right)$$

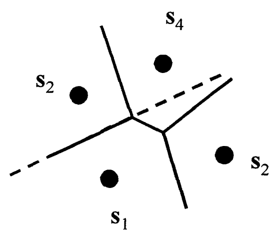
- Let's denote $d_{\min} := \min\{d_i : i = 1, 2, ..., M\}$

$$\leq (M-1) Q\left(\frac{d_{\min}}{\sqrt{2N_o}}\right)$$

- This is a useful upper bound often used.

## Lower Bounds

- Equally likely symbols




$$P(e) = \sum_{i=1}^{M} P\{\mathbf{r} \in \mathcal{D}_i^c \mid \mathbf{s}_i\}\ p(\mathbf{s}_i)$$

- Note $P\{\mathbf{r} \in \mathcal{D}_i^c \mid \mathbf{s}_i\} \geq P\{\mathbf{s}_i \to \mathbf{s}_j\}$ for any $j$

$$\geq \sum_{i=1}^{M} \frac{1}{M} Q\left(\frac{d_i}{\sqrt{2N_o}}\right) \geq \sum_{i=1}^{M} \frac{1}{M} Q\left(\frac{d_{i,j}}{\sqrt{2N_o}}\right)$$

- To make the bound tight we may want to choose $d_i$.
- We continue with the tighter one and use only those terms that $d_i = d_{min}$

$$\geq \frac{N_{\min}'}{M} Q\left(\frac{d_{\min}}{\sqrt{2N_o}}\right)$$

where $N_{min}$' is $\#\{i: d_i = d_{min}, i = 1, 2, \ldots, M\}$

©200x Heung-No Lee

## Upper and Lower Bounds for General $M$-ary Constellations

$$\frac{N_{\min}}{M} Q\left(\frac{d_{\min}}{\sqrt{2N_o}}\right) \leq P(e) \leq (M-1) Q\left(\frac{d_{\min}}{\sqrt{2N_o}}\right)$$

- This relation can be used for any constellation
- Note there are only two parameters
  - $d_{min}$ : minimum Euclidean distance of constellation
  - $N_{min} := |\{j : d_j = d_{min}, j = 1, 2, \ldots, M\}|$

©200x Heung-No Lee

## Example

- Obtain P(e) for the following constellation
  - $s_1 = \sqrt{E_s}$ [ 1 -1 -1 1 1 -1];
  - $s_2 = \sqrt{E_s}$ [ -1 -1 1 -1 -1 1];
  - $s_3 = \sqrt{E_s}$ [ -1 1 1 1 1 -1];
  - $s_4 = \sqrt{E_s}$ [ 1 -1 1 -1 -1 -1];
- Obtain $d_{min}$ and $N_{min}$
- Simulate the system and find symbol error rate, and compare it with the upper and lower bound as SNR is varied.



## Appendix --

Often Used Identities

- $\sin(x \pm y) = \sin(x)\cos(y) \pm \cos(x)\sin(y)$
- $\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y)$
- $\cos(x)\cos(y) = ½(\cos(x+y) + \cos(x-y))$
- $\sin(x)\sin(y) = -½(\cos(x+y) + \cos(x-y))$
- $\cos(x)\sin(y) = ½(\sin(x+y) - \sin(x-y))$
- $\sin(x) = (1/2j)(e^{jx} - e^{-jx})$
- $\cos(x) = (1/2)(e^{jx} + e^{-jx})$

## Problems

1. (the Q(x) function) Assume $X$ is a Gaussian distributed random variable with mean 1 and variance of $\sigma^2$. Find the probability $\Pr\{X > 5\}$. Express the probability with the Gaussian Q function (see definition in Section 2.3).
2. Find out if sin and cos waveforms are orthogonal to each other. If yes, under what condition?

3. P2.3 (KL Decomp),
4. P2.11 (Rep. of Signals),
5. P2.25 (Bounds on Q(x) function),
6. P2.51 (Sampling theorem)

7. P3.2 (Signal Representation)
8. P3.6 (Power Efficient Constellation)

9. P4.5 (Signal Representation/Constellation)



## Problems

- Consider a communications system with the following conditions:
  - There are eight users and one access point.
  - All eight users make accesses to the access point simultaneously.
  - They use the same frequency band as well. The bandwidth is 1MHz.
  - Each user sends 1Mbps with arbitrarily small errors.
- Is it possible to design a set of waveforms for such a multiple access system which support all the statements above ? If yes, please provide one design. For full credits, justification to the level of this lecture note should be given.

# Information Theory and Gallager's Random Coding Bound

 1

## Agenda

- Shannon's Information Theoretic Capacity and Channel Coding Theorem
  - Read his 1948 paper
- Random Coding Bounds by Gallager (Chapter 5 of Gallager; Chapter 6 of Proakis/Salehi)
  - Read Gallager's 1965 paper

 2

## Brief Discussion on Channel Capacity

- Shannon introduced two measures of information
  - Entropy
  - Mutual Information



## Definition of Entropy

- Entropy is the average *measure* of uncertainty of a distribution, $p_1, p_2, \ldots, p_n$ .

$$H(p_1, p_2, \ldots, p_n) := \Sigma_{j=1}^{n} p_j \log(1/p_j)$$

- If $p_i < p_j$, then $\log(1/p_i) > \log(1/p_j)$.
  - Less probable event means larger uncertainty.
  - More probable event means smaller uncertainty.
  - The sure event has zero uncertainty.
- Uncertainty = Amount of Information = The number of bits needed in representation.

## Entropy of a RV

- Let X be a random variable with alphabet $A = \{x_1, x_2, \ldots, x_n\}$ and its probability mass function $p(x) = \Pr\{X=x_i \in A\}$.
- We define entropy for r.v. X

  $$H(X) := \sum_{x\in\mathcal{X}} p(x) \log(1/p(x))$$

  - When the base of the logarithm is 2, the unit is "bits."
  - When the base is e, "nats."

©200x Heung-No Lee 5

## H(X) is the Average Uncertainty (Information) of X

- Let's take some examples
- Ex1) When X is binary
  - X ~ Uniform [1/2, 1/2].
  - X ~ Bernoulli($p = 2^{-4}$). $H(X) = \frac{1}{16}\log_2(16) + \frac{15}{16}\log_2\left(\frac{16}{15}\right) = 0.337$
- Ex2) When X is quaternary

As a simple example of some of these results consider a source which produces a sequence of letters chosen from among $A, B, C, D$ with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$, successive symbols being chosen independently. We have

$$H = -\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{2}{8}\log\tfrac{1}{8}\right)$$
$$= \tfrac{7}{4} \text{ bits per symbol.}$$

[Shannon1948, pg. 18]

©200x Heung-No Lee 6

## Maximum Entropies

- $H(X) \le \log|\mathcal{X}|$, where $|\mathcal{X}|$ is the size of alphabet, with equality iff X is uniform over $\mathcal{X}$.
- Gaussian distribution gives max. entropy under average energy constraint. For any $X$ whose 2nd moment is $N$, we have

$$h(X) \le \frac{1}{2}\log_2(2\pi e N) = h(Y)$$

where $Y \sim \mathcal{N}(0, N)$

©200x Heung-No Lee 7

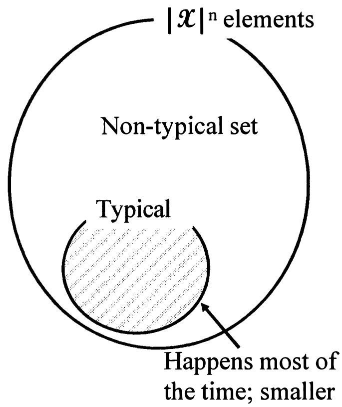
## Meaning of Entropy

- Let X be a 1-0 Bernoulli with 1 appearing with prob. 1/16.
  - H(X) = 0.337 from previous page.
- Consider a sequence of X's of length $n$, $(X_1, X_2, \ldots, X_n)$.
- For a large $n$, due to the LLN, the set of sequences can be divided into two sets.
  - A *typical* set of sequences which occur in real experiment
  - An *atypical* set of sequences which almost never occur
- Shannon noted that there are only $2^{nH(X)}$ typical ones.

©200x Heung-No Lee 8

## Typical Set

- Consider the sequences of length 32 (n = 32).
- Let's count the number of 1s in any sequence.
  - What is the number of 1s in a sequence you would see typically?
- What is the probability to see a sequence of all 1s?
- Atypical sequence happens very rarely. Why?
  - Law of large numbers




 9

## Summary on Entropy

- Entropy measures the degree of randomness.
  - Uniform distribution
  - Gaussian distribution
- An information generating source can be modeled as a random variable with entropy.
  - The amount of information a source produces can be quantified by its entropy.
- There are about $2^{nH(X)}$ typical sequences of length $n$ when a source with entropy $H(X)$ generates $n$ symbols.

 10

## Recall Communications System

Digital Tx → TX Radio → MIMO Channel → RX Radio → Digital Rx

- Input/Output Relationship between Digital Tx and Rx
  - $\mathbf{Y} = \mathbf{HX} + \mathbf{W}$
  - Rec'd signal matrix $\mathbf{Y} = [N_r \text{ x } T]$,
  - Ch. matrix is $\mathbf{H} = [N_r \text{ x } N_t]$,
  - Noise matrix $\mathbf{W} = [N_r \text{ x } T] \sim$ i.i.d. $\mathcal{N}(0, N_o/2)$ each dimension
  - Ch. symbol matrix is $\mathbf{X} = [N_t \text{ x } T]$
  - One tx and one rx antennas, then $\mathbf{Y}_{(1,:)} = \mathbf{H}_{(1,:)}\mathbf{X}_{(1,:)} + \mathbf{W}_{(1,:)}$ .
- Let's assume **H** is the identity matrix for now.
- Information Theory tells us how close $\mathbf{Y} = \mathbf{X} + \mathbf{W}$ is to $\mathbf{X}$.

## Joint Entropy and Conditional Entropy

- Joint Entropy: The joint entropy H(X, Y) of a pair of discrete random variable (X, Y) with a joint distribution p(x, y) is defined as

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

- Conditional Entropy:

$$H(Y \mid X) := -\sum_x \sum_y p(x, y) \log p(y \mid x)$$
$$= -E\{\log p(Y|X)\}$$
$$= -\sum_x p(x) H(Y \mid X = x)$$
$$= -\sum_x p(x) \sum_y p(y|x) \log p(y|x)$$

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

- $H(X, Y) := - \sum_x \sum_y p(x, y) \log p(x, y)$

$$= - \sum_x \sum_y p(x, y) \log[p(x)\, p(y|x)]$$
$$= - \sum_x \sum_y p(x, y) [\log p(x) + \log p(y|x)]$$
$$= - \sum_x p(x) \log p(x) - \sum_x \sum_y p(x, y) \log p(y|x)$$
$$= H(X) + H(Y|X)$$

or similarly

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



## Self Exercise

| Y \ X | 0 | 1 |
|---|---|---|
| 0 | 2/8 | 4/8 |
| 1 | 1/8 | 1/8 |

- $H(X) = 3/8 * \log_2(8/3) + 5/8 * \log_2(8/5) = 0.9544$
- $H(Y) = 6/8 * \log_2(8/6) + 2/8*\log_2(8/2) = 0.8113$
- $H(Y|X) = \sum_x p(x)\, H(Y|X{=}x)$

$$= 3/8*H(Y|X{=}0)+5/8*H(Y|X{=}1)$$
$$= 3/8*H(2/3, 1/3)+5/8*H(4/5, 1/5)$$
$$= 3/8*0.9183+5/8*0.7219$$
$$= 0.7955$$

- $H(X, Y) = H(X) + H(Y|X) = 1.75$
- $H(X, Y) = - E\{\log p(X, Y)\}$

$$= 2/8*\log_2(4) + (4/8)*\log_2(2) + 2*1/8*\log_2(8)$$
$$= 1/4*2 + 1/2 + 2*3/8 = 1 + 3/4 = 1.75$$

The units are [bit].

## Mutual Information I(X; Y) = H(X) – H(X|Y)

- The measure of amount of information about X we can have knowing Y (vise versa).
  - Cf) Measure of correlation between X and Y
- Ex) Suppose Y = X, then H(X|Y) = 0 (no uncertainty). → Self-mutual information is entropy.
  - Thus, knowing Y means knowing X exactly (the full information H(X) = H(Y) is obtained)
- Ex) Suppose Y and X independent, then H(X|Y) = H(X), then I(X;Y) = H(X) – H(X) = 0.
  - Knowing Y cannot tell anything about X.

 15

## I(X, Y) = H(X) – H(X | Y)

- $I(X; Y) = \sum_{x\in X} \sum_{y\in Y} p(x, y) \log[p(x, y)/p(x)p(y)]$
  $= \sum_{x\in X} \sum_{y\in Y} p(x, y) \log[\cancel{p(y)}\, p(x|y)/p(x)\cancel{p(y)}]$
  $= \sum_{x\in X} \sum_{y\in Y} p(x, y) \{\log[p(x|y)] - \log[p(x)]\}$
  $= H(X) - H(X|Y)$
- Reduction in uncertainty of X due to the knowledge of Y
- Also, I(X; Y) = H(Y) – H(Y|X)
- How much can I tell about X knowing Y?
- How much can I tell about Y knowing X?
- I(X; Y) = I(Y; X)

 16

## Conditioning reduces entropy

- $H(X|Y) \leq H(X)$, with equality *iff* X and Y independent
  - $I(X; Y) = H(X) - H(X|Y) \geq 0$
  - cf) $I(X; Y) = 0$ *iff* X and Y independent.
- Meaning of conditional entropy $H(X|Y)$ or $H(Y|X)$
  - Given the output $Y = y$, $H(X|Y=y)$ tells the uncertainty on input X. This means that there are about $2^{nH(X|Y=y)}$ typical inputs for a particular input y.
  - Given the input $X = x$, $H(Y|X=x)$ tells the uncertainty on output Y. This means that there are about $2^{nH(Y|X=x)}$ typical inputs for a particular input x.

©200x Heung-No Lee 17

## Shannon Capacity

- We aim to show that $R < C$ for very small $P(e)$.
- See page 23 and 24 in his paper and read the key insight in his own words.

©200x Heung-No Lee 18

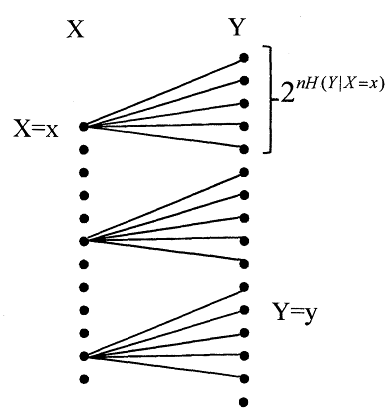
## Main Story in Shannon's Paper

- Given a channel relation (Y = X + W), find out the size *M* of the input message set which results in very small P(e).
  - You are allowed to use the same channel many times, say *n* times.
- The strength of the noise limits the size of the input message set. (Obvious)
- Determine the range of rates $R = \log_2(M)/n$ that gives P(e) very small.
- There are $2^{nH(X)}$ typical input sequences of length n.
- We choose $2^{nR}$ messages randomly out of total $2^{nH(X)}$ typical words.
- We want only one message out of total $2^{nR}$ messages falls into the fan of $2^{nH(X|Y)}$.

©200x Heung-No Lee 19

## Shannon's Key Idea

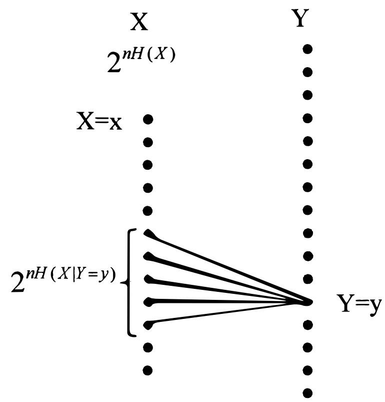
- There are $2^{nH(Y)}$ *typical* outputs.
- Each input message fans out to $2^{nH(Y|X)}$
- If we select only $\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(X;Y)}$ number of messages, no *equivocation* would occur.
  - P(e) is close to 0 (LLN)
- So, such a codebook *exists* and can be constructed, though difficult.
  - What if it is constructed randomly?




©200x Heung-No Lee 20

## Shannon's Key Idea: P(e) in Random Codebook Construction

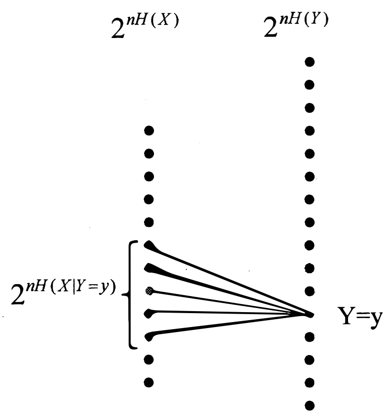
- Let's select the message set(Codebook) randomly.
- And, see if we can make P(e) very small.
- Given a fan of size $2^{nH(X|Y=y)}$, decoding error occurs if there are more than one messages.
  - See the analysis in the following page




 21

## Shannon's Key Idea: P(e) in Random Codebook Construction (2)

- Steps:
  - Select the first message (the red dot) and send.
  - With probability close to 1, we get the typical output $y$.
  - Randomly select the rest of messages.
  - Consider the *fan* of $y$ and find out the probability of decoding error.
  - Decoding error occurs when any one of $2^{nR} - 1$ other messages falls within the fan.




 22

no error

## P(e) in Random Codebook Construction (3)

error

- $$P(e) = 1 - \left(1 - \frac{2^{nH(X|Y)}}{2^{nH(X)}}\right)^{2^{nR}-1}$$
  $$\le 1 - \left(1 - 2^{-n[H(X)-H(X|Y)]}\right)^{2^{nR}}$$
  $$\approx 1 - \left(1 - 2^{nR}2^{-n[H(X)-H(X|Y)]}\right)$$
  $$= 2^{-n[I(X;Y)-R]}$$

- Thus, as long as $R$ is chosen slightly smaller than I(X; Y), P(e) decreases to zero as $n$ increases.
  - Now we maximize I(X; Y) by selecting the best input distribution, and obtain the capacity, $C = \max_{p(x)} I(X; Y)$.

- Note that our objective has been achieved.

 23

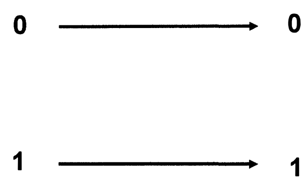
## "*Information*" Channel Capacity

- $C = \max_{p(x)} I(X; Y)$
  - The maximum is taken over all input distr. p(x)
- I(X; Y) is the mutual information between X and Y
  - I(X;Y) = H(X) – H(X|Y),
    ---- maximum input-size which causes no equivocation on X given an output Y
    = H(Y) – H(Y|X)
    ---- maximum output-size which causes no uncertainty on Y given an input X
  - H(X) = Size of uncertainty (information)
  - H(X|Y) = Size of uncertainty (information) on X given Y

 24

## Noiseless Binary Channel

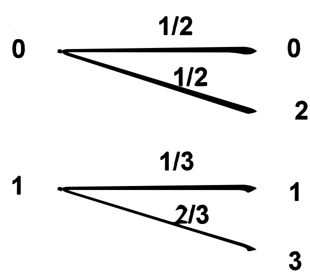
- $C = \max_{p(x)} I(X;Y) = ?$




©2004 Heung-no Lee 25

## Noisy Channel with Non-Overlapping Output

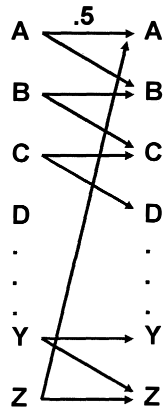
- C = ?




©2004 Heung-no Lee 26

## Noisy Typewriter

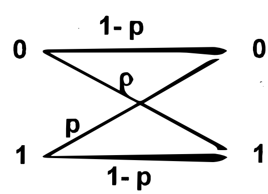
- The typewriter writes the input letter with prob. 1/2 or the next letter in the alphabet with prob. 1/2
- C = ?
- $C = \max_{p(x)} I(X ; Y)$
  $= \max[H(Y) - H(Y|X)]$
  $= \max H(Y) - 1$
  $= \log_2(26) - 1$
  $= \log_2(26/2)$
  $= \log_2(13)$




©2004 Heung-no Lee 27

## Binary Symmetric Channel

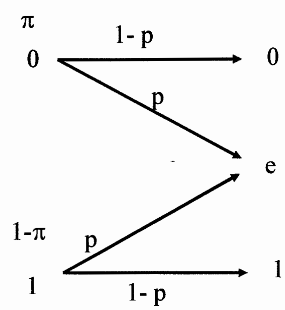
- $I(X; Y) = H(Y) - H(Y|X)$
  $= H(Y) - \sum p(x) H(Y|X=x)$
  $= H(Y) - H(p)$
  $\leq 1 - H(p)$
- The equality is when $H(Y) = 1$
  - Y is uniform when X is uniform
- $p(x) \sim$ uniform
- $C = 1 - H(p)$ bits




©2004 Heung-no Lee 28

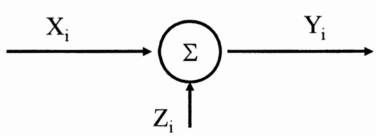
## Binary Erasure Channel

- Some bits are lost (no decision)
- $C = \max_{p(x)} I(X;Y)$
  $= \max_{p(x)} H(Y) - H(Y|X)$
  $= \max_{p(x)} H(X) - H(X|Y)$




©200x Heung-no Lee 29

## The Additive White Gaussian Noise Channel




- $Y_i = X_i + Z_i$,  $Z_i \sim \mathcal{N}(0, N)$
- $X_i$ and $Z_i$ are independent
- Constraint on the input $X_i$
  - $E(X_i^2) \leq P$
  - Successive use of the channel (trying to exploit the LLN)
  - Send a codeword $(x_1, x_2, \ldots, x_n)$: $(1/n) \sum_i x_i^2 = P$

©200x Heung-No Lee 30

## Information Capacity of AWGN

- $C = \max_{p(x):\, E(X^2) = P}\ I(X; Y)$
- $I(X; Y) = h(Y) - h(Y|X)$

$= h(Y) - h(X+Z|X)$

$= h(Y) - h(Z)$

$= h(Y) - 0.5 \log(2\pi e N)$

--- $E(Y^2) = E(X^2) + E(Z^2)$ – X, Z indep

$= P + Z$

$\leq 0.5 \log(2\pi e (P+N)) - 0.5 \log(2\pi e N)$

$= 0.5 \log(1+P/N)$

When do we have the equality?

 31

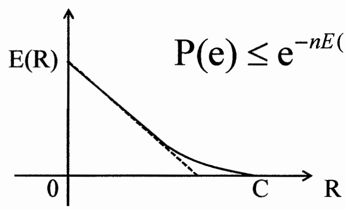
## Some Properties of Channel Capacity

1. $C \geq 0$, since $I(X; Y) \geq 0$
2. $C \leq \log|X|$, why?
3. $C \leq \log|Y|$, why?

 32

## Summary of Shannon's Result

- Capacity
  - What is it?
  - How to achieve it?
- Motivation for Gallager's Channel Coding Theorem
  - P(e) due to Maximum Likelihood Receiver
  - Error exponent
  - Block length n
  - $R = \log_2(M)/n$

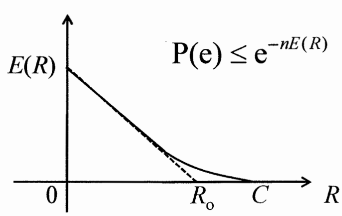



$$P(e) \le e^{-nE(R)}$$



## Motivation for Gallager's Channel Coding Theorem

- It provides P(e) due to Maximum Likelihood receiver
  - Error exponent $E(R)$ gets smaller as $R$ gets closer to $C$.
  - Block length $n \sim$ complexity
  - Transmission rate, $R = \log_2(M)/n$ [bits/symbol]
  - $R_o$ is called the cut-off rate.
- It shows how difficult it is to achieve a rate $R$ closer to $C$.




$$P(e) \le e^{-nE(R)}$$

## Comparison

- Sketch of ideas Shannon (Gallager) used
  - Allowing arbitrarily small but non-zero error probability
  - Successive uses of channels (utilize the LLN)
  - Calculate the average performance of codebooks, rather than that of a single codebook
  - Use typical set decoder (Use ML detector)

 35

## Channel Coding Theorem

- All rates below capacity C are achievable. Specifically, for rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.
- Conversely, any sequence of $(2^{R}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

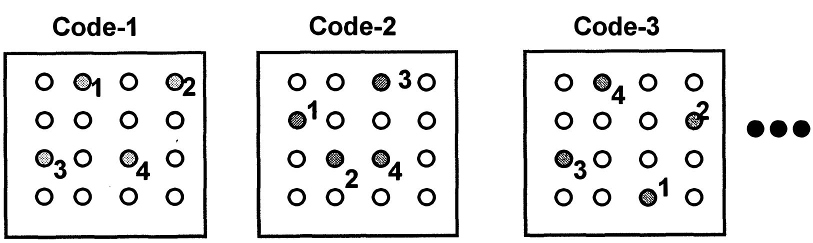
 36

## Ensemble Average over All Codebooks

- Consider a prob. of error for a particular word $w$ of a particular code $\mathcal{C}$
  - It's difficult to calculate $\lambda_w(\mathcal{C})$ because codewords in $\mathcal{C}$ are chosen random
- But if averaged over all codebooks
  - $\Sigma_C P(\mathcal{C})\, \lambda_i(\mathcal{C}) = \Sigma_{\mathcal{C}} P(\mathcal{C})\, \lambda_j(\mathcal{C})$, for any i and j, by symmetry

 37

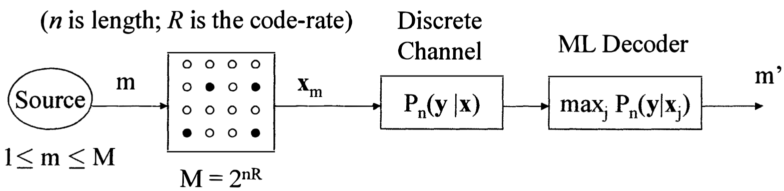
## $P_w$ in Random Coding Experiment




- Ensemble of codes
  - Select one, according to a prob. dist. $\prod_{m=1}^{M} Q_n(\mathbf{x}_m)$
  - Each code has its own probability of error
- Ensemble avg. P(e), instead of a particular P(e)
  - Easier to calculate
  - Quite a tight bound can be obtained

 38

## The Coding Theorem




- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y}=(y_1, y_2, \ldots, y_n)$
- $Q_n(x)$: probability of selecting a particular word as a codeword
- $\Pr(m \neq m') = P_{e,\,m} \leq (M\text{-}1)^{\rho} \sum_{\mathbf{y}} [\sum_{\mathbf{x}} Q_n(\mathbf{x})\, P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}]^{1+\rho}$ for some $0 < \rho \leq 1$.

©200x Heung-No Lee

39

## The Probability of Word Error

- $P_w := \Pr\{\text{word error}\}$

$= \sum_{m=1}^{M} P(m)[\sum_{\mathbf{x}m} \sum_{\mathbf{y}} Q_n(\mathbf{x}_m)\, P_n(\mathbf{y}\,|\mathbf{x}_m) \Pr\{\text{error}\,|m, \mathbf{x}_m, \mathbf{y}\}]$

$= \sum_{m=1}^{M} P(m)\, P_{e,m}$

Prob. of error assuming message *m* is transmitted

- Suffice to calculate $P_{e,\,m=1}$ due to the argument of random selection of codewords.

©200x Heung-No Lee

40

## $P_{e,\, m=1}$

- $P_{e,1} = \Sigma_{\mathbf{x}1}\, \Sigma_{\mathbf{y}}\, Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}|\mathbf{x}_1)\, \Pr\{\text{error} \mid m=1, \mathbf{x}_1, \mathbf{y}\}$

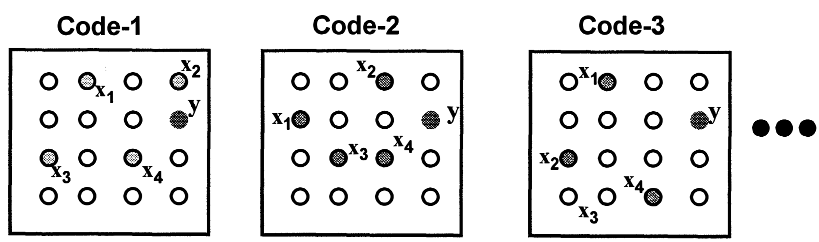
Prob. of choosing $\mathbf{x}_1$ as the first codeword in a code

Prob. of observing the channel output y

- Now, how do we calculate $\Pr\{\text{error} \mid m=1, \mathbf{x}_1, \mathbf{y}\}$?
  - Use bounds and approximations, and so on
  - Parameters involved are *n*, *R*, and P(e)



## Considering $\Pr\{\text{error} \mid m=1, \mathbf{x}_1, \mathbf{y}\}$

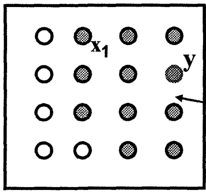



- Given $\mathbf{x}_1$ and $\mathbf{y}$, the error event depends on the selection of the rest of (M-1) codewords
  - Choose one, out of $2^n$ possible selections (allow repetitive selections for simplicity)
  - Any of the (M-1) selections will result in an error whenever a selection gives rise to

$$P_n(\mathbf{y}|\mathbf{x}_1) \le P_n(\mathbf{y} \mid \mathbf{x}_j) \qquad \text{for } j = 2, 3, \ldots, M-1$$

## Considering the selection of $\mathbf{x}_2$

**Random Code**



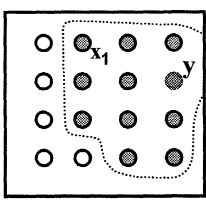

$\sum_{\{\mathbf{x}_2:\, P(\mathbf{y}|\mathbf{x}_1) \le P(\mathbf{y}|\mathbf{x}_2)\}} Q_n(\mathbf{x}_2)$

- With prob. $Q_n(\mathbf{x}_2)$, a particular location will be selected out of $2^n$ possible locations (allow a multiple selection for simplicity)
- If the selection $\mathbf{x}_2$ falls within the error region $E_2$, an error will occur.

©200x Heung-No Lee 43

## Chernoff Bounds and $\Pr\{E_2|m=1, \mathbf{x}_1, \mathbf{y}\}$

**Random Code**




Error region: $E_j = \{\mathbf{x}_j:\ 1 \le P_n(\mathbf{y}|\mathbf{x}_j)/P_n(\mathbf{y}|\mathbf{x}_1)\}$, for j=2, 3, …, M

No-error region: $E_j^c = \{\mathbf{x}_j:\ 1 > P_n(\mathbf{y}|\mathbf{x}_j)/P_n(\mathbf{y}|\mathbf{x}_1)\}$

- $\Pr\{E_2|m=1, \mathbf{x}_1, \mathbf{y}\} = \sum_{E_2} Q_n(\mathbf{x}_j)$

  --- now, apply the Chernoff bound

  $\le \sum_{\mathbf{x}_2} Q_n(\mathbf{x}_2)\, [P_n(\mathbf{y}|\mathbf{x}_2)/P_n(\mathbf{y}|\mathbf{x}_1)]^s$

  » We will choose the optimum $0<s<1$ later on

©200x Heung-No Lee 44

## Events of other selections are similarly obtained

- $E_j := \{\mathbf{x}_j: P_n(\mathbf{y}|\mathbf{x}_1) \le P_n(\mathbf{y}|\mathbf{x}_j)\}$, j=2, 3, …, M
- Thus, we have

$$\Pr\{E_j | m=1, \mathbf{x}_1, \mathbf{y}\} = \sum_{Ej} Q_n(\mathbf{x}_j)$$
$$\le \sum_{\mathbf{x}j} Q_n(\mathbf{x}_j)[P_n(\mathbf{y}|\mathbf{x}_j)/P_n(\mathbf{y}|\mathbf{x}_1)]^s$$

- This is the same for any j, thus we may drop the index on the right-hand side

$$\Pr\{E_j | m=1, \mathbf{x}_1, y\} \le \sum_{\mathbf{x}} Q_n(\mathbf{x})[P_n(\mathbf{y}|\mathbf{x})/P_n(\mathbf{y}|\mathbf{x}_1)]^s$$

 45

## Union Bounds

- $\Pr\{\text{error}| m=1, \mathbf{x}_1, \mathbf{y}\} \le \min\{1, \sum_{m\neq 1}^{M} \Pr(E_m| m=1, \mathbf{x}_1, \mathbf{y})\}$
- Then, for $0 < \rho \le 1$, we have

$$\sum_{m\neq 1}^{M} \Pr(E_m | m=1, \mathbf{x}_1, \mathbf{y})\} \le [\sum_{m\neq 1}^{M} \Pr(E_m| m=1, \mathbf{x}_1, \mathbf{y})]^{\rho}$$

  - Ex) see $0.9^{1/2} = 0.9487 > 0.9$, or see $0.9^{1/4} = 0.97$

- Thus, we have

$$\Pr\{\text{error}| m=1, \mathbf{x}_1, \mathbf{y}\} \le [\sum_{m\neq 1}^{M} \Pr(E_m| m=1, \mathbf{x}_1, \mathbf{y})]^{\rho}$$
$$= [\sum_{m\neq 1} \sum_{\mathbf{x}} Q_n(\mathbf{x})[P_n(\mathbf{y}|\mathbf{x})/P_n(\mathbf{y}|\mathbf{x}_1)]^s]^{\rho}$$
$$= [(M-1) \sum_{\mathbf{x}} Q_n(\mathbf{x})[P_n(\mathbf{y}|\mathbf{x})/P_n(\mathbf{y}|\mathbf{x}_1)]^s]^{\rho}$$

 46

## Combining All terms

- $P_{e,1} = \sum_{\mathbf{x}1} \sum_{\mathbf{y}} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}|\mathbf{x}_1)\, \Pr\{\text{error} \mid m=1, \mathbf{x}_1, \mathbf{y}\}$
  $\leq (M-1)^\rho \sum_{\mathbf{y}} \sum_{\mathbf{x}1} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}|\mathbf{x}_1) \times$
  $\{\sum_{\mathbf{x}} Q_n(\mathbf{x})\, [P_n(\mathbf{y}|\mathbf{x})/P_n(\mathbf{y}|\mathbf{x}_1)]^s\}^\rho$

$= (M-1)^\rho \sum_{\mathbf{y}} \sum_{\mathbf{x}1} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}|\mathbf{x}_1) \times$
$P_n(\mathbf{y}|\mathbf{x}_1)^{-s\rho} \cdot \{\sum_{\mathbf{x}} Q_n(\mathbf{x})\, P_n(\mathbf{y}|\mathbf{x})^s\}^\rho$

$= (M-1)^\rho \sum_{\mathbf{y}} [\sum_{\mathbf{x}_1} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}|\mathbf{x}_1)^{1-s\rho}]\{\sum_{\mathbf{x}} Q_n(\mathbf{x}) P_n(\mathbf{y}|\mathbf{x})^s\}^\rho$
---- (1)

Note that $\mathbf{x}_1$ also is a dummy

Let's try to minimize the RHS (blue)

 47

## Combining All terms

$P_{e,1} \leq (M-1)^\rho \sum_{\mathbf{y}} [\sum_{\mathbf{x}1} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}|\mathbf{x}_1)^{1-s\rho}]\{\sum_{\mathbf{x}} Q_n(\mathbf{x}) P_n(\mathbf{y}|\mathbf{x})^s\}^\rho$
---- (1)

Use Holder's Inequality $\sum_i a_i b_i \leq (\sum_i a_i^{1/q})^q [\sum_i b_i^{1/(1-q)}]^{1-q}$, the equality is achieved when $a_i^{1-q} = b_i^q$ for all i.

$[\sum_{\mathbf{x}} Q_n(\mathbf{x}) P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}]^{(1+\rho)} \leq [\sum_{\mathbf{x}1} Q_n(\mathbf{x}_1) P_n(\mathbf{y}|\mathbf{x}_1)^{1-s\rho}] \times [\sum_{\mathbf{x}} Q_n(\mathbf{x}) P_n(\mathbf{y}|\mathbf{x})^s]^\rho$

Consider $[\sum_{\mathbf{x}1} Q_n(\mathbf{x}_1) P_n(\mathbf{y}|\mathbf{x}_1)^{1-s\rho}]^{1/(1+\rho)} \times [\sum_{\mathbf{x}} Q_n(\mathbf{x}) P_n(\mathbf{y}|\mathbf{x})^s]^{\rho/(1+\rho)}$

Let $q = 1/(1+\rho)$ & $1 - q = \rho/(1+\rho)$,

$a_i^{1/q} = Q_n(x)\, P_n(y|x)^{1-s\rho} = [Q_n(x)^q\, P_n(y|x)^{(1-s\rho)q}]^{1/q}$,

$b_i^{1/(1-q)} = Q_n(x)\, P_n(y|x)^s = [Q_n(x)^{(1-q)}\, P_n(y|x)^{s(1-q)}]^{1/(1-q)}$,

$a_i b_i = Q_n(x)\, P_n(y|x)^{(1-s\rho)q + s - sq}$

the equality is achieved when $a_i^{1-q} = b_i^q$ for all i.

Thus, $s = (1-s\rho)$

 48

## Minimization of the RHS

- For a tighter bound, minimize the RHS wrt s
- Use the Holder's inequality
  - $\sum_i a_i b_i \le (\sum_i a_i^{1/s})^s [\sum_i b_i^{1/(1-s)}]^{1-s}$, --- (2)
    for $a_i \ge 0$ and $b_i \ge 0$ for all i, and $0 < s < 1$,
    with equality iff for some c, $a_i^{1-s} = c\, b_i^s$ for all i,
  - Special case is the Schwartz inequality (s=1/2)
- Applying the holder's inequality, we can prove that the RHS is minimized at $s = 1/(1+\rho)$ or $s = 1 - s\rho$, (See HW#5)
  $$[\sum_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}]^{(1+\rho)} \le [\sum_{\mathbf{x}1} Q_n(\mathbf{x}_1)P_n(\mathbf{y}|\mathbf{x}_1)^{1-s\rho}] \times [\sum_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}|\mathbf{x})^{s}]^{\rho}$$
  --- (3)

©200x Heung-No Lee 49

## Finally, we obtain the first coding theorem (or Gallager '65 paper Eq. (11))

- Continuing the derivation for $P_{e,1}$,

$$P_{e,1} \le (M-1)^{\rho} \sum_{y} [\sum_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}] \cdot [\sum_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}]^{\rho}$$

$$\boxed{P_{e,1} \le (M-1)^{\rho} \sum_{y} [\sum_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}]^{(1+\rho)}} \quad \text{--- (4)}$$

- Eq. (4) is the union bound for either memory or memoryless channels, and can be generalized to continuous channel outputs.
- Let's continue the derivation for the memory-less channel
  - Memory-less channel means that $P_n(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{n} P(y_i|x_i)$

©200x Heung-No Lee 50

## Signal and Channel-Output Alphabets

- $x_i \in$ input alphabet
  - Index k = 1, 2, …, K
- $y_i \in$ channel-output alphabet
  - Index j = 1, 2, …, J
- Let Q(k) be an arbitrary prob. distr. on the input
  - $Q_n(k) = \prod_{i=1}^{i} Q(x_i)$
- Using these indices, Eq. (4) can be rewritten

## Use Memory-less Channel: $P_n(\mathbf{y}|\mathbf{x}) = \prod P(y_i|x_i)$

- $P_{e,1} \le (M-1)^{\rho}\Sigma_{\mathbf{y}} [\Sigma_{\mathbf{x}} Q_n(\mathbf{x}) P_n(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}]^{\rho+1}$
  $= (M-1)^{\rho} \Sigma_{y1} \ldots \Sigma_{yn}[\Sigma_{x1} \ldots \Sigma_{xn} \prod_{i=1}^{n} Q(x_i) P(y_i|x_i)^{1/(1+\rho)}]^{\rho+1}$
  $= (M-1)^{\rho} \prod_{i=1}^{n} \Sigma_{yi} [\Sigma_{xi} Q(x_i) P(y_i|x_i)^{1/(1+\rho)}]^{\rho+1}$
  $= (M-1)^{\rho} \{ \Sigma_{j=1}^{J} [\Sigma_{k=1}^{K} Q(k) P(j \mid k)^{1/(1+\rho)}]^{\rho+1} \}^{n}$
- P(j | k) is the channel transition probability
- Q(k) is the prob. of selecting a symbol in the alphabet

- Ex) $n$=2, $j$=1,2, (ignore $k$ loop for this example)
  $\Sigma_{j1=1}^{2} \Sigma_{j2=1}^{2} Q(j_1)Q(j_2)P(y_1|j_1)^a P(y_2|j_2)^a = [\,\Sigma_j Q(j)P(y_1|j)^a\,]^2$

## Gallager's Random Coding Bounds for Memory-less channels

- We want the relationship between $R$ and $n$
  - If R closer to C, then n needs to be longer (for P(error) small)
- $R = \ln(M)/n$, [nats] (cf. $\log_2(M)/n$ [bits])
  - $M = e^{nR}$ or $ceil(e^{nR})$
- Continuing the derivation

$$P_{e,1} \le (M-1)^{\rho} \{ \textstyle\sum_{j=1}^{J} [\sum_{k=1}^{K} Q(k)\, P(j \mid k)^{1/(1+\rho)}]^{\rho+1} \}^{n}$$

$$< e^{nR\rho} \exp\{n \ln[\textstyle\sum_{j=1}^{J} [\sum_{k=1}^{K} Q(k)\, P(j \mid k)^{1/(1+\rho)}]^{\rho+1}]\}$$

-- $E_0(\rho, Q) := -\ln(\sum_{j=1}^{J} [\sum_{k=1}^{K} Q(k)\, P(j \mid k)^{1/(1+\rho)}]^{\rho+1})$

$$= e^{nR\rho} \exp\{-n\, E_0(\rho, Q)\}$$

droped 1, why?

$$\boxed{P_e \le \exp[\,-n\,(E_0(\rho, Q) - \rho R)\,]} \quad \text{--- (5)}$$

©200x Heung-No Lee 53

## Coding Theorem for Discrete Memoryless Channels (see Gallager's 1965 paper Eq. (19)-(22))

- Memoryless Channel: $P_n(\mathbf{y} \mid \mathbf{x}) = \prod_i P(y_i|x_i)$
- Discrete: the value of $y_i$ is quantized $\rightarrow$ $P(y_i|x_i) = P(j \mid k)$
- Theorem is

$$\boxed{P_e \le \exp(-n\, E(R)) \le \exp[-\, n(\, E_0(\rho, R) - \rho\, R\,)]} \quad \text{--- (6)}$$

- where $E_0(\rho, R) := -\log \sum_j [\sum_k Q(k)\, P(j|k)^{1/(1+\rho)}]^{1+\rho}$
- For a tighter bound, we maximize the exponent such that

$$E(R) = \max_{\rho} \max_{Q} (E_0(\rho, Q) - \rho R)$$

where $0 \le \rho \le 1$ and the selection of distribution{Q(1), Q(2), …, Q(K)}

©200x Heung-No Lee 54

## Implications of The Channel Coding Theorem

- $P_e$ can be made approach 0 exponentially fast as $n \uparrow \infty$
- This was the ensemble average over all possible codes. Thus, there exists at least one code whose error probability P(e) is smaller than or equal to the ensemble average $P_e := E\{P(e)\}$
- A random selection of a code will do a good job most of the time

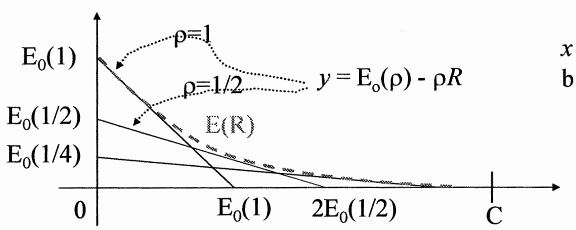
  $\forall\, a >> 1, \quad \Pr[P(e|C) > a\, P_e] < 1/a$

  (Markov Inequality: $a\, P_e\, I_{\{C\,:\,P(e|C) > a\, Pe\}} < P(e|C)$ )
- Finding a good code is not a problem, but finding a practical encoding and decoding scheme for such a code is.

 55

## Random Coding Exponent

$E(R) = \max_\rho \max_Q (E_0(\rho, Q) - \rho R)$



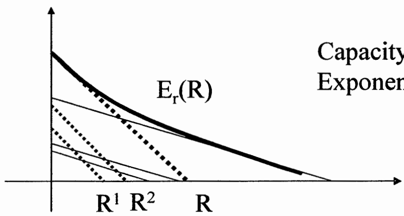

$x = R$
$b = E_0(\rho)$

- E(R) is a convex-U, decreasing, positive function of $R$, $0 \le R \le C$ (See section III of the paper)
  - For symmetric channel, the optimum Q is uniform.
  - Calculate $E_0(\rho, Q)$ for $0 \le \rho \le 1$.
  - Draw the tangential lines—from $(E_0(\rho)/\rho, 0)$ to $(0, E_0(\rho))$—for each $\rho$
  - Then, E(R) is the least upper bound of the tangential lines.

 56

## Interesting Example: Parallel Channels (Theorem 5 of the paper)

- Two parallel channels with two transitional prob. $P^1(j|k)$ and $P^2(q|p)$
- Consider the coding theorem for the combination channel where input is (k, p) and output is (j, q) with trans. prob. $P^1(j|k)P^2(q|p)$

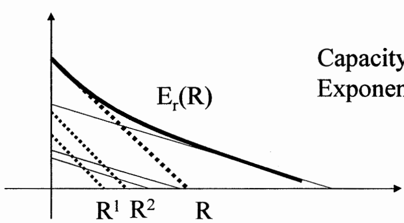



Capacity adds up
Exponents adds up

©200x Heung-No Lee 57

## Interesting Example: Parallel Channels (Theorem 5 of the paper)

- Two parallel channels with two transitional prob. $P^1(j|k)$ and $P^2(q|p)$
- Consider the coding theorem for the combination channel where input is (k, p) and output is (j, q) with trans. prob. $P^1(j|k)P^2(q|p)$




Capacity adds up
Exponents adds up

©200x Heung-No Lee 58

## Interesting Example (2): (Proakis 7.1-52)

- Recall that we have derived the P(e) union bound for M-ary orthogonal signals in Lecture-4 over (infinite bandwidth) AWGN channels
- $P_M \le 2^k \exp(-k\, E_b/2N_o)$

  $= 2^k\, 2^{\log2(\exp(-kEb/2No))} = 2^k\, 2^{-k\, Eb/2No\, \log2(e)}$

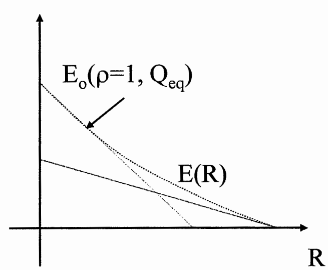
  $= 2^k\, 2^{-k\,(1/2)\, Eb/(No\, \ln(2))} = 2^{-k[0.5\, Eb/(No\, \ln(2)) - 1]}$

  - where $k = \log_2(M)$, where M is again the size of the message set
  - R = k/T, bits/channel use, T = duration of signal in time
  - Thus, k = TR.
  - Note that the infinite bandwidth capacity
    - $C_\infty := \lim_{W\to\infty} W \log_2[1 + P/(N_o W)]$ (See Proakis pg. 386)
    - $C_\infty = RE_b/(N_o \ln(2)) = P/(N_o \ln(2))$
- $P_M \le 2^{-T[0.5C\infty - R]}$

©200x Heung-No Lee 59

## Example (3): Binary Symmetric Channels

- BSC with cross-over prob. p
- Equally likely Q is optimum




©200x Heung-No Lee 60

## Other Use of Random Coding Bounds

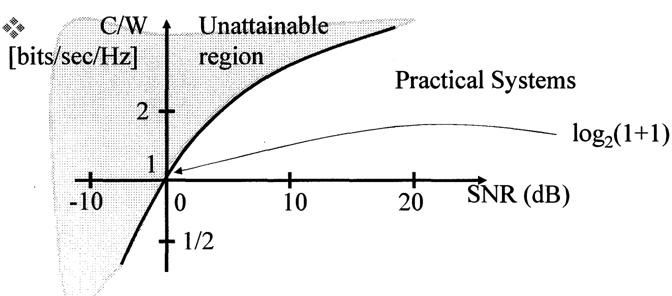
- Criteria for Code Construction
- Select a code at random and discard codewords which results in large $P_e$
  - Not very useful when M is very large
- We will see Gallager's example next week on constructing a random ensemble code (LDPC code)

©200x Heung-No Lee 61

## Shannon Capacity (2)

- The channel capacity in [bits/sec] is

$$C = W\,log_2(1+\frac{P}{N}) = W\,log_2(1+\frac{P}{N_oW})$$

- 


©2004 Heung-no Lee 62

## Shannon Limit

- The bit rate of the transmission
  R [bits/sec] = $R_s$ [symbols/sec] $\times$ $k$ [bits/symbol]
- Signal Power P = Energy per symbol $\times$ Baud = $E_s \times R_s$
- Energy per bit $E_b = E_s/k$
- $P = E_b \times R$
- $P/N = (E_b R)/(N_o W)$
- Thus, the spectral efficiency [C/W] is

$$C/W = \log_2[1+(E_b/N_o)(R/W)] \qquad \text{---------(1)}$$

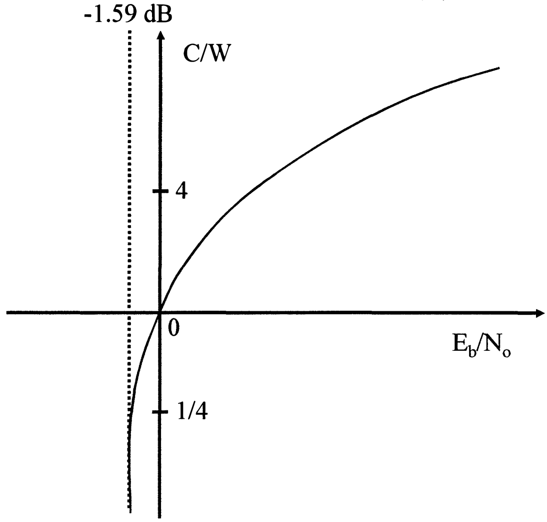
 63

## Shannon Limit (2)

- We are interested in finding out the smallest $E_b/N_o$ that error free transmission is possible
- In Eq. (1), substitute R for C in the right side of Eq.
- Then, we have
  $C/W = \log_2(1+(E_b/N_o)(C/W))$,
- Arranging it for $E_b/N_o$, we have
  $E_b/N_o = (W/C)(2^{C/W} - 1)$
- Let $x:=(C/W) \rightarrow 0$
- $E_b/N_o = \lim_{x \rightarrow 0} (1/x)(2^x - 1) = \log_e 2 = 0.693 = -1.6$ dB
- This is the ultimate limit below which no error-free transmission is possible no matter how small R/W we may choose

 64

## Shannon Limit (3)




 65

## Codes that Achieve the Shannon Limit Very Closely

- Turbo code (1993)
- Low Density Parity Check Code (1963, 1995)
- Both approaches Shannon limit within a fraction of dB over AWGN channels

 66

## Gallager Bounds

Motivation: Gallager bounds are good for Maximum Likelihood performance evaluation. They lead to tight performance bounds.

Purpose of this note: Derive Maximum Likelihood performance evaluation for AWGN channel.

Let

- M is the size of the codebook. Use $m$ as the index.
- P($m$) is the probability of sending an index $m$.
- $Q_n(\mathbf{x}_m)$ be the selection probability that an $n$-tuple vector $\mathbf{x}_m$ is selected to carry the message index $m$.
- The received vector is $\mathbf{y} = \mathbf{x}_m + \mathbf{w}$ where $\mathbf{w}$ is multivariate Gaussian noise whose mean is zero and $Cov(\mathbf{w}) = \mathbf{R}_{\mathbf{w}}$.
- $P_n(\mathbf{y} \mid \mathbf{x}_m)$ be the likelihood function for message index $m$.

The probability of error can be calculated by

$$
\begin{aligned}
P_e &= \Pr\{error\} \\
&= \sum_{m=1}^{M} P(m)\Pr\{error \mid m\}
\end{aligned}
\tag{1.1}
$$

Assume equally likely transmission of symbol index, i.e., $P(m) = 1/M$.

We assume the first codeword $\mathbf{x}_1$ is selected and sent and $\mathbf{y}$ is observed as the result. Then, we would like to evaluate the probability of error, which can be written by

$$
\begin{aligned}
P_e &= \Pr\{error \mid m = 1\} \\
&= \int_{\mathbf{x}_1}\int_{\mathbf{y}} \Pr\{e \mid m = 1, \mathbf{x}_1, \mathbf{y}\} Q_n(\mathbf{x}_1) P_n(\mathbf{y} \mid \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{y}
\end{aligned}
\tag{1.2}
$$

We note that error happens when we choose a second codeword $\mathbf{x}_2$, or any other codeword, closer to $\mathbf{y}$ than $\mathbf{x}_1$ is to $\mathbf{y}$. The same goes for the other codeword selections.

Now working on the first part and use the union bound idea, we may write the first part as

$$
\Pr\{e \mid m = 1, \mathbf{x}_1, \mathbf{y}\} \le (M-1)\int_{\mathbf{x}_2 : \frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)} \le 1} Q_n(\mathbf{x}_2) d\mathbf{x}_2
\tag{1.3}
$$

We would like the upper bound to be tight. We note that the L.H.S. can grow as big as only up to 1 since it's a probability. But the union upper bound is not a probability and thus can grow unbounded as $M$ grows. That is, $M$ can be very large while the pairwise error could be a small number less than 1. Thus, we would like to make an upper bound which would give us the following effect

$$\Pr\{e \mid m=1,\mathbf{x}_1,\mathbf{y}\} = \min\left\{1,(M-1)\int_{\mathbf{x}_2:\frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)}\le 1} Q_n(\mathbf{x}_2)d\mathbf{x}_2\right\} \quad (1.4)$$

This effect can be made by raising the union bound by a power of $\rho \ge 0$:

$$\Pr\{e \mid m=1,\mathbf{x}_1,\mathbf{y}\} \le \left[(M-1)\int_{\mathbf{x}_2:\frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)}\le 1} Q_n(\mathbf{x}_2)d\mathbf{x}_2\right]^{\rho} \quad (1.5)$$

We choose $0 \le \rho \le 1$. When $(M-1)\int_{\mathbf{x}_2:\frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)}\le 1} Q_n(\mathbf{x}_2)d\mathbf{x}_2 \ge 1$, we can choose $\rho$ very close to 0 so that the value stays close to 1. For example, $100^{0.0001} \approx 1.00046$. When $(M-1)\int_{\mathbf{x}_2:\frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)}\le 1} Q_n(\mathbf{x}_2)d\mathbf{x}_2 < 1$, on the other hand, we can choose $\rho$ arbitrarily close to 1 so that it can stay smaller than 1. For example, $0.001^{0.99} \approx 0.00107$.

On the other hand, the selected integration in the union bound is difficult to evaluate. Thus, we use the idea of Chernoff bound. That is, we replace each integration with a selected geometry with the following one:

$$\int_{\mathbf{x}_2:\frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)}\le 1} Q_n(\mathbf{x}_2)d\mathbf{x}_2 \le \int Q_n(\mathbf{x}_2)\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)}d\mathbf{x}_2 \quad (1.6)$$

Note that the integration on the R.H.S. is not restricted to any region. This bound can be tightened by raising the ratio with power of $s \ge 0$. That is, we have

$$\int_{\mathbf{x}_2:\frac{P_n(\mathbf{y}|\mathbf{x}_1)}{P_n(\mathbf{y}|\mathbf{x}_2)}\le 1} Q_n(\mathbf{x}_2)d\mathbf{x}_2 \le \int Q_n(\mathbf{x}_2)\left(\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)}\right)^s d\mathbf{x}_2 \quad (1.7)$$

By choosing a right value $s$, we can make the ratio to work for us as if it is an indicator function. When $\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)} > 1$, we can choose $s$ very close to 0, i.e., $\left(\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)}\right)^s \xrightarrow{s \to 0} 1$. On the other hand, when $\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)} < 1$, we can choose $s$ close to infinity, i.e., $\left(\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)}\right)^s \xrightarrow{s \to \infty} 0$. Thus, we can tighten the bound, i.e. making the R.H.S. of (1.6) arbitrarily close to the L.H.S. of (1.6).

Now we substitute (1.7) into (1.5). Thus, we have

$$\Pr\{e \mid m=1,\mathbf{x},\mathbf{y}\} \le \left[(M-1)\int Q_n(\mathbf{x}_2)\left(\frac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)}\right)^s d\mathbf{x}_2\right]^{\rho} \quad (1.8)$$

Further substituting (1.8) in to (1.2), we have

$$
\begin{aligned}
P_e &\le \int_{\mathbf{x}_1}\int_{\mathbf{y}} Q_n(\mathbf{x}_1)P_n(\mathbf{y}\,|\,\mathbf{x}_1)\left[(M-1)\int Q_n(\mathbf{x}_2)\left(\tfrac{P_n(\mathbf{y}|\mathbf{x}_2)}{P_n(\mathbf{y}|\mathbf{x}_1)}\right)^s d\mathbf{x}_2\right]^\rho d\mathbf{x}_1 d\mathbf{y} \\
&= (M-1)^\rho \int_{\mathbf{y}}\int_{\mathbf{x}_1} Q_n(\mathbf{x}_1)P_n(\mathbf{y}\,|\,\mathbf{x}_1)^{1-s\rho} d\mathbf{x}_1 \left[\int Q_n(\mathbf{x}_2)P_n(\mathbf{y}\,|\,\mathbf{x}_2)^s d\mathbf{x}_2\right]^\rho d\mathbf{y}
\end{aligned} \tag{1.9}
$$

Now note $\mathbf{x}_1$ and $\mathbf{x}_2$ are dummy variables each. Thus we can write

$$
P_e \le (M-1)^\rho \int_{\mathbf{y}}\left[\int_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{1-s\rho} d\mathbf{x}\right]\left[\int Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^s d\mathbf{x}\right]^\rho d\mathbf{y} \tag{1.10}
$$

By choosing the optimal values of $s$ and $\rho$, one can tighten the upper bound (make the smallest upper bound). In Lemma, we have done such an optimization. We use the Holder's inequality and show that when $s = \dfrac{1}{1+\rho}$, the R.H.S. of (1.10) is minimized. With this result, the upper bound can then be written as

$$
\begin{aligned}
P_e &\le (M-1)^\rho \int_{\mathbf{y}}\left[\int_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]\left[\int Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^\rho d\mathbf{y} \\
&= (M-1)^\rho \int_{\mathbf{y}}\left[\int Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{\rho+1} d\mathbf{y}
\end{aligned} \tag{1.11}
$$

Now it should be noted that the value $s$ is restricted to the interval from 1/2 to 1 since $0 \le \rho \le 1$. Note that this bound is for vector input vector output channel and thus is very general.

**Lemma**. We want to show that the integration $\left(\left[\int_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{1-s\rho} d\mathbf{x}\right]\left[\int Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^s d\mathbf{x}\right]^\rho\right)$ is minimized when $s = \dfrac{1}{1+\rho}$.

Proof: We note that $\left(\left[\int_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{1-s\rho} d\mathbf{x}\right]\left[\int Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^s d\mathbf{x}\right]^\rho\right)$ is minimized when $\left(\left[\int_{\mathbf{x}} Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{1-s\rho} d\mathbf{x}\right]\left[\int Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^s d\mathbf{x}\right]^\rho\right)^{\frac{1}{1+\rho}}$ is minimized. Then, by applying the Holder's inequality, $\int_{\mathbf{x}} a(\mathbf{x})b(\mathbf{x})d\mathbf{x} \le \left[\int_{\mathbf{x}} a^{1+\rho}(\mathbf{x})d\mathbf{x}\right]^{\frac{1}{1+\rho}}\left[\int b^{\frac{1+\rho}{\rho}}(\mathbf{x})d\mathbf{x}\right]^{\frac{\rho}{1+\rho}}$ with the equality achieved when $a^{1-h}(\mathbf{x}) = c\, b^h(\mathbf{x})$, we notice

- $h = \dfrac{1}{(1+\rho)}$ and $1-h = \dfrac{\rho}{(1+\rho)}$
- $a^{1/h}(\mathbf{x}) = Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{1-s\rho}$
- $b^{1/(1-h)}(\mathbf{x}) = Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{s}$

Then, we have

- $a(\mathbf{x}) = \left(Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{1-s\rho}\right)^h = Q_n^{\frac{1}{1+\rho}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1-s\rho}{1+\rho}}$
- $b(\mathbf{x}) = \left(Q_n(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{s}\right)^{1-h} = Q_n^{\frac{\rho}{1+\rho}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{s\left(\frac{\rho}{1+\rho}\right)}$

By equating $a^{1-h}(\mathbf{x}) = b^h(\mathbf{x})$, the inequality is met with equality. We have

$$\left(Q_n^{\frac{1}{1+\rho}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1-s\rho}{1+\rho}}\right)^{\frac{\rho}{(1+\rho)}} = \left(Q_n^{\frac{\rho}{1+\rho}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{s\left(\frac{\rho}{1+\rho}\right)}\right)^{\frac{1}{(1+\rho)}} \tag{1.12}$$

which leads to

$$\left(Q_n^{\frac{1}{1+\rho}\frac{\rho}{(1+\rho)}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1-s\rho}{1+\rho}\frac{\rho}{(1+\rho)}}\right) = \left(Q_n^{\frac{\rho}{1+\rho}\frac{1}{(1+\rho)}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{s\left(\frac{\rho}{1+\rho}\right)\frac{1}{(1+\rho)}}\right) \tag{1.13}$$

Equating the exponent of $P_n(\mathbf{y}\,|\,\mathbf{x})$, we have

$$\frac{1-s\rho}{1+\rho}\frac{\rho}{(1+\rho)} = s\left(\frac{\rho}{1+\rho}\right)\frac{1}{(1+\rho)} \tag{1.14}$$

Thus, we have

$$1-s\rho = s \tag{1.15}$$

Thus, the optimal value $s$ is $s = \dfrac{1}{1+\rho}$.

Then, the L.H.S. of the Holder's inequality is

$$\int_{\mathbf{x}} a(\mathbf{x})b(\mathbf{x})d\mathbf{x} = \int_{\mathbf{x}} Q_n^{\frac{1}{1+\rho}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{1-s\rho}{1+\rho}} Q_n^{\frac{\rho}{1+\rho}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{s\left(\frac{\rho}{1+\rho}\right)} d\mathbf{x}$$
$$= \int_{\mathbf{x}} Q_n^{\frac{\rho}{(1+\rho)^2}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{s\rho(1-s\rho)}{(1+\rho)^2}} d\mathbf{x} \tag{1.16}$$
$$\geq \left( \int_{\mathbf{x}} Q_n^{\frac{\rho}{(1+\rho)}}(\mathbf{x})P_n(\mathbf{y}\,|\,\mathbf{x})^{\frac{s\rho(1-s\rho)}{(1+\rho)}} d\mathbf{x} \right)^{\frac{1}{(1+\rho)}}$$

In fact, we do not need to obtain the expression for the L.H.S. of the Holder's inequality for this problem. We can obtain the minimum by substituting $s = \dfrac{1}{1+\rho}$ to the R.H.S. of the given problem.

**End of Proof**

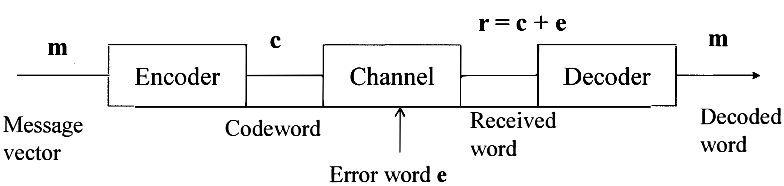
End of Document

# Linear Block Codes

1



## Coded System




- **m**, **c**, **r** and **m'** are vectors: From Shannon's work, we know block processing of information over noisy channel helps, rather than processing them in a bit-by-bit manner.
- As the size of the block increases, the larger is the potential to achieving the capacity; but the greater is the difficulty in decoding.

## $(n, k)$ Block Codes

- The block length $n$.
- The message length $k$.

- The code is $q$-ary, if each tuple is taking a value from GF($q$).
  - This course we only consider binary GF($q$=2) codes.

- Rate of the code $R = \log_q(M)/n$
- Redundancy $r = n - \log_q(M)$

3



## $(n, k)$ Linear Block Code

- A linear code $\mathcal{C}$ spans a vector space with dimension $k$. *It is a collection of M codewords.*

- *Linearity*: For any $a, b \in$ GF($q$) and any $\mathbf{v}, \mathbf{u} \in \mathcal{C}$, $a\mathbf{v} \in \mathcal{C}$ and $a\mathbf{v}+b\mathbf{u} = \mathbf{c} \in \mathcal{C}$.
  - If $\mathbf{c}$ is a codeword, $0\mathbf{c} = \mathbf{0}$ is a codeword.
  - Let $d(\mathbf{v}, \mathbf{u})$ denote Hamming distance between any two different codewords $\mathbf{v}, \mathbf{u} \in \mathcal{C}$ and $w(\mathbf{v})$ Hamming weight of codeword $\mathbf{v}$ respectively.
  - Then, $d_{min} = \min d(\mathbf{v}, \mathbf{u})$
    $= \min w(\mathbf{v} + \mathbf{u})$
    $= \min w(\mathbf{c}=\mathbf{v} + \mathbf{u})$
    $= \min w(\mathbf{c})$, over all non-zero $\mathbf{c} \in \mathcal{C}$
  - It is the minimum weight of non-zero codeword.

## A linear code = a generator matrix = a parity check matrix

- A linear block code can be defined by either a generator matrix or a parity-check matrix.
- Generator matrix is obtained by $k$ linearly independent codewords.
  - The rows of **G**, generator matrix [$nR$ x $n$] of a code, span the code space
- The rows of **H**, parity check matrix [$n(1\text{-}R)$ x $n$], span the linear space perpendicular to the row space of G.
  - There are $n(1\text{-}R)$ number of simultaneous linear homogeneous parity check equations.
  - There are $n(1\text{-}R)$ rows of **H** which span the null space of the code.

$$\mathbf{G}\mathbf{H}^T = 0$$

5



## Syndrome

- Given the I/O relation $\mathbf{r} = \mathbf{c} + \mathbf{e}$, a *syndrome* vector **s** is obtained by

$$\mathbf{s} = \mathbf{r}\mathbf{H}^T = (\mathbf{c}+\mathbf{e})\mathrm{H}^T = \mathbf{e}\mathbf{H}^T$$

  - Non-zero $s$ indicates "there is a problem."

6

## Distance Spectrum

- Hamming weight of a codeword is the number of non-zero coordinates.
- $A_h$ is the number of codewords with weight $h$, $h$=0, 1, 2, …, $n$, in a code $C$.
- Distance spectrum $\{A_h, h=0, 1, 2, \ldots, n\}$ is a collection of $A_h$.
- Polynomial representation is useful.
  - $A(z) = \sum_{h=0}^{n} A_h z^h$
  - $A(z)\big|_{z=1} = \sum_{h=0}^{n} A_h = 2^k$

7

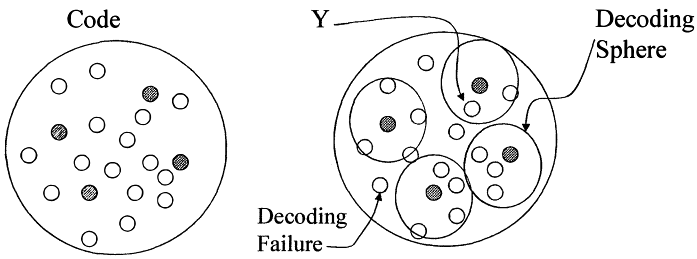


## Minimum Distance of a Linear Code

- $d_{min}$ of a linear code $C$ is the smallest weight of codeword in a code.
  - It is the minimum distance between any two codewords in a code
    $d_{min} = \min_{\mathbf{c}_i, \mathbf{c}_j \in C} d(\mathbf{c}_i, \mathbf{c}_j)$ (This definition is most general and works for non linear codes as well)
  - With respect to distance spectrum, it can also be written as
    $d_{min} = \min\{h = 1, 2, \ldots, n: A_h \neq 0\}$
  - It can also be obtained by investigating the parity check matrix **H**.
    - $d_{min}$ is the smallest number $d$ such that there exists $d$ linearly dependent columns of **H**.

8

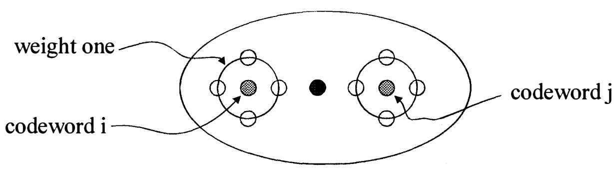
## Max. Likelihood Decoding




- Encoding: **m→ c**, it is a mapping from a block of message bits to a codeword.
- ML Decoding: make decision in favor of a message index *m* that maximizes Pr{Y|*m*}.
  - This leads to a minimum distance decoding rule.
  - Minimum distance error event dominates the error performance.



## Minimum Distance Decoding and Correctable Errors




- Note $d_{min}$ = 4. The code can detect all error patterns up to weight 3.
- MLD = minimum distance decoding.
- If the blue ball is a received word, it leads to decoding failure because it has the same distance to codeword *i* and to codeword *j*.
- **A code with $d_{min}$ can correct all error patterns of weight <= *floor*(($d_{min}$- 1)/2)**

## Error Correction and Detection Capability of a Code

- A code can correct up to all $t$ error patterns if

$$t \le \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor$$

- A code can detect up to all $d_{min} - 1$ error patterns.

- A code can correct up to $e_1$ error patterns and $e_2$ erasures if
  - $e_1 + e_2 \le d_{\min} - 1$
    and
  - $e_1 \le e_2$

11



## Simultaneous Correction & Detection

- Example with $d_{min} = 7$
- Correct 0 error and detect 6 errors
- Correct 1 error and detect 5 errors
- Correct 2 errors and detect 4 errors
- Correct 3 errors and detect 3 errors
- $d_{min} = e_1 + e_2 + 1$ and $e_2 >= e_1$

## Syndrome Decoding (2)

22

- There are $2^{n-k}$ syndromes
- Example : (5,2) code

  Let $H^T = \begin{bmatrix} P \\ I_3 \end{bmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and

  and let $\underline{r} = (\underbrace{1\,0}_{\text{info bits}}\ \underbrace{1\,1\,1}_{\text{parity check bits}}) \oplus (\underbrace{0\,1\,0\,0\,0}_{\text{error pattern}}) = (\underbrace{1\,1\,1\,1\,1}_{\text{rec. word}})$

  $\underline{s} = \underline{r}\,H^T = (1\,1 \vdots 1\,1\,1)\,H^T = [1\;1]\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \oplus [1\;1\;1]\,I_3$

  $= \underset{\substack{\uparrow \\ \text{Generated} \\ \text{Parity} \\ \text{Checks}}}{(0\,0\,1)} \oplus \underset{\substack{\uparrow \\ \text{Rec.} \\ \text{Parity} \\ \text{checks}}}{(1\,1\,1)} = \underset{\substack{\uparrow \\ \text{Syndrome}}}{(1\,1\,0)}$

  Or $\underline{s} = \underline{e}\,H^T$

  $= (0\,1\,0\,0\,0)\,H^T = 1\,1\,0 \Leftarrow$ store this in table look up.

  Correct error by adding the error pattern to $\underline{r}$

©200x Heung-No Lee

## Syndrome Decoding (3)

23

- Choose $2^{n-k}$ most likely error patterns
  - start from smallest weight patterns
- Find all syndromes of most likely error patterns
- Construct a table look-up with k columns and $2^{n-k}$ rows ( First syndrom, Second Error Patten)
- Decoding Steps
  1 – Find syndrome $\underline{s} = \underline{r}\,H^T$
  2 – Use table (look up) to find the error pattern for the calculated syndrome in 1.
  3.– Add the correctable error pattern to the rec. word

©200x Heung-No Lee

## Standard Array

- It is a $[2^{(n-k)} \times 2^k]$ table of words on length n.
- All $2^n$ words appear once and only once.
- It shows decoding cells.
- The first row lists up all possible $2^k$ codewords.
- The first column lists up all $2^{(n-k)}$ distinct error patterns **e** that can be recovered.

15



## Standard Array for (5, 2) LBC

| Info bits | 00 | 01 | 10 | 11 | Syndrome |
|---|---|---|---|---|---|
| Codewords | 00000 | 01110 | 10111 | 11001 | 000 |
| Correctable Single Error Patterns | 00001 | 01111 | 10110 | 11000 | 001 |
| | 00010 | 01100 | 10101 | 11011 | 010 |
| | 00100 | 01010 | 10011 | 11101 | 100 |
| | 01000 | 00110 | 11111 | 10001 | 110 |
| | 10000 | 11110 | 00111 | 01001 | 111 |
| Correctable Error Patterns (w=2) | 10100 | 11010 | 00011 | 01101 | 011 |
| | 10010 | 11100 | 00101 | 01011 | 101 |

$$H^T = \begin{pmatrix} 111 \\ 110 \\ 100 \\ 010 \\ 001 \end{pmatrix}$$

BSC

- Construct it row by row by adding the error pattern (the 1st column element) to the codewords.
  - Make sure not to select the error pattern which have appeared already
- When choosing the error patterns (1st column), make sure they lead to distinct syndrome.

16

## $d_{min}$ and Detectable Errors

- Received = codeword(i) + error pattern
- Undetectable error patterns:
  - When the error pattern itself is a codeword, then the received is a codeword but is not the transmitted codeword (i).
  - This is an undetectable error event.
  - It happens IFF the error pattern itself is a codeword.
  - The syndrome is zero in such a case.
  - Minimum weight of such error patterns is $d_{min}$.
- An error pattern with weight less than $d_{min}$ can always be detectable.
- **A code with $d_{min}$ can detect all error patterns of weight <= $d_{min}$-1**



## The Singleton Bound (Linear Codes)

- The minimum distance $d_{min}$ of a linear $(n, k)$ code is bounded from above by $d_{min} \leq n - k + 1$.
  - An $(n, k)$ code has a parity matrix which contains $(n - k)$ linearly independent rows.
  - The dimension of row space, and thus that of the column space, is $(n - k)$.
  - Thus, any collection of $(n - k) + 1$ columns of **H** has to be linearly dependent.

## Example

Example

$$H = \begin{array}{c} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \\ \begin{matrix} H^1 & H^2 & H^3 & H^4 & H^5 \end{matrix} \end{array}$$

- the column vectors of H span a 3-dim. vector space.
- note $H^1 = H^3 + H^4 + H^5$

  $H^2 = H^3 + H^4$

  $d_{min} = 3$

$n - k + 1 = 5 - 2 + 1 = 4$



Example of Hamming Code (7,4)

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

true in general

- Note every pair of columns of H is independent.
- But, some collection of 3 columns are dependent.

  $\Rightarrow d_{min} = 3$

## Decoding of (7, 4) Hamming Code

- Compute the syndrome s = r $H^T$ = e $H^T$
- Can correct all error patterns of weight = 1
- For a single error occurred at *j*-th coordinate, the syndrome is equal to the *j*-th column of H
- Thus, decoding step is
  - Compute the syndrome. If zero, then the rec. word is a codeword.
  - If not equal to zero, examine if any match can be found from columns of H. Record the index j of the column.
  - Complement the *j*-th bit of the received word.



④

### Decoding of Hamming Code

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\underline{r} = (1\ 0\ 1\ 0\ 1\ 1\ 0)$$

$$\underline{s} = \underline{r}\, H^T = (0\ 1\ 1)$$

$\therefore$ Error correction $\underline{r} + (0\ 0\ 1\ 0\ 0\ 0\ 0)$

$$= \underbrace{(1\ 0\ 0\ 0\ 1\ 1\ 0)}$$

Decided codeword

# HW#5

- P.1 (Gallager bound for binary input AWGN channel) Let's consider a binary input {+1, -1} AWGN channel. The PSD of the noise is $N_o/2$.
  - (a) Obtain Gallager random coding bound, similar to lecture note, for this channel.
  - (b) Obtain the error exponent expression of this channel . What type of Q is desirable? Why?
  - (c) Obtain the expression for error exponent E(R) and sketch it.

- Proakis/Salehi:
  - P7.13
  - P7.25
  - P7.28
  - P7.29
  - P7.32
  - P7.33
  - P7.34

23

# LDPC Codes



## Review on Linear Code

- Here we define $R$ as the rate of a linear block code.
- The block length is $n$.
- H, parity check matrix [$n$(1-$R$) x $n$]
  - $n$(1-$R$) rows, [$n$ x 1], of H span the null space of the code
- G, generator matrix [$nR$ x $n$] of a code
  - Row-space of G is space of the code
- $GH^T = 0$
- There are $n$(1-$R$) number of simultaneous linear homogeneous parity check equations.
- $rH^T = (c+e)H^T = eH^T = s$
  - Non-zero $s$ indicates "problem"

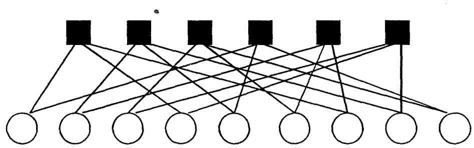
## Gallager's Thesis ('63)

- ($n, j, k$) low density parity check code.
- Parity check matrix $H$ [$n(1-R)$ x $n$] of the code has
  - $j$ number of 1's in each column
  - $k$ number of 1's in each row
  - The code rate is R = 1 - j/k
- Min. distance of a typical (n, j, k) code for j $\geq$ 3
  - increases linearly with $n$ for fixed $j$ & $k$.
- Practical decoding method exists
  - Simple or probabilistic

Heung-No Lee 3

## Parity Check Matrix on Bipartite Graph

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_6 \end{bmatrix}$$



Heung-No Lee 4

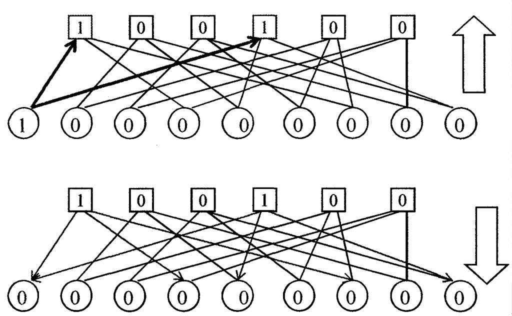
## Illustration of Decoding Concept with Simple Hard-Decision Decoder

- Works only for BSC
- Compute all parity checks
- Change the digit involved in more than a fixed number of unsatisfied parity checks
- Re-compute all parity checks
- Repeat

Heung-No Lee 5

## Simple Decoding Example (j=2, k=3)

- Suppose we have r=[1 0 0 0 0 0 0 0 0]
- Violates check nodes (equations) 1 and 4
- Check nodes 1 and 4 send down instruction to correct
- Error-node 1 corrects it, having two instructions
- Error-nodes 4, 5, 7 & 9 do not correct, since it has only one instruction to correct




Heung-No Lee 6

# Probabilistic Decoding

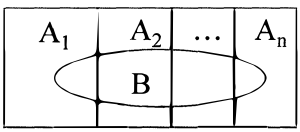
- Bayes' Theorem on The Total Probability
- Total Probability: If A={$A_1$, $A_2$, …, $A_n$} is a partition of S and B is an arbitrary event

$$\Pr\{B\} = \Sigma^{n}_{i=1}\Pr\{B \cap A_i\} = \Sigma^{n}_{i=1}\Pr\{B \mid A_i\} \Pr\{A_i\}$$

- Bayes' Theorem: We know

$$Pr\{A_i|B\} = \frac{Pr\{A_i \cap B\}}{Pr\{B\}}$$

- The aposteriori probability is then given by




Likelihood

$$\underbrace{Pr(A_i|B)}_{\text{aposteriori}} = \frac{Pr(B|A_i)Pr(A_i)}{\sum_{i=1}^{n} Pr(B|A_i)Pr(A_i)}$$

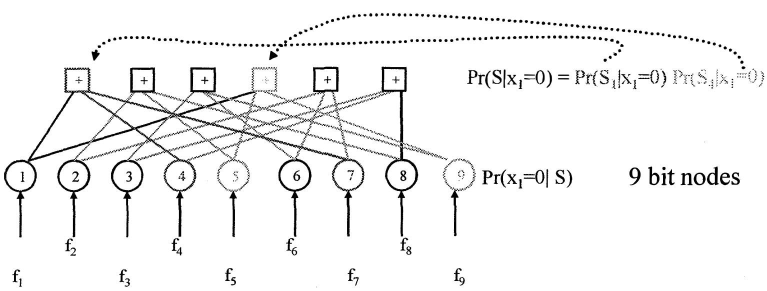
aposteriori

prior

Heung-No Lee

7

# The Iterative Graph Decoding Theorem




Get the input probabilities
from the channel output $y_k = (2x_k - 1) + n_k$, $n_k \sim \mathcal{N}(0, N_0/(2E_s))$

Gallager's Decoding Theorem

$$\frac{Pr(x_d = 0|y, S)}{Pr(x_d = 1|y, S)} = \frac{1 - Pr(x_d = 1|\ y)}{\underbrace{Pr(x_d = 1|\ y)}_{p_d :=}} \prod_{i=1}^{j} \frac{1 + \prod_{l=1}^{k-1}(1 - 2p_{il})}{1 - \prod_{l=1}^{k-1}(1 - 2p_{il})}$$

$$= \frac{1 - p_d}{p_d} \prod_{i=1}^{j} \frac{1 + \prod_{l=1}^{k-1}(1 - 2p_{il})}{1 - \prod_{l=1}^{k-1}(1 - 2p_{il})}$$

Heung-No Lee

## Decoding Theorem

- Bayes Theorem:

$$\begin{aligned} Pr(x_d = 1|\mathrm{y}, S) &= \frac{Pr(x_d = 1, \mathrm{y}, S)}{p(\mathrm{y}, S)} \\ &= \frac{Pr(S|x_d = 1, \mathrm{y})\ p(x_d = 1, \mathrm{y})}{p(\mathrm{y}, S)} \\ &= \frac{Pr(S|x_d = 1, \mathrm{y})\ Pr(x_d = 1|\mathrm{y})\ p(\mathrm{y})}{p(\mathrm{y}, S)} \end{aligned}$$

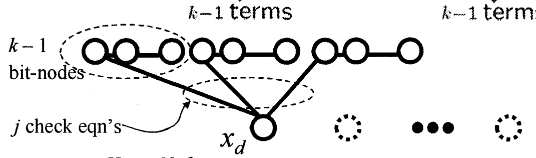
- The ratio is of our interest

$$\frac{Pr(x_d = 0|\mathrm{y}, S)}{Pr(x_d = 1|\mathrm{y}, S)} = \frac{Pr(S|x_d = 0, \mathrm{y})\ Pr(x_d = 0|\mathrm{y})}{Pr(S|x_d = 1, \mathrm{y})\ Pr(x_d = 1|\mathrm{y})}$$

Heung-No Lee 9

## Decoding Principle

- A posteriori probability: $Pr(x_d = 1|\mathrm{y}, S)$
  - Event S: All participating check equations are satisfied.
  - Event {y}: Observed output of the channel.
- First, let us think about a codeword in a sub-code.
  - A sub-code is a collection of codewords which satisfy all *j* parity checks.
  - Each of the *j* parity check equations involves (*k*-1) bit nodes.
  - Note a codeword in this sub-code is comprised of *j**(*k*-1)+1 bit nodes.
  - A codeword $\mathrm{c} = (x_d, \underbrace{x^1_{d_1}, x^1_{d_2}, \ldots, x^1_{d_{k-1}}}_{k-1 \text{ terms}}, \cdots, \underbrace{x^j_{d_1}, x^j_{d_2}, \ldots, x^j_{d_{k-1}}}_{k-1 \text{ terms}})$




Heung-No Lee 10

## Decoding

- Let $S_i$ : = {the $i$-th check is satisfied}
- Then, $S = S_1$ and $S_2$ and … and $S_j$
- Thus, we have

$$\begin{aligned} Pr(S|x_d = 0, \mathrm{y}) &= \prod_{i=1}^{j} Pr(S_i|x_d = 0, \mathrm{y}) \\ &= \prod_{i=1}^{j} \frac{1 + \prod_{l=1}^{k-1}(1 - 2p_{il})}{2} \end{aligned}$$

- Similarly, we have

$$Pr(S|x_d = 1, \mathrm{y}) = \prod_{i=1}^{j} \frac{1 - \prod_{l=1}^{k-1}(1 - 2p_{il})}{2}$$

Heung-No Lee 11

## Decoding Theorem

- The decoding theorem is

$$\begin{aligned} \frac{Pr(x_d = 0|\mathrm{y}, S)}{Pr(x_d = 1|\mathrm{y}, S)} &= \frac{1 - Pr(x_d = 1|\ \mathrm{y})}{\underbrace{Pr(x_d = 1|\ \mathrm{y})}_{p_d :=}} \prod_{i=1}^{j} \frac{1 + \prod_{l=1}^{k-1}(1 - 2p_{il})}{1 - \prod_{l=1}^{k-1}(1 - 2p_{il})} \\ &= \frac{1 - p_d}{p_d} \prod_{i=1}^{j} \frac{1 + \prod_{l=1}^{k-1}(1 - 2p_{il})}{1 - \prod_{l=1}^{k-1}(1 - 2p_{il})} \end{aligned}$$

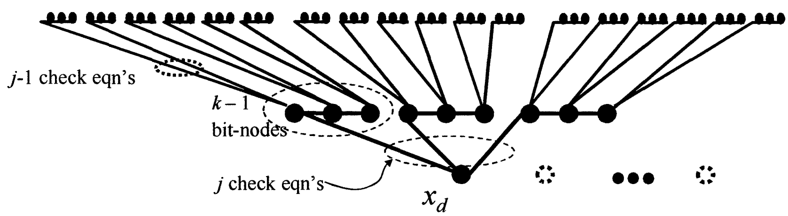
- Note $p_d$ and $p_{il}$, i=1, 2, …, j, l = 1, 2, …, $k$-1 are posterior probabilities of having digit "1" at the particular location given the complete output y

$$\begin{aligned} p_d &:= Pr(x_d = 1 \mid \mathrm{y}) \\ p_{il} &:= Pr(x^i_{d_l} = 1 \mid \mathrm{y}) \end{aligned}$$

Heung-No Lee 12

## Iterative Decoding

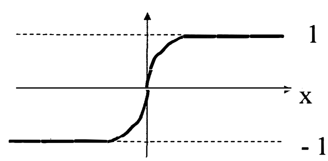
- We can consider the sub-code scenario for each of the $j*(k\text{-}1)$ first tier bit nodes
- Now, we have $(j - 1)$ check equations for each bit node (why?)
  - Each check equation checks $(k - 1)$ bit nodes




Heung-No Lee 13

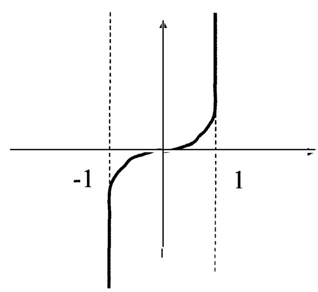
## tanh and tanh$^{-1}$ function

- Note $tanh(\frac{x}{2}) = \frac{e^x-1}{e^x+1}$




- Note

$$
\begin{aligned}
tanh^{-1}(x) &= \frac{1}{2}log\frac{1+x}{1-x} \\
&= sign(x)\frac{1}{2}log\frac{1+|x|}{1-|x|}
\end{aligned}
$$




Heung-No Lee 14

## Log Ratio Algorithm

- Take the log of the ratio of the posteriors

$$log\frac{Pr(x_d = 1|\mathrm{y}, S)}{Pr(x_d = 0|\mathrm{y}, S)} = log\frac{p_d}{1-p_d} + \sum_{i=1}^{j} log\frac{1-\prod_{l=1}^{k-1}(1-2p_{il})}{1+\prod_{l=1}^{k-1}(1-2p_{il})}$$

- Using $tanh(\frac{x}{2}) = \frac{e^x-1}{e^x+1}$, the summand of the second term is

$$log\frac{1-(-1)^{k-1}\prod_{l=1}^{k-1} tanh(\frac{LR(p_{il})}{2})}{1+(-1)^{k-1}\prod_{l=1}^{k-1} tanh(\frac{LR(p_{il})}{2})} = log\frac{1+(-1)^{k}\prod_{l=1}^{k-1} tanh(\frac{LR(p_{il})}{2})}{1-(-1)^{k}\prod_{l=1}^{k-1} tanh(\frac{LR(p_{il})}{2})}$$

- Making use of $tanh^{-1}(x) = \frac{1}{2}log\frac{1+x}{1-x}$ , it becomes

$$\sum_{i=1}^{j} 2\, tanh^{-1}((-1)^k \prod_{l=1}^{k-1} tanh(\frac{LR(p_{il})}{2}))$$

$$= \sum_{i=1}^{j} (-1)^k\, 2\, tanh^{-1}(\prod_{l=1}^{k-1} tanh(\frac{LR(p_{il})}{2}))$$

Making use of tanh$^{-1}$ being odd function

Heung-No Lee 15

## Product of Real Numbers

- $\prod_i \alpha_i = [\prod_i sign(\alpha_i)] \cdot exp(\sum_i log(|\alpha_i|))$

- $a\,b$ = sign($a$) sign(b) exp(log(|$a$|)) exp(log(|b|))

- $\prod_{l=1}^{k-1} tanh(\frac{\alpha(p_{il})}{2})$

$$= [\prod_{l=1}^{k-1} sign(LR(p_{il}))] \cdot exp(\sum_{l=1}^{k-1} log(tanh(\frac{|LR(p_{il})|}{2}))$$

Heung-No Lee 16

$$f(x) := -\log(\tanh(x/2)) = \log\frac{e^x+1}{e^x-1}$$

- Use the identity of product of real numbers to get rid of product

$$\sum_{i=1}^{j} (-1)^k\, 2\, tanh^{-1}\Big(\prod_{l=1}^{k-1} tanh\Big(\frac{LR(p_{il})}{2}\Big)\Big)$$

$$= \sum_{i=1}^{j} (-1)^k \Big[\prod_{l=1}^{k-1} sign(LR(p_{il}))\Big]\, 2\, tanh^{-1}\Big[exp\Big(\sum_{l=1}^{k-1} log\Big(tanh\Big(\frac{|LR(p_{il})|}{2}\Big)\Big)\Big)\Big]$$

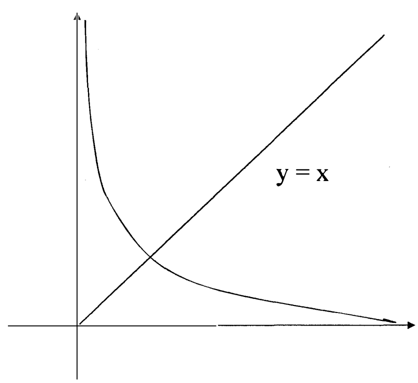
$$= \sum_{i=1}^{j} \underbrace{\Big[\prod_{l=1}^{k-1} sign(LR(p_{il}))\Big] \cdot f^{-1}\Big(\sum_{l=1}^{k-1} f(|LR(p_{il})|)\Big)}_{\text{Information generated by the i-th check node}}$$

Log ratio: info. from bit nodes

Heung-No Lee 17

$$f^{-1}(x) = f(x)$$

- Symmetric wrt y=x




Heung-No Lee 18

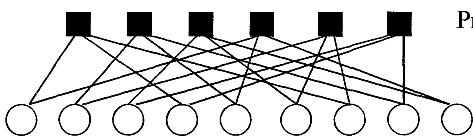
## Finally, the Log Ratio Algorithm

- Note the ratio here is Pr(x=1)/Pr(x=0), which is the inverse of the ratio used in Gallagar's thesis
- With the following definitions
  - $LR(p_d) := log\frac{p_d}{1-p_d} \quad LR(p_{il}) := log\frac{p_{il}}{1-p_{il}}$
  - $LR(p'_d) := log\frac{Pr(x_d=1|S,y)}{Pr(x_d=0|S,y)}$
- Theorem becomes

$$LR(p'_d) := LR(p_d)+\sum_{i=1}^{j}[\prod_{l=1}^{k-1} sign(LR(p_{il}))]f[\sum_{l=1}^{k-1} f(|LR(p_{il})|)]$$

Heung-No Lee 19

## Number of Edges in Bipartite Graph



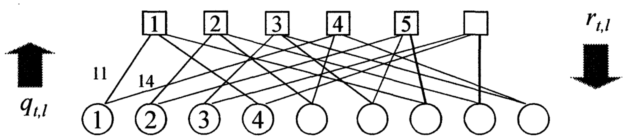
$Pr(S_i|x_d=1,y)$

$Pr(x_d=1|S,y)$

- There are *n* bit nodes and *L* check equations
  - Then there are $E=n*j = L*k$ edges
  - The total number of messages flowing from bit to check, and also from check to bit, is $E$

Heung-No Lee 20

## Let's Label the Edges




- Starting from the left of the graph
  - For the edges connecting bit-1 to check-1 and check-4, let's name them to be '11' and '14' respectively
  - For the edges connecting bit-2 to check-2 and check-5, let's name them to be '22' and '25' respectively
  - And so on
- Now, let's define $q_{t,l}$ the message from the bit-$t$ to check-$l$
- And, define $r_{t,l}$ the message from the check-$l$ to bit-$t$

Heung-No Lee 21

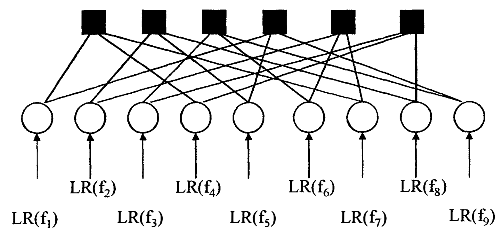
## Likelihoods as Input

- $y_t = (2x_t - 1) + n_t$, where $n_t$ is AWGN for $t$=1, 2, …, n where $n_t$ is $\mathcal{N}(0, N_0/(2E_s))$ with $E_s = E_b * R$
- Obtain the likelihood function $p(y_t|x_t)$
- The log likelihoods are used to start the iteration
- Let's denote the likelihood functions
  - $f_t(1) = p(y_t|x_t{=}1)$ and $f_t(0) = p(y_t|x_t{=}0)$
- The Log Ratio of Likelihood Probability is

$$LR(f_t) = \log(f_t(1)/f_t(0)) = (4E_s/N_0)y_t$$

Heung-No Lee 22

## In the beginning, we have $LLR(f_k)$




- The log likelihood ratios (or $\{f_t(1)\}$) are the input to the message passing decoder, to start the iterative decoding

Heung-No Lee 23

## Until the last iteration

- Do the second tier calculation
- Each bit node generates *j* bit-to-check messages
  - Each bit-to-check message is generated by checking *j*-1 check equations, excluding the check equation to which the message flows
- Each check node generates *k* check-to-bit messages
  - Each check-to-bit message is generated by utilizing *k*-1 posteriors, excluding the edge connecting to the bit node to which the message flows

Heung-No Lee 24

# Parity Check Matrix on Bipartite Graph
## ($n$=9, $j$=2, $k$=3) example

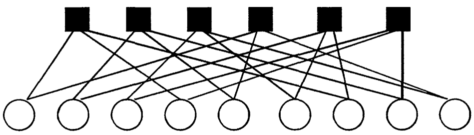
$$\begin{array}{c} \\ l \downarrow \end{array}\ \overset{t \longrightarrow}{\begin{bmatrix} 1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0 \\ 0&0&1&0&0&1&0&0&1 \\ 1&0&0&0&1&0&0&0&1 \\ 0&1&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&1&0 \end{bmatrix}} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_6 \end{bmatrix}$$

Q1(m,t): Row in a column

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| m=1 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| m=2 | 4 | 5 | 6 | 6 | 4 | 5 | 5 | 6 | 4 |

$j$

$L$ the number of checks



Q2($m$, $l$): Column in a row

| $l$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| m=1 | 1 | 2 | 3 | 1 | 2 | 3 |
| m=2 | 4 | 5 | 6 | 5 | 6 | 4 |
| m=3 | 7 | 8 | 9 | 9 | 7 | 8 |

$k$

Heung-No Lee 25

# Summary of Decoding (Log Ratio)

($n$, $j$, $k$) code with $f(x) := -\log(\tanh(x/2)) = \log[(e^x + 1)/(e^x - 1)]$

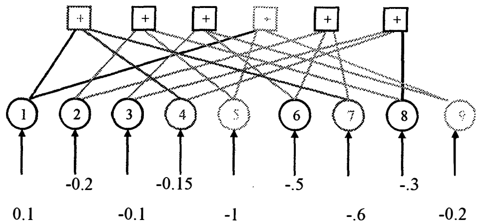
- Initialize:
  - $LR(f_t) = (4E_s/N_0)y_t$,
  - $LR(r_{tl}) = 0$, $t$=1, 2, …, $n$ and $l$=1,2,…, $k$
- Iteration:
  - Bit-to-Check messages: $LR(q_{t,Q1(m,t)})$, $t$=1,2,…, $n$; $m$=1,2, …, $j$

    $LR(q_{t,Q1(m,t)}) = LR(f_t) + \sum_{m' \neq m} LR(r_{t,Q1(m',t)})$
  - Check-to-Bit messages: $LR(r_{Q2(m,l),l})$, $l$=1,2,…,$L$; $m$=1,2,…, $k$

    $LR(r_{Q2(m,l),l}) = \prod_{m' \neq m} \mathrm{sgn}(LR(q_{Q2(m',l),l})) \times f[\sum_{m' \neq m} f(|LR(q_{Q2(m',l),l})|)]\ (-1)^k$
- Output:
  - $LR(p_t) = LR(f_t) + \sum_m LR(r_{t,Q1(m,t)})$
- Decision:
  - *if* $LR(p_t) > 0$ $x_t = 1$ ; *else* $x_t = 0$

Heung-No Lee 26

## Calculate Check-to-Bit messages




Do $l$ =1, $m$=1 ➔ Q2(1,1) = 1 = t

$r_{1,1}= (-1)^3 (-1)(-1) f[f(|-0.15|)+f(|-.6|)] = (-1) f(2.59 + 1.23) = -.44$

Do $l$ =1, $m$=2 ➔ Q2(2,1) = 4 = t

$r_{4,1}= (-1)^3 (+1)(-1) f[f(|0.1|)+f(|-.6|)] = f(3.0 + 1.23) = 0.029$

Do $l$ =1, $m$=3 ➔ Q2(3,1) = 7 = t

$r_{7,1}= (-1)^3 (+1)(-1) f[f(|0.1|)+f(|-.15|)] = f(3.0 + 2.59) = 0.0075$

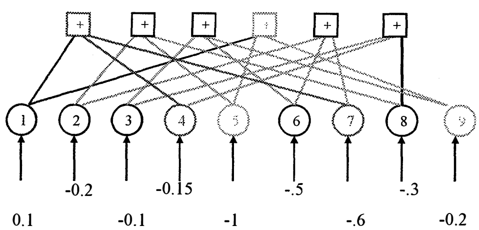
Do $l$ =4, $m$=1 ➔ Q2(1,4) = 1 = t

$r_{1,4}= (-1)^3 (-1)(-1) f[f(|-1|)+f(|-.2|)] = (-1) f(0.77 + 2.3) = -.093$

Heung-No Lee 27

## Calculate Bit to Check messages

$r_{1,1} = -.44$
$r_{1,4} = -.093$




Do $t$ =1, $m$=1 ➔ Q1(1,1) = 1 = $l$

$q_{1,1}= 0.1 + r_{1,4} = .007$ (m' = 2 → Q1(2,1) =4)

Do $t$ =1, $m$=2 ➔ Q1(2,1) = 4 = t

$q_{1,4}= 0.1 + r_{1,1} = -0.33$ (m' = 1 → Q1(1,1) =1)

Heung-No Lee 28

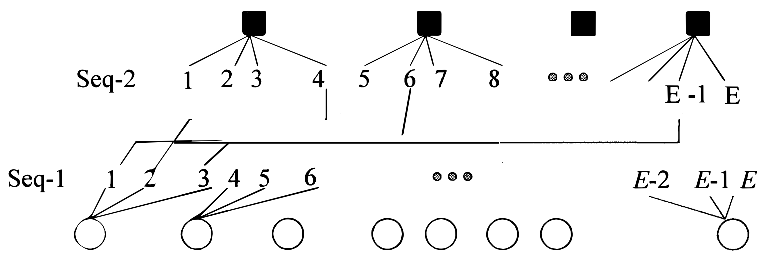
## Generation of **H** matrix

- There are many ways to generate the H matrix
- Any solution that satisfies the two constraints will do the job (Some bad choice will increase the code rate)

1. *j* number of ones in each column
2. *k* number of ones in each row

Heung-No Lee 29

## Using s-random interleaver to generate the parity matrix H: (n, j=3, k=4) example




- We know there are $E=n*j = L*k$ edges
  - Obtain the edge index sequence, say this is Seq-1.
  - Find a randomly interleaved (e.g. s-random interleaver: $s>k$) sequence $\pi$
  - Let Seq-2 = $\pi$(Seq-1) = (6 4 E-1 …)
  - Make connections according to Seq-1 and Seq-2

Heung-No Lee 30

## Generate **H** directly

- Randomly select the *j* positions of 1s in each column, while making sure that the number of ones in that particular row is not greater than *k*
- Equivalently, we can construct the Q1(m, t) matrix or Q2(m, *l*) matrix

Heung-No Lee 31

## Example with (n=9, j=2, k=3)

- Let's generate Q1(m,t) matrix
  - Rows={1,2,3,4,5,6}
- At t=1, randomly select two numbers from Rows
  - Suppose the two were 2, 5
- At t=2, select two numbers again from Rows
  - Suppose they were 2, 4
- At t=3, select two numbers again from Rows
  - Suppose it were 2, 3
  - Remove '2' from Rows
- At t=4, select two numbers from Rows-{2}
- And so on

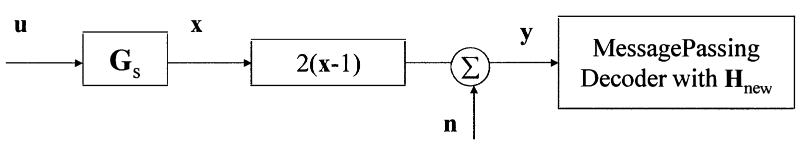
Heung-No Lee 32

## Gaussian Elimination on $\mathbf{H}$

- Once $\mathbf{H}$ is found, perform Gaussian Elimination on $\mathbf{H}$ and find a systematic $\mathbf{H}$, say $\mathbf{H}_s$
  - Keep track of the column exchanges made in GE
- Do the same set of column exchanges on $\mathbf{H}$ and obtain new $\mathbf{H}$ matrix $\mathbf{H}_{new}$
- Find $\mathbf{G}_s$ using the relationship $\mathbf{G}_s\mathbf{H}_s^T = \mathbf{0}$
- Use $\mathbf{G}_s$ in encoding the data
- Design the message passing decoder according to $\mathbf{H}_{new}$ (Why?)
  - $\mathbf{G}_s\mathbf{H}_{new}^T = \mathbf{0}$

Heung-No Lee 33

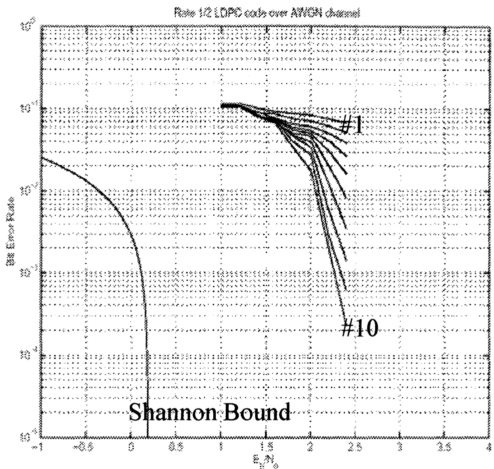
## Finally, we have




- Noise is AWGN with mean 0 and variance $N_0/(2E_s)$

Heung-No Lee 34

# Results of the rate ½ LDPC code, N=4096




Try to obtain BER=$10^{-5}$

Heung-No Lee 35

# HW #5

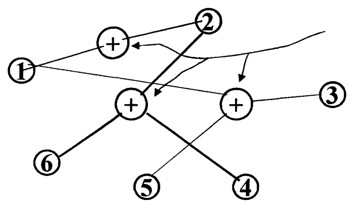
- The following homework questions are due by April 29.

Heung-No Lee 36

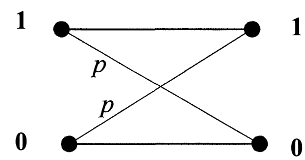
## Prob. 1: Design of Any Code/Performance Evaluation

6-tuple vectors $(x_1, x_2, x_3, x_4, x_5, x_6)$, where $x_i \in \{0, 1\}$

**Modulo-2 addition = 0 (coding constraint)**




**BSC**



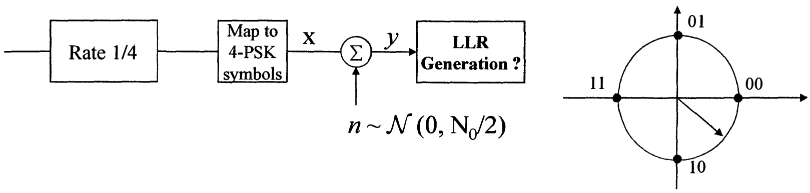

Heung-No Lee 37

## Prob. 1 cont'd

- From the linear code class, we can say that coding is nothing but choosing some words in a bigger dimensional space. Now consider a set of codewords which are related with each other as shown in the above graph.
- A) Find all valid codewords. List them and identify their Hamming weights (the number of ones in each codeword).
- B) How many bits can each codeword carry? And what is the coding rate of the code?
- C) Assume the use of the code over a discrete memory-less binary symmetric channel with the parameter $p$ $(0 < p < 1)$ as shown above:
  - What is the word error probability expression due to the minimum distance neighbors?
  - What is the probability of bit-error due to the minimum distance error event? Give a general expression as well as your answer at p=0.01.

Heung-No Lee 38

## Prob. 2: Log Likelihood Generation




- Derive the expression for the initial log-likelihood ratio as a function of $E_b/N_o$ in the following problem set up:
  - The channel noise is AWGN ~ $\mathcal{N}(0, N_0/2)$.
  - The modulation scheme is binary phase shift keying.
  - The channel code rate is 1/4.

- Now, make the following modification and do it again:
  - The modulation scheme is 4-Phase Shift Keying.
  - The bit-to-symbol mapping is shown in the constellation.
  - The channel noise is also two dimensional; each is AWGN ~ $\mathcal{N}(0, N_0/2)$.
  - Note that each channel symbol carries two coded bits. So, obtain LLR for each bit upon receiving $y$ (= $x + n$).

Heung-No Lee 39

## Prof. 3: Obtain Systematic Parity Check Matrix

- Perform the Gaussian elimination and bring the parity check matrix into the systematic form.

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Heung-No Lee 40

## Prob. 4: MATLAB programming of LDPC code

- Design an ($n$=12,$j$=3, $k$=6) Gallager code in MATLAB. The objective of this problem is to implement the LDPC decoder algorithm in MATLAB, and verify its successful operation.
  - (a) Obtain H.
  - (b) Program Gaussian elimination and obtain $G_{sys}$.
  - (c) Use $E_b/N_o$ = 10dB and obtain the receive vector **y**. From **y**, obtain the LLR vector LR(**f**).
  - (d) Do one iteration of Bit-to-Check and Check-to-Bit calculations. Print out your example, and verify your calculation results.
  - (e) Neatly write the procedure (a) to (e) carefully. The aim is to illustrate that you have verified a successful operation of your decoder.
    - State the method to obtain H
    - State how Gaussian elimination procedure were implemented
    - Explain the decoding procedure in a step-by-step manner
    - Do you think your decoder algorithm is programmed right and it works? Why?
  - (e) Now you are 100% sure from the procedure so far that your decoder is programmed right. Evaluate the performance of your code via Monte Carlo simulation over the AWGN channel. Do not simulate WERs < $10^{-3}$.

Heung-No Lee 41

## Prob. 4 (MATLAB) 2$^{nd}$ part

- (f) Obtain a tight upper (or lower) bound expression on the word error rate, which depends on the block length $n$, and use it to compare the simulation result. If necessary, use quantization on the received signal (and increase the level of quantization and see the effect).

- (g) Increase $n$ to 1024 and investigate its word error performance (word-error rate vs. $E_b/N_o$). Also, obtain the bit error rate and compare it with the analytical P(e) result for uncoded BPSK. This part is due by May 6$^{th}$.

Heung-No Lee 42

**Wireless Communications 2010**

**Prof. Heung-No Lee**

In this note, my aim is to illustrate the decoding procedure in a step-by-step manner. The example code used here is a Gallager (n=12, j=3, k=6) code.

We want to simulate at $E_b/N_o$ = 4dB = 2.5119. We let Es = 1.

The Rate of this code is 0.6667. It is not exactly equal to 1 – j/k = 1 – 3/6 = 1/2. It is a little bit higher than ½. Why?

The parity-check matrix of Gallager code always has j-1 dependent rows. Thus, the number of independent parity-check equations is always less by j-1. The actual rate is 1 – j/k + (j-1)/n. When n = 12, the rate R = 1/2 +1/6 = 0.6667.

Now, let's calculate $N_o$. $E_s$ = R*$E_b$. $E_b$ = $E_s$/R = 1/0.6667 = 1.4999. Thus, $N_o$ at $E_b/N_o$ = 4dB is equal to

$N_o$ = (1/2.5119)*(Es/R) = 0.5972.

txSymbols =

1 0 1 1 1 1 0 0 0 1 1 1

txSignal =

1 -1 1 1 1 1 -1 -1 -1 1 1 1

rxSignal = txSignal + noise

0.5873 -1.2776 0.9279 0.8469 1.5371 1.5159 -0.9929 -1.1936 -0.5111 0.5562 0.9401

-0.4926

LRft = (4*Es/No)*rxSignal

3.9339 -8.5578 6.2157 5.6731 10.2957 10.1539 -6.6506 -7.9953 -3.4236 3.7259 6.2974

-3.2999

LRb2c =

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3.9339 | -8.5578 | 6.2157 | 5.6731 | 10.2957 | 10.1539 | -6.6506 | -7.9953 | -3.4236 | 3.7259 | 6.2974 |
| 3.9339 | -8.5578 | 6.2157 | 5.6731 | 10.2957 | 10.1539 | -6.6506 | -7.9953 | -3.4236 | 3.7259 | 6.2974 |
| 3.9339 | -8.5578 | 6.2157 | 5.6731 | 10.2957 | 10.1539 | -6.6506 | -7.9953 | -3.4236 | 3.7259 | 6.2974 |

| |
|---|
| -3.2999 |
| -3.2999 |
| -3.2999 |

LRc2b =

| | | | | | |
|---|---|---|---|---|---|
| 3.2766 | 2.8952 | 2.8604 | -2.8826 | -3.6997 | 2.8924 |
| 2.7376 | -2.7889 | -2.7353 | 2.7399 | -3.1969 | 3.5722 |
| -3.2092 | -3.1618 | 3.1920 | 3.1690 | -3.2054 | 5.2025 |
| 2.8085 | -2.5310 | 2.5724 | 2.5291 | -3.0513 | 2.8862 |
| -2.6367 | -2.5905 | 2.5945 | 3.1576 | -2.6147 | 3.2640 |
| 3.5895 | -3.0584 | 3.0976 | 3.0552 | -3.0821 | 3.7684 |

LRpt =

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 13.6085 | -11.4096 | 8.9992 | 2.8200 | 7.4038 | 13.0028 | -9.4233 | 0.5080 | -7.0169 | 7.1836 | 3.3697 |

| |
|---|
| 8.7388 |

If we make decision right now, it should be

1 -1 1 1 1 1 -1 1 -1 1 1 1

Thus, the 8th bit 1 is an error, comparing the decided symbols with the txSignal.

txSignal =

1 -1 1 1 1 1 -1 -1 -1 1 1 1

Now, let's see what happens in the second iteration.

LRb2c =

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 10.3319 | -14.1473 | 6.1040 | 5.6088 | 4.5435 | 15.7381 | -6.5407 | -2.2318 | -3.3173 | 10.3805 | 0.4774 |
| 10.8001 | -8.8786 | 12.2084 | 0.2476 | 10.5657 | 10.4738 | -12.6152 | -2.6610 | -3.9657 | 4.2974 | 6.5751 |
| 10.0190 | -8.3512 | 5.9016 | 5.4566 | 9.9943 | 9.9477 | -6.3412 | -2.0864 | -10.1745 | 3.4152 | 5.9844 |

5.1666

3.5363

5.4748

LRc2b =

| | | | | | |
|---|---|---|---|---|---|
| 0.4299 | 0.4319 | 0.4394 | -0.4312 | -0.4634 | 2.9844 |
| -2.1489 | 2.1804 | 2.1489 | -4.6669 | 2.1491 | 2.1984 |
| 2.3005 | 2.3007 | -2.3005 | -3.4884 | 2.3143 | 2.6408 |
| 0.2318 | -0.2318 | 3.4192 | 0.2318 | -0.2408 | 0.2382 |
| 2.0340 | 2.0028 | -4.5044 | -2.0027 | 2.0210 | 2.0334 |
| 3.2795 | -3.2846 | 3.3536 | 3.2796 | -3.3262 | 5.3338 |

LRpt =

7.8751 -14.2232 12.3016 13.3067 15.0387 15.8141 -12.7084 -20.6551 -6.1305 11.4471 13.6171 3.5727

Now making decision with this sequence, we have

1 -1 1 1 1 1 -1 -1 -1 1 1 1

We note that this is equal to the txSignal. Thus, the error has been corrected.

The matrices for this illustration were

H =

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |

Hs =

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |

Gs =

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

>> Q1

Q1 =

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 4 | 4 | 3 | 4 | 3 | 4 | 3 | 3 | 4 | 4 | 3 | 3 |
| 6 | 6 | 6 | 5 | 5 | 6 | 6 | 5 | 5 | 6 | 5 | 5 |

>> Q2

Q2 =

| | | | | | |
|---|---|---|---|---|---|
| 1 | 3 | 5 | 7 | 9 | 11 |
| 2 | 4 | 6 | 8 | 10 | 12 |
| 3 | 5 | 7 | 8 | 11 | 12 |
| 1 | 2 | 4 | 6 | 9 | 10 |
| 4 | 5 | 8 | 9 | 11 | 12 |
| 1 | 2 | 3 | 6 | 7 | 10 |

I hope with this step-by-step illustration everyone can design their own LDPC encoding/decoding system.

End of illustration.

# Fading Channels/Multiple Antenna Diversity

Diversity Benefits

 1

## Agenda

- Fading Channels
- Chi-Square Random Variables
- Multipath channel
- Diversity system
  - Multiple receiver
  - Multiple transmitter

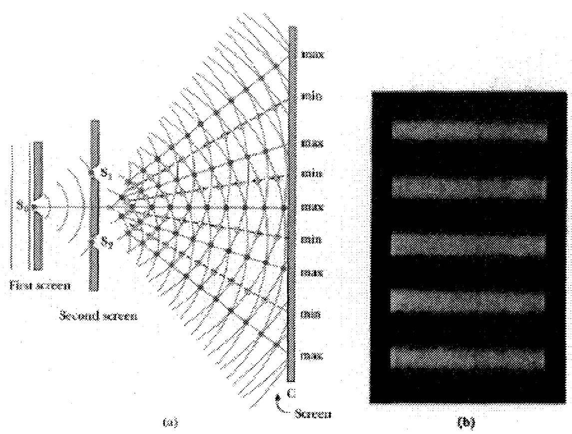
 2

## Fading Channels

- One of the characteristics of terrestrial wireless channels is the signal-strength fading
- The culprit of the fading is multi-path propagation of the radio-wave signals
- Thus, the fading phenomenon becomes significant in rich scattering environment where a large number of reflectors and scatters are around the receiver
- In this course, we model the channel as the tapped delay line with fading taps
- Good transceiver design provide a solution that is not limited by fading but rather takes advantage of it
- Recent results show that we can utilize the signal strength fading to boost up the channel capacity

 3

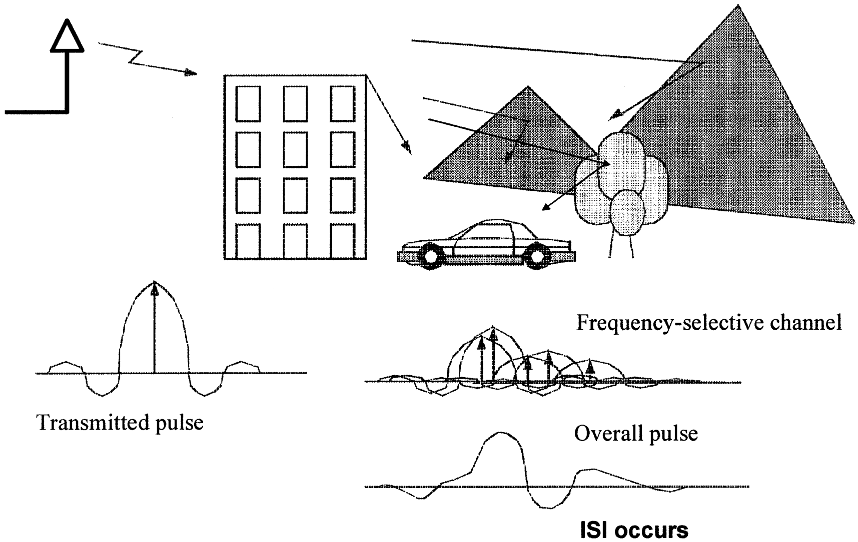
## Young's Double Slit Experiment




*A Google Picture by Simon Connell 2006-02-21*

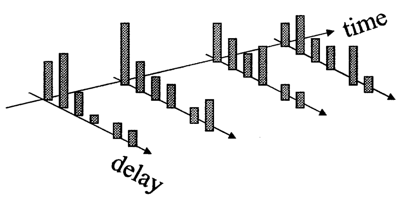
 University of Pittsburgh 4

## Multipath Fading ISI Channels




## Statistical Channel Characterization

- Channel response is parameterized with delay and time $c(\tau, t)$.




Variation of channel response over time

- We can use the two dimensional auto-correlation function $\phi_c(\tau_1, \tau_2; t_1, t_2) = E\{c^*(\tau_1, t_1)\, c(\tau_2, t_2)\}$.
  - Delay characterization: Multipath delay profile
  - Time variation: Doppler spread

©200x Heung-No Lee 6

## Wiener Kintchine Theorem

- How do you measure the frequency content or the degree of time-variation (frequency content) of random processes (signals)?
  - Autocorrelation function ⇔ Power Spectral Density (Fourier Transform pair, Wiener Kintchine Theorem)
  - Wide-sense stationary
- $R(\tau) = E\{G^*(t+\tau)\, G(t)\}$, $-\infty \leq \tau \leq \infty$
- $P(f) = \mathcal{F}\{R(\tau)\}$ due to WK theorem
- In many engineering problems, we also make the assumption that the process is ergodic and use the time-averaging operator to estimate $R(\tau)$ or $P(f)$
  - $R_e(k) = (1/K) \sum_{k=1}^{K} g^*(k+n)\, g(k)$ where $g(k)$ is a sample-path

 7

## Two Types of Frequencies in this note

- Frequency response of a channel at a fixed time instance
  - Fourier transform of the impulse response at the particular time
  - Measures channel's response(magnitude, phase) to a sinusoidal input as its frequency is varied
  - Channel's response to different tones.
- Frequency of channel variation
  - How quickly the channel changes?
  - Doppler frequency

 University of Pittsburgh 8

## Two Autocorrelation Functions of Importance

- Multipath Power Delay Profile
  - $\phi_c(\tau) = E\{c^*(\tau)\, c(\tau)\}$
- Time autocorrelation function
  - $\phi_c(\Delta t) = E\{c^*(\tau, t_1)\, c(\tau, t_1 + \Delta t)\}$
  - Doppler spectrum

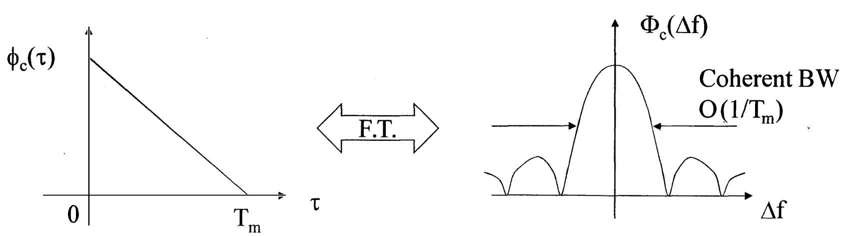


## Multipath Power Delay Profile

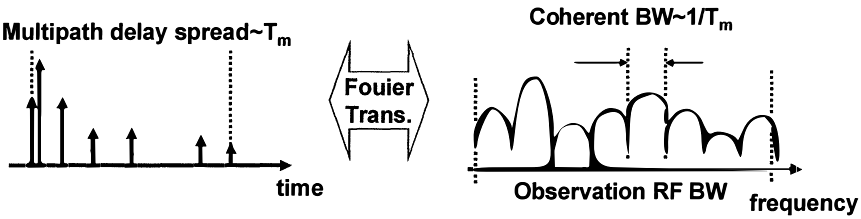
- One of them is the multipath power delay profile

$$\phi_c(\tau) = E\{c^*(\tau)\, c(\tau)\}$$

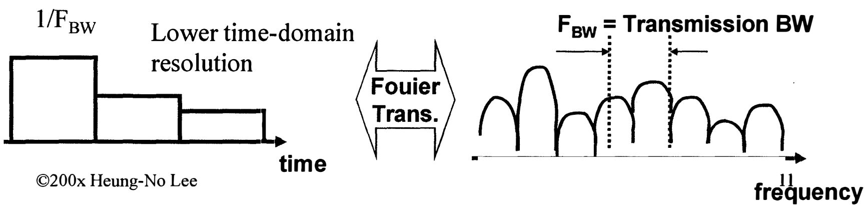
  - It's the average power of a multi-path component at delay $\tau$.
  - Multiplication in time domain = Autocorrelation in Frequency Domain



## Delay dispersive/Frequency-selective channel

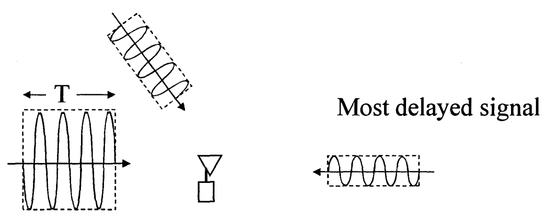



- $T_m$ ~ the time difference between the ***first*** and the ***last*** arriving pulse
  **Indoor ~ up to a couple of meters ---> 10s of nano seconds (10s of MHz)**
  **Outdoor ~ up to 1000s meters ---> 10s of micro seconds (10s of KHz)**




©200x Heung-No Lee 11

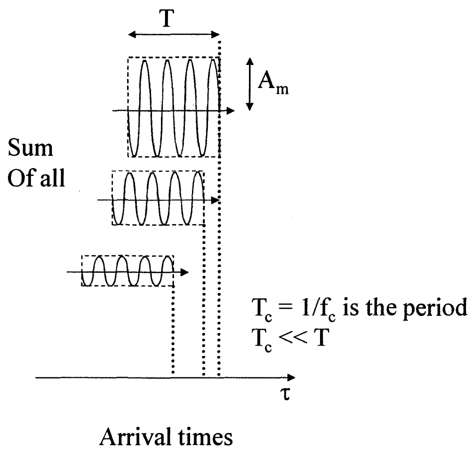
## What is happening at the receiver antenna?




- The phase and strength of the electro-magnetic excitation at the receiver's antenna is proportional to the summation of all the incidental waves arriving to the receiver antenna with different phases and magnitudes due to multipath propagation

©200x Heung-No Lee 12

## Delay Spread vs. Gaussian Fading Model




- When arrival-time differences are relatively small ($T_m <<$ T), the signal perceived at the receiver is the addition of all arriving sinusoids.
- $Y = \sum_m A_m \exp(j2\pi\, \theta_m)$
- Add constructively or destructively depending on the phase
- The delay $\tau_m$ is translated to the phase difference $\theta_m \in [0, 2\pi]$, uniform distributed.
- Invoking the Central Limit Theorem, Y is modeled as a complex-valued Gaussian.

 13

## **Flat-Fading**: When delay-spread is negligible ($T_m << T$)

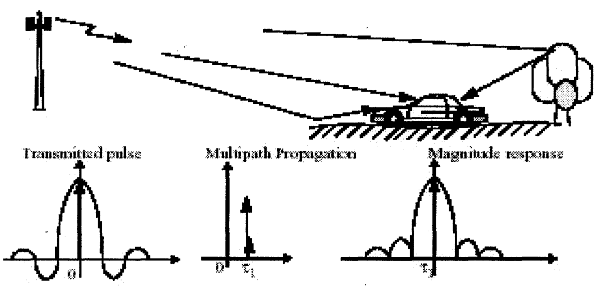
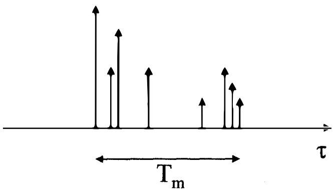



Figure 2-2 Outdoor wireless communication channels in cellular structure for the typical rural area.

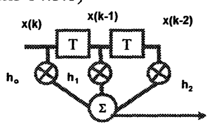
Only local scattering objects, then we have a single Rayleigh fading tap Such that $y_k = a_k x_k + n_k$, where $|a_k|$ is Rayleigh distr.

 14

## Frequency-Selective Fading:
When delay-spread is significant ($T_m \approx T$)




Tapped-delay-line channel model
(Proakis 14.5.1)




$$y_n = \Sigma_k\, x(k-1)\, h_k + n_k$$

- Shown left is a wideband channel response
- The pulse used to sound the channel is very wide so that it is *resolvable* at the receiver as shown
- If a pulse of shape g(t) is used then the received signal is convolution of g(t) with the wideband channel
- Ex) g(t) = Rect(t/T)

©200x Heung-No Lee 15

---

When delay is significant ($T_m = O(T)$)
We model the channel with tapped delay line filter

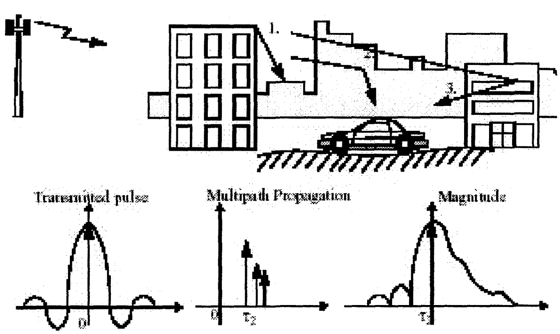



Figure 2-3 Outdoor wireless communication channels in cellular structure for the typical urban area.

E.M. Propagation Characteristics
-- Scattering
-- Refraction
-- Reflection

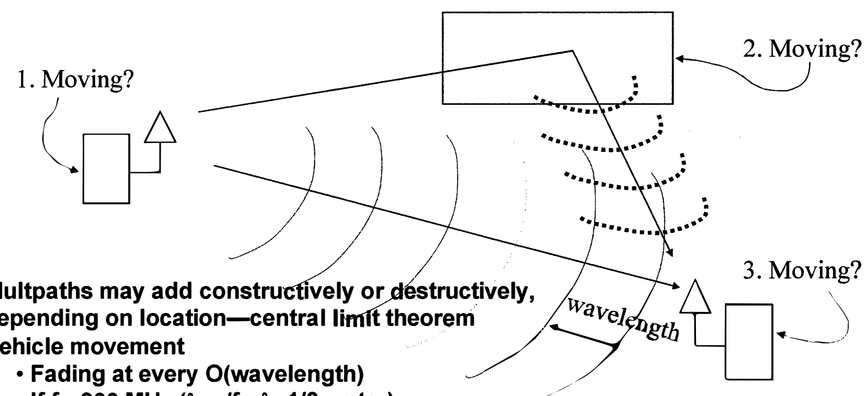
©200x Heung-No Lee 16

## Summary on Delay Spread

- If delay spread is small, $T_m << T$, then we say the channel is frequency non selective and use a single tap channel model.
  - This happens when the transmission bandwidth is smaller than the coherence BW of the channel.
- If delay spread is large, $T_m >= T$, we say the channel is frequency-selective and use multi-tap FIR channel model.
  - This happens when the transmission bandwidth is larger than the coherence BW of the channel.

©200x Heung-No Lee University of Pittsburgh 17

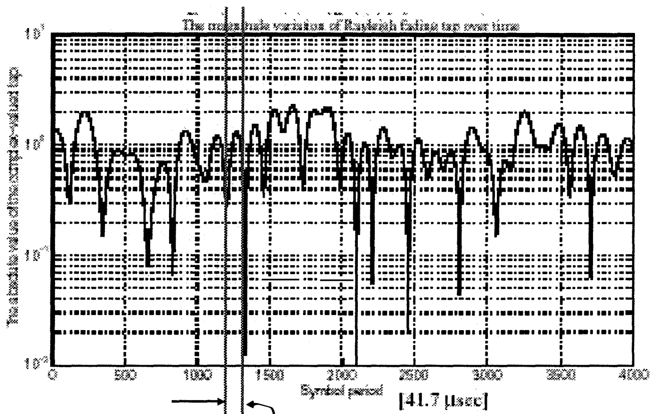
## Time-Varying Fading Channel due to Mobility




- **Multpaths may add constructively or destructively, depending on location—central limit theorem**
- **Vehicle movement**
  - **Fading at every O(wavelength)**
  - **If $f_c$=900 MHz ($\lambda=c/f_c$, $\lambda$=1/3 meter)**
- **Doppler spreading (measured as the max. Doppler frequency)**
  - **$f_{dm}=f_c(v/c)=v/\lambda$ = 33 [m/sec] / 1/3 [m] ~ 100 Hz**
    - **$v_{max}$= 120 km/hr = 33 m/sec**

©200x Heung-No Lee 18

## A sample path of a single fading tap over time, of a moving receiver




$\lambda/2$ — 0.165 m = 120 ticks * 33 [m/sec] * 41.7 [μ sec/tick]

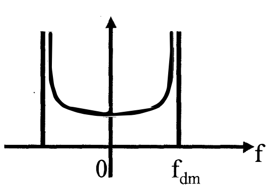
 

## Time Autocorrelation Function vs. Doppler Power Spectrum

- The time auto-correlation function measures the degrees of time-variation in the channel response

  $\phi_c(\Delta t) = E\{c^*(\tau, t_1)\, c(\tau, t_1 + \Delta t)\}$
- The F.T. of this function is called the Doppler power spectrum $S_C(f)$
- For the outdoor Doppler spectrum, we usually follow the Jakes' model

$$S_c(\lambda) = (1 - |f/f_{dm}|^2)^{-1/2}, \text{ if } |f| \le f_{dm}$$
$$= 0, \text{ o.w.}$$

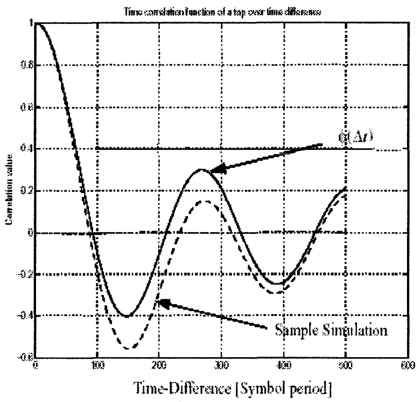



F.T.

$\phi_c(\Delta t) = J_0(2\pi\, f_{dm}\, \Delta t)$,
the zeroth order Bessel function of the first kind

## Time Auto-correlation Function (Outdoor channels)




The zero-th order Bessel function of the first kind

Obtainable from inverse F.T. of $S_C(f)$




## Doppler Power Spectrum

- The F.T. of this Time-Autocorrelation function is called the Doppler power spectrum $S_C(f)$
- It gives you the idea of how fast the channel is time-varying
- For a discrete-time example, we can estimate the $S_C(f)$
  - choose a time-varying tap and collect the tap values over a specific amount of time forming a sequence
  - Take ‘xcorr’ and take ‘fft’ of the results

## A fading tap is a band-limited random process

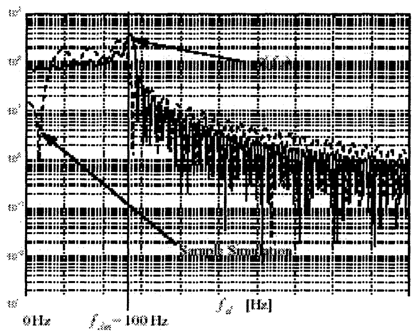


Figure 2-10 The autocorrelation function and the Doppler power spectrum of Rayleigh fading tap.

Determined by
The speed of the receiver
$f_{dm} = f_c v/\lambda$

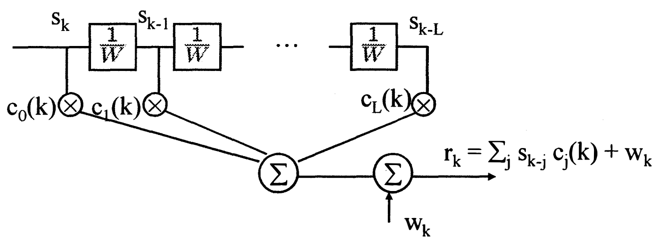
©200x Heung-No Lee 23

## Further Reading on Fading Channels

- Proakis Chapter 14, Digital Comm.
- Stein and Sklar, Comm Magazine, Sept. 97
- In general, you can use IEEExplore to find any IEEE papers
  - http://www.ieeexplore.ieee.org/
- MATLAB has a Rayleigh Fading Channel Simulator

©200x Heung-No Lee 24

## Time-Varing Frequency-Selective Fading Channel Model [The Bottom Line]




- $c(k) := (c_0(k)\ c_1(k) \ldots c_L(k))$ may represent the samples of the overall wireless channel's impulse response sounded by a pulse whose bandwidth is W Hz.
- Once we have a wideband pulse channel model, we can make narrow-band channel model very easily
  - Use the samples of $g(t) * c(t) * g(t)$, where $g(t)$ is the shaping filter
- Wide-sense stationary uncorrelated scattering wide-band channel model (WSSUS)
  - Each tap is a Rayleigh fading process; each fading process is uncorrelated with each other
  - Multi-path delay profile describes the powers of each tap, $[E\{|c_o|^2\}\ E\{|c_1|^2\} \ldots E\{|c_L|^2\}]$
  - How fast each tap is fading is described by the Doppler spectrum; the maximum Doppler rate

 25

## Proakis/Salehi Ch 2: Some Useful Probability Distributions

- Gaussian random variables: $X_i$, i.i.d. $\sim \mathcal{N}(m_x, \sigma^2)$
- Chi-square random variable: $Y := X_1^2 + X_2^2$
- $R = Y^{1/2}$
  - Rayleigh-distributed r.v. when $m_x = 0$
  - Rice-distributed r.v. when $m_x \neq 0$

 26

## Ricean Fading Simulation

- K ~ Ricean factor
- m = sqrt(K/(2*(K+1))) and s = sqrt(1/(2*(K+1)));
- Ricean simulation

  r = {s*randn(1,1)+m} +j*{s*randn(1,1) + m};
- R = abs(r) is in general a Ricean variable.
- We note $m^2 + s^2 = 1/2$ and $E(r^2) = 1$.

 27

## Random Variables (Processes) to Represent Fading Channel Statistics

- Zero-mean Complex-valued Gaussian process G(t)
- For fixed $t$, $G(t)$ is complex-valued Gaussian r.v.
  - The distribution of the power, $|G(t)|^2$, is Chi-square
  - The distribution of the mag., $|G(t)|$, is Rayleigh
- PDF of Chi-square r.v. with two deg. of freedom
  - $Y := G(t)^* G(t) = G_r^2 + G_i^2$ where $G_r$ and $G_i$ are independent Gaussians with zero mean and variance $\sigma^2$
  - $P(y) = (1/2\sigma^2) \exp(-y/(2\sigma^2))\, U(y)$
- PDF of Rayleigh r.v.
  - $R := Y^{1/2}$ with $P(r) = (r/\sigma^2) \exp(-r^2/(2\sigma^2))\, U(r)$
  - The magnitude of the signal

 28

# Chi-Square Distributed R.V.

- Let

  $Y := \sum_{j=1}^{n} X_j^2$

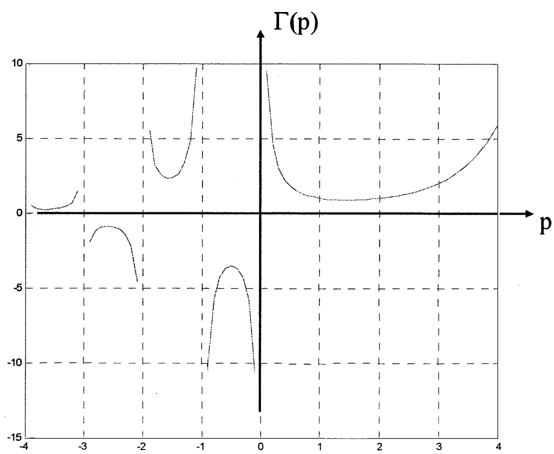
  where $\{X_j\}$ are iid Gaussian rvs with zero mean and variance $\sigma^2$
- It's pdf is

$$p_Y(y) = \frac{1}{\sigma^n 2^{n/2} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}} U(y) \qquad \text{--- (1)}$$

  where $\Gamma(p)$ is the gamma function
  - $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt$, $p > 0$
  - $\Gamma(p) = (p-1)!$, p an integer $> 0$
  - $\Gamma(1/2) = \pi^{1/2}$, $\Gamma(3/2) = 0.5\, \pi^{1/2}$

 29

# The Gamma Function




 30

## So much for the fading channel-model

- Let's move on to the communications theory again

©200x Heung-No Lee 31

## Simple Communications Model

- Consider the *flat-fading* channel case where the received signal is

  $y_k = r\, e^{j\theta}\, x_k + n_k$

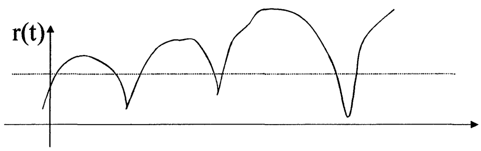
  where $r$ is Rayleigh distr., $\theta$ is uniform over $(0, 2\pi)$, and $n_k$ is complex-valued Gaussian with mean and variance $2(N_0/2)$
- Consider the coherent communications where $\theta$ can be estimated perfectly; then what matters is the magnitude-part only (multiply $y_k$ with $e^{-j\theta}$)
- Thus, it is equivalently

  $y_k = r\, x_k + w_k$, $w_k \sim$ real valued $\mathcal{N}(0, N_0/2)$ now (why?)

  which we already know it's P(e)

©200x Heung-No Lee 32

## The Adverse Effect of Fading




- Communication when the instantaneous channel gain is below a certain threshold → P(e) is large and the channel capacity is zero (useless)

©200x Heung-No Lee 33

## P(e) averaged over the single fading tap

- Let the instantaneous power be $a = r^2$ which we know that it's a Chi-square r.v. with two degrees of freedom.
- $P_b(a) = Q((2a/N_0)^{1/2})$, note that $2a/N_0$ is the instantaneous SNR
- $P(e) = \int \Pr\{\text{error} \mid a\}\, p(a)\, da = \int_0^\infty P_b(a)\, p(a)\, da$
  where $p(a)$ is the Chi-square pdf with $p(a) = e^{-a}\, U(a)$
- In fact, it's convenient to obtain the averaged P(e) in terms of the averaged SNR
  - Let $\gamma = 2a/N_0$;
  - Let the channel power $E(a) = 1$
    - then $E\{\gamma\} = 2/N_0$

©200x Heung-No Lee 34

## $P(e) \approx SNR^{-1}$

- Then, we can represent the P(e) as a function of SNR=E($\gamma$)
- First, $Q((2a/N_0)^{1/2}) = Q((a\ SNR)^{1/2}) < \exp(-a\ SNR/2)$
- Evaluate $P(e) = \int_0^\infty P_b(a)\ p(a)\ da$

## Using the result

- Using the Chernov bound over AWGN channel

$$\text{P(error)} \le Q\left(\sqrt{\frac{d_E^2}{2N_0}}\right) \le exp\left(-\frac{d_E^2}{4N_o}\right)$$

- It leads to

$$P_b(a) = Q\left(\sqrt{2a/N_0}\right) = Q\left(\sqrt{a\ SNR}\right) < exp(-\tfrac{a\ SNR}{2})$$

$$\begin{aligned}
P(e) &= \int_0^\infty P_b(a)\ p(a) da \\
&< \int_0^\infty exp(-\frac{a\ SNR}{2})\ exp(-a)\ da \\
&= \int_0^\infty exp(-\frac{a\ (2+SNR)}{2}) da \\
&= \frac{2}{2+SNR} \\
&\approx 2\ SNR^{-1}
\end{aligned}$$

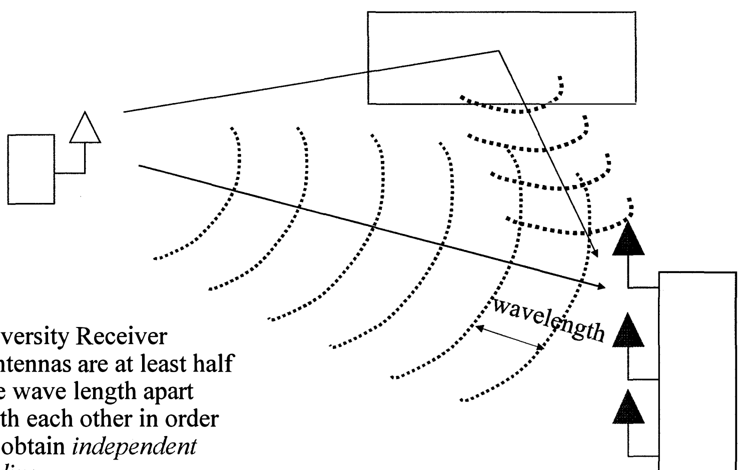
 

## Receiver Diversity

- Add more receive-antennas
  - to scoop up additional signal energy
    - This seems to be equivalent to increasing the diameter of your dish antenna in order to increase the SNR
  - And to achieve higher diversity order
    - The diversity benefit is obtained only when the receive antenna picks up the signal whose strength is independently faded

©200x Heung-No Lee 37

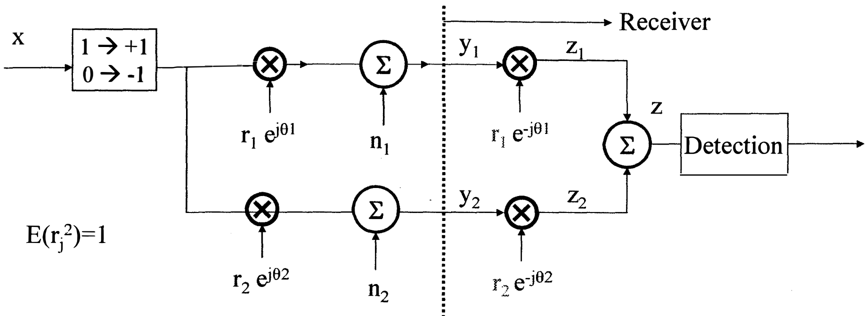
## Diversity Antenna Technique (Receiver Diversity)




- Diversity Receiver Antennas are at least half the wave length apart with each other in order to obtain *independent fading*

©200x Heung-No Lee 38

## Diversity Receiver Scenarios ($N_r = 2$ example)
## (So called Maximal Ratio Combiner)




- $y_d = r_d \exp(j\theta_d)\, x + n_d$, $d = 1, 2, \ldots, N_r$, where $N_r$ is the number of diversity antennas and $n_d$ is complex-valued Gaussian with zero mean, variance $2(N_0/2)$
- Coherent communications at each of the receivers

 39

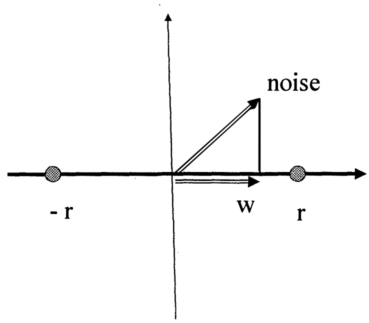
## Signal Space View

- Let $a = |r_1|^2 + |r_2|^2$
- Thus, the decision statistic is

  $$y = a\,x + w$$

  where $w = r_1 w_1 + r_2 w_2$ and $w_1$ and $w_2$ are i.i.d. real-valued Gaussian with variance $N_0/2$.
- Thus, $w$ is real valued Gaussian with variance $aN_0/2$.
- Thus, the instantaneous SNR $\gamma$ is
- $\gamma = a/N_0$
- SNR=E($\gamma$) = E($a/N_0$)

  $$= (E(r_1^2) + E(r_2^2))/N_0$$
  $$= 2/N_0$$




 40

## Two Orders of Diversity with Two Receive Antennas

- Now consider the P(e) over such setting

  $y = a\,x + w$, $w \sim$ real. $\mathcal{N}(0, aN_0/2)$
- It's

  $Q((2a/N_0)^{1/2}) = Q((a\ \mathrm{SNR})^{1/2}) < \exp(-\ a\ \mathrm{SNR}/2)$
- Thus, we again can evaluate the integral

  $P(e) = \int_0^\infty P_b(a)\ p(a)\ da$
  - $p(a)$ can be easily obtained from using the characteristic function (convolution of pdfs = multiplication in characteristic functions)
  - Using (1), n=4, $\sigma$=1/2, $P(a) = a\ e^{-a}\ U(a)$
- $P(e) \approx \mathrm{SNR}^{-2}$

 41

## Using the Chernov bound $P_b(a) \leq \exp(-\ a\ \mathrm{SNR}/2)$

- $P(e) = \int_0^\infty P_b(a)\ p(a)\ pa$

  $\leq \int \exp(-a\ \mathrm{SNR}/2)\ a\ \exp(-a)\ pa$

  $= \int x \exp(-\ x\ [2 + \mathrm{SNR}]/2\ )\ px$

  Let $c := (2+\mathrm{SNR})/2$

  $= \int x \exp(-\ cx)\ px$

  Let $y := cx$

  $= c^{-2} \int_0^\infty y \exp(-y)\ dy$

  $= c^{-2}$

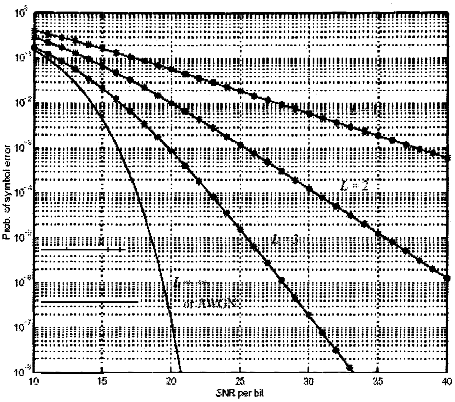
  $= 4/(2+\mathrm{SNR})^2$

  $\approx 4\ \mathrm{SNR}^{-2}$ High SNR

 42

# P(e) ≈ $\mathrm{SNR}^{-N_r}$ for $N_r$ receive antennas

- The bottom line is that P(e) is *power-law* decaying with order $N_r$, for $N_r$ receive antennas and 1 Tx
- As $N_r$ increases to $\infty$, P(e) converges to that of AWGN channels
- If you pick up signals coherently from many diversity antennas, you will get rid of the fading effect (of the signal strength of the decision statistic y at all)




Note here L = $N_r$

©200x Heung-No Lee 43

# Summary of MRC

- The averaged probability of making bit error in antipodal modulation is

$$P_2 \approx (4SNR_c)^{-N_r}\binom{2N_r-1}{N_r}$$

for L = 1, 2, 3, … where $SNR_c = E\{r_1^2\}/N_0$ is the per-channel SNR (or use the full expression in pg. 825 Proakis)

- Note that the above result applies whenever we have a decision variable in the form

  $y = a\,x + w$,

  where $a = r_1^2 + \ldots + r_L^2$ ($r_l$ is Rayleigh or $r_l^2$ is Chi-square with 2 degree of freedom) and $w \sim \mathcal{N}(0, aN_0/2)$

©200x Heung-No Lee 44

## Transmit Diversity

- Use multiple transmit antennas in order to obtain the diversity, why?
- Alamouti code
  - Only length-two code
  - Some extensions of this code, but only up to a certain length (say length 8)
- Delay diversity code
  - Use the same symbol across all the transmit antennas, but delayed by one symbol epoch at each antenna

 45

## Matched Filter Bounds On Multiple Transmit Antennas

- Now consider at least two signaling epochs, say $k = 0$ and $k = 1$ ($N_t$ signaling epochs for $N_t$ transmit signals)
- $y_0 = h_1 x^1 + h_2 x^2 + n_0$
- $y_1 = h_1 x^2 + h_2 x^1 + n_1$
- Thus, we may write $\mathbf{y} = H\mathbf{x} + \mathbf{n}$

$$\begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ h_2 & h_1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} n_0 \\ n_1 \end{pmatrix} \qquad \text{-----} (1)$$

- $z = H^* y = H^* H x + H^* n$

 46

## Matched Filter Bounds On Multiple Transmit Antennas

- $z = (z_1\ z_2)^t = H^*H\,x + H^*\,n$
- $\chi^2 = H^*H$
  - Diagonal terms of $\chi^2 = |h_1|^2 + |h_2|^2$
  - Off diagonal terms of $\chi^2 = h^*_1 h_2 + h^*_2 h_1 = 2\,\mathrm{Re}\{h^*_1\, h_2\}$
- $q = H^*n$, is a complex-Gaussian vector with zero mean vector and correlation matrix $E\{qq^*\} = N_0\,\chi^2$
- Ignoring the cross-correlation, we could have the optimal decision variable (the matched filter bound)

  $$z_k = a\,x_k + w_k$$

  where $w_k \sim \mathcal{N}(0, aN_0/2)$ and $a$ is $\chi^2_{2N_t}$
- Thus, the optimum system will achieve $N_t$ order of diversity (If there is one)

©200x Heung-No Lee 47

## Optimal Transceivers That Achieve the MFBs Very Closely

- A simple delay-diversity scheme: with a simple transmission scheme that avoids repeating the same channel symbol in other antennas at the same epoch (send repetitive symbols at other antennas in different time-epoch)
  - Maximum Likelihood Sequence Detection
    - The Viterbi Algorithm
  - Maximum A Posteriori Algorithm
    - The BCJR Algorithm
- With special schemes:
  - Alamouti Scheme (Alamouti, S.M. JSAC 98')---A simple transmit diversity technique for wireless communications
  - Space-Time Trellis Code or Space-Time Block Codes (Tarokh et al, ...)

©200x Heung-No Lee 48

## Alamouti Scheme

- Use a special repetition scheme
- $y_0 = h_1 x^1 + h_2 x^2 + n_0$ ------ (2)
- $y_1 = h_1 x^{2*} + h_2 (-x^{1*}) + n_1$
- Obtain $z_1 = h_1^* y_0 - h_2 y^*_1$
  $= |h_1|^2 x^1 + h_1^* h_2 x^2 - h_2 h_1^* x^2 + |h_2|^2 x^1 + h_1^* n_0 - h_2 n_1^*$
  $= (|h_1|^2 + |h_2|^2) x^1 + h_1^* n_0 - h_2 n_1^*$
- $z_2 = h^*_2 y_0 + h_1 y^*_1$
  $= h^*_2 h_1 x^1 + |h_2|^2 x^2 + |h_1|^2 x^2 - h_1 h_2^* x^1 + h_2^* n_0 + h_1 n_1^*$
  $= (|h_2|^2 + |h_1|^2) x^2 + h_2^* n_0 + h_1 n_1^*$
- Thus, the scheme perfectly achieves the MFB performance
- This scheme, however, exists only up to $N_t = 4$ systems (Refer Tarokh on Space-Time Block Codes)

 49

## Delay Diversity Coding

- Consider $N_t = 2$, $N_r = 1$ and $L=1$ case
- $y_k = h_1 x^1_k + h_2 x^2_k + n_k$
  where $n_k$ is a complex-Gaussian with zero mean and variance $N_0$
- Consider the case where $x^2_k = x^1_{k-1}$ (Delay diversity):
  $y_k = h_1 x^1_k + h_2 x^1_{k-1} + n_k$
- This looks like an ISI channel
  - Use the maximum likelihood sequence detection method
  - Obtain the full diversity with the VA

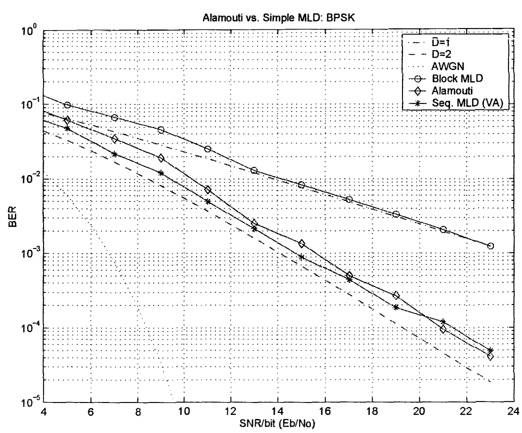
 50

## Simple Maximum Likelihood Detection

- Use Eq. (1) and use MLD
- For an example of binary modulation
- $x_i \in \{(-1,-1), (-1, 1), (1, -1), (1, 1)\}$ for i=0, 1, 2 and 3
- $\mathbf{y} = \mathbf{H}\,\mathbf{x} + \mathbf{n}$ where the CSC Gaussian channel matrix $\mathbf{H}=(h_1\ h_2\ \ldots\ h_{N_t})$ and $\mathbf{n}$ is a multivariate CSC Gaussian $[N_r \times 1]$ with zero mean and variance $R_n = N_0\, I_{N_r \times N_r}$
- The MLD then is to find the best input $\mathbf{x}_i$ which maximizes $P(\mathbf{y} \mid \mathbf{x}_i) \sim \min (\| \mathbf{y} - \mathbf{x}_i \|)$

©200x Heung-No Lee 51

## Delay-Diversity with VA detection




- Seq. MLD (VA) is the delay-diversity case

©200x Heung-No Lee 52

## Diversity Analysis

- Read the following paper
  - **Space-time codes for high data rate wireless communication: performance criterion and code construction**
    Tarokh, V.; Seshadri, N.; Calderbank, A.R.; March, 1998.
  - Read the introduction.
  - Pay attention to the derivation of diversity benefit
  - Try to understand equations (1) thru (20).

 53

## Summary of Transmit Diversity

- With more transmit antennas, the diversity order can be improved.
  - Order of diversity = $N_t$
  - $P(e) \sim SNR^{-N_t}$

 54

# Chapter

# Chi-Square Random Variables and Diversity Benefits using Multiple Antennas

This chapter provides some useful results on Chi-square random variables and their relationship to the probability of making error averaged over the fading ensemble.

Recall that the error probability can be expressed as the Gaussian $Q$-function. For example, the bit error probability of binary antipodal modulation is given by $Q\left(\sqrt{\frac{2E_s}{N_o}}\right)$. Letting $SNR \triangleq \frac{E_s}{N_o}$, $P_2(SNR) = Q\left(\sqrt{2SNR}\right)$. Note this is a little bit different definition than what I have done in my ppt chart note.

Now, suppose the communication takes place over a fading channel. The gain of the channel is a random variable. Let's denote y the channel gain and its distribution $p_Y(y)$. Suppose for now the mean of distribution is 1. We are interested in calculating the probability of error averaged over the fading channel. This problem becomes the evaluation of the Gaussian $Q$-function over the distribution of Chi-square random variables. The average probability of error at a certain average SNR is then

$$P_2(SNR) = \int_0^\infty Q(\sqrt{2 \cdot SNR \cdot y})p_Y(y)dy, \tag{1.1}$$

where $p_Y(y)$ denotes the pdf of Chi-square random variable of a certain degree.

For this purpose, the Gaussian $Q$-Function can be simplified as an exponential function of the SNR using the Chernoff bound, i.e.

$$Q(\sqrt{2 \cdot SNR \cdot y}) \le \exp(-SNR \cdot y). \tag{1.2}$$

Or, the Craig-Identity can be used

$$Q(\sqrt{2SNR \cdot y}) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(\frac{-SNR \cdot y}{\sin^2\theta}\right)d\theta. \tag{1.3}$$

Note that in both expressions we have the exponential function of $y$. We take advantage of this fact in evaluating the integral. Namely, the expression in (1.1) can turn in to the form of the moment generating function of $Y$ using (1.2) or (1.3) into (1.1).

The moment generating function is defined as $M(t) := E\left[e^{tY}\right] = \int_0^\infty e^{ty}p_Y(y)dy$.

For example, substituting the Craig Identity of $Q(x)$ in (1.3) into (1.1), one have

$$\begin{aligned} P_2(SNR) &= \frac{1}{\pi}\int_0^{\pi/2}\int_0^\infty \exp\left(\frac{-SNR \cdot y}{\sin^2\theta}\right)p_Y(y)dyd\theta \\ &= \frac{1}{\pi}\int_0^{\pi/2} M\left(t = \frac{-SNR}{\sin^2\theta}\right)d\theta \end{aligned}, \tag{1.4}$$

where $M(t)$ is the moment generating function of pdf $p_Y(y)$. Even though it is an exact result, it should be noted that using CI involves an integration over the variable theta. Using the Chernoff bound (1.2), however, one can obtain the result in a compact form, i.e.,

$$\begin{aligned} P_2(SNR) &\le \int_0^\infty \exp\left(-SNR \cdot y\right) p_Y(y) dy d\theta \\ &= M\left(t = -SNR\right). \end{aligned} \tag{1.5}$$

Therefore, the error probability averaged over the fading ensemble can be obtained using the moment generating function.

In fact, the fading distribution takes the form of Chi-square distribution. Depending upon the degrees of the Chi-square random variable, varying results can be obtained.

**The Chi-square distribution with *n* degrees of freedom**: Let's denote $Y$ as the sum of squares of independent Gaussian random variables $\sim \mathcal{N}(m_i, s^2)$, i.e.

$$Y = \sum_{i=1}^{n} X_i^2 . \tag{1.6}$$

Note that we will be dealing with the case where $n$ is even. (Why?)
The moment generation function of this random variable is

$$M(t) = \frac{1}{(1-t2s^2)^{n/2}} \cancel{\exp} \left[ \frac{t\sum_{i=1}^{n} m_i^2}{1-t2s^2} \right], \tag{1.7}$$

and the pdf is

$$p_Y(y) = \frac{1}{2s^2} \left( \frac{y}{\sum_{i=1}^{n} m_i^2} \right)^{(n-2)/4} \exp\left( -\frac{\sum_{i=1}^{n} m_i^2 + y}{2s^2} \right) I_{\frac{n}{2}-1} \left( \sqrt{y} \frac{\left(\sum_{i=1}^{n} m_i^2\right)^{1/2}}{s^2} \right), \tag{1.8}$$

where $I_k(x)$ is the $k$-th order modified Bessel function of the first kind which can be evaluated in the following infinite series

$$I_k(x) = \sum_{i=0}^{\infty} \frac{(x/2)^{k+2i}}{i!\,\Gamma(k+i+1)} \tag{1.9}$$

where $\Gamma(p)$ is the Gamma function, i.e.,

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \qquad \text{for } p > 0 \tag{1.10}$$

$$\Gamma(p) = (p-1)! \qquad \text{for positive integer p,} \tag{1.11}$$

and few evaluations are $\Gamma(1/2)=\pi$ 1and $\Gamma(3/2)=\frac{1}{2}\sqrt{\pi}$. The alternate expression for $I_0(x)$ is useful, i.e., $I_0(x)=\frac{1}{\pi}\int_0^{\pi} e^{\pm x\cos\phi}d\phi$. Thus, we know $I_0(x=0)=1$.

**Application of Chi-square distribution to obtain the probability of error averaged over fading channel**

**The single tap fading case:**
Consider a Ricean random variable $R$ defined as $R=|\alpha|$ where

$$\alpha := (sX_1+m)+j(sX_2+m) \tag{1.12}$$

and $s$ and $m$ can be chosen as

$$s=\sqrt{\frac{1}{2(K+1)}}, \tag{1.13}$$

$$m=\sqrt{\frac{K}{2(K+1)}}, \tag{1.14}$$

for a *Ricean Factor*, $0\le K\le\infty$. $X_1$ and $X_2$ are mutually independent normal ~ $\mathcal{N}(0, 1)$. We define the Ricean factor with respect to $s$ and $m$ as

$$K=\frac{m^2}{s^2}. \tag{1.15}$$

Note that the Ricean factor $K$ implies the power ratio between the direct and the diffuse components. If $K=0$ (i.e. $m=0$), then we note that there is no direct component; then, $R$ is Rayleigh distributed. If $s=0$ or $K=\infty$; then, $R=1$ for sure; then the channel is AWGN channel (after phase equalization is done).

In addition, we note that $Y:=R^2=|\alpha|^2$ is in general a non-central Chi-square random variable with two degrees of freedom with $E(Y)=2(s^2+m^2)=1$.

The moment generating function $M(t):=E\left[e^{tY}\right]=\int_0^{\infty} e^{ty}p_Y(y)dy$ of the random variable $Y$ is given by

$$\begin{aligned} M(t) &= \left(\frac{1}{1-t2s^2}\right)\exp\left(\frac{t2m^2}{1-t2s^2}\right) \\ &= \left(\frac{1}{1-t\left(\frac{1}{(K+1)}\right)}\right)\exp\left(\frac{t\left(\frac{K}{(K+1)}\right)}{1-t\left(\frac{1}{(K+1)}\right)}\right). \end{aligned} \tag{1.16}$$

The pdf of $Y$ is obtained by letting $n = 2$, $E(Y) = 2(s^2 + m^2) = 1$ and $K = \frac{m^2}{s^2}$ into (1.8),

$$\begin{aligned} p_Y(y) &= \frac{1}{2s^2} \exp\left(-K - \frac{y}{2s^2}\right) I_0\left(\sqrt{y}\frac{\left(2m^2\right)^{1/2}}{s^2}\right) \\ &= \frac{1}{2s^2} \exp\left(-K - \frac{y}{2s^2}\right) I_0\left(\sqrt{y}\frac{\left(2Ks^2\right)^{1/2}}{s^2}\right) \\ &= \frac{1}{2s^2} \exp\left(-K - \frac{y}{2s^2}\right) I_0\left(\frac{1}{s}\sqrt{2Ky}\right) \end{aligned} \quad . \tag{1.17}$$

Alternatively, we sometimes represent the Ricean variable by fixing $s^2 = 1/2$, and thus $K = 2\,m^2$. We note that unlike previous case, $E(Y) = 2(s^2 + m^2) = 1$, in this case the expected value of the power variable $Y$ increases with the Ricean factor $K$ such that

$$E(Y) = 2s^2 + 2m^2 = 1 + K\ . \tag{1.18}$$

In this case, the moment generating function can be written as

$$M(t) = \left(\frac{1}{1-t}\right) \exp\left(\frac{tK}{1-t}\right). \tag{1.19}$$

And the pdf of Y is

$$\begin{aligned} p_Y(y) &= \exp\left(-\frac{2m^2 + y}{1}\right) I_0\left(\sqrt{y}\frac{\left(2m^2\right)^{1/2}}{1/2}\right) \\ &= \exp\left(-(K + y)\right) I_0\left(2\sqrt{yK}\right) \end{aligned} \tag{1.20}$$

We may use any of the two representation systems as long as we are aware that the average channel power is different, 1 vs. (1+$K$).

Note that the pdf can also be obtained by taking the inverse Fourier transform of the characteristic function (obtained by letting $t = j\omega$), which is given by

$$P_Y(y) = \frac{1}{2s^2}\left(\frac{y}{2m^2}\right)^{1/2} e^{-\frac{(2m^2+y)}{2s^2}} I_0\left(\sqrt{2y}\frac{m}{s^2}\right), \tag{1.21}$$

where $I_0(x)$ is the zeroth-order modified Bessel function of the first kind.

We are now ready to evaluate the probability of error averaged over the fading ensemble for the single tap case. First, using the Chernoff bound (1.2) we have

$$\begin{aligned} P_2(SNR) &= \int_0^{\infty} Q(\sqrt{2 \cdot SNR \cdot y}) p_Y(y) dy \\ &\leq \int_0^{\infty} \exp(-SNR \cdot y) p_Y(y) dy \\ &= M(t = -SNR). \end{aligned} \tag{1.22}$$

Using the un-normalized representation given in (1.19), the answer is

$$\begin{aligned} P_2(SNR) \leq M(t = -SNR) &= \left(\frac{1}{1+SNR}\right) \exp\left(\frac{-SNR \cdot K}{1+SNR}\right) \\ &= \left(\frac{1}{1+SNR}\right) \exp\left(\frac{-SNR \cdot K}{1+SNR}\right). \end{aligned} \tag{1.23}$$

Setting $K = 0$ (Rayleigh case), it becomes

$$P_2(SNR) \leq \left(\frac{1}{1+SNR}\right), \tag{1.24}$$

where $SNR := \frac{E_s}{N_o}$ as we have defined in the beginning of this note.

Note that this result is consistent with the one we obtained in class using the direct integration method.

## Two or more fading tap cases:

We now want to consider the two tap fading channel case in which the signal is coherently received. For example, take look at the chart shown below:

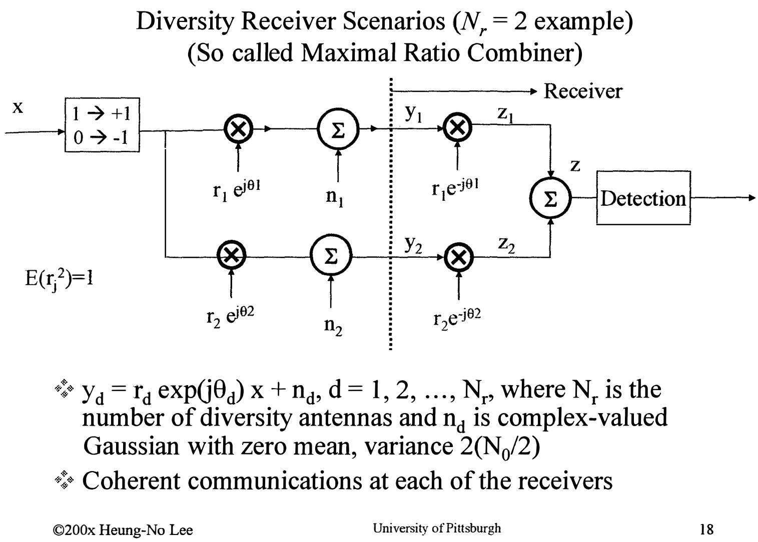



Figure 1: Maximal Ratio Diversity Reception of signal over fading channel

As depicted in the chart in Figure 1, let's assume we have a bank of receiver antennas available and the received signal $y_d$ in the $d$-th antenna, i.e.,

$$y_d = r_d e^{j\theta_d} x + n_d, \text{ for } d=1, 2, \ldots, N_r. \tag{1.25}$$

We define $\alpha_d := r_d e^{j\theta_d}$. There are total $N_r$ receive antennas. Consider coherent detection of signals such that the receiver has perfect estimations of the channel's fading coefficients. Then, the optimal maximal ratio combining is to multiply each received signal with the conjugate of the complex valued channel fading coefficient. Denoting $z_d := \alpha_d^* y_d$, we have

$$z_d = |r_d|^2 x + r_d e^{j\theta_d} n_d. \tag{1.26}$$

We note that the sum of all $z_d$, $d = 1, 2, \ldots, N_r$, is a constructive addition for the perspective of the signal component but not for the noise component. Now, by defining $z := \sum_{d=1}^{N_r} z_d$, it is written as

$$\begin{aligned} z &:= \sum_{d=1}^{N_r} z_d = \underbrace{\sum_{d=1}^{N_r} |r_d|^2}_{=:a} x + \sum_{d=1}^{N_r} r_d e^{j\theta_d} n_d \\ &= a\,x + \sum_{d=1}^{N_r} r_d e^{j\theta_d} n_d \end{aligned} \tag{1.27}$$

We note that the multiplicative factor $a$ is the sum of all channel gain squares, i.e.,

$$a := \sum_{d=1}^{N_r} |r_d|^2 . \tag{1.28}$$

Thus, $a$ is Chi-square random variable with $2N_r$ degrees of freedom.

Now, making use of our assumption that the signal $x$ is real valued binary antipodal $\{+\sqrt{E_s}, -\sqrt{E_s}\}$, we don't need to worry about the noise along the imaginary axis. Then, the equivalent real-valued channel model for (1.27) becomes

$$z = ax + w \tag{1.29}$$

where $w$ is real-valued Gaussian $\sim \mathcal{N}(0, aN_0/2)$. Thus, the probability of making bit error is simply $Q\left(\frac{2\sqrt{aE_s}}{\sqrt{2N_0}}\right) = Q\left(\sqrt{\frac{2aE_s}{N_0}}\right)$.

Let's denote $\gamma := \frac{aE_s}{N_o}$ to be the instantaneous SNR. Then, the average $SNR_t$ is defined as

$$SNR_t := E(\gamma) = E(a)\frac{E_s}{N_0}. \tag{1.30}$$

Using the normalized version, it is

$$SNR_t := E(\gamma) = E(a)\frac{E_s}{N_0} = N_r \frac{E_s}{N_0}. \tag{1.31}$$

The other definition of SNR we could use is the average SNR per receive channel, i.e.,

$$SNR := SNR_t / N_r = \frac{E_s}{N_0}. \tag{1.32}$$

With the average SNR per channel, the binary error probability is

$$Q\left(\sqrt{\frac{2aE_s}{N_0}}\right) = Q\left(\sqrt{2aSNR}\right) \le \exp\left(-a \cdot SNR\right). \tag{1.33}$$

Making use of the un-normalized approach given in (1.7) (i.e. $s^2 = 1/2$ and $2m^2 = K$), the moment generation function of Chi-square random variables with 2 $N_r$ degrees of freedom is

$$
\begin{aligned}
M(t) &= \frac{1}{(1-t)^{N_r}}\exp\left[\frac{t\sum_{i=1}^{2N_r} m_i^2}{1-t}\right] \\
&= \frac{1}{(1-t)^{N_r}}\exp\left[\frac{t2N_r m^2}{1-t}\right] \\
&= \frac{1}{(1-t)^{N_r}}\exp\left[\frac{tN_r K}{1-t}\right].
\end{aligned}
\tag{1.34}
$$

This result is good for showing how performance can be varied as $K$ is reduced down to 0 (Rayliegh fading). That is, the probability of making bit errors averaged over the Rayleigh fading channel can be obtained by setting $K = 0$ of the following expression:

$$
\begin{aligned}
P_2(SNR) &= \int_0^\infty Q(\sqrt{2\cdot SNR\cdot y})p_Y(y)dy \\
&\le \int_0^\infty \exp(-SNR\cdot y)p_Y(y)dy \\
&= M(t=-SNR) \\
&= \frac{1}{(1-t)^{N_r}}\exp\left[\frac{t\,N_r K}{1-t}\right]\Bigg|_{t=-SNR} \\
&= \frac{1}{(1+SNR)^{N_r}}\exp\left[\frac{-(SNR)N_r K}{1+SNR}\right].
\end{aligned}
\tag{1.35}
$$

On the other hand, using the normalized approach ($s = \sqrt{\frac{1}{2(K+1)}}$ and $m = \sqrt{\frac{K}{2(K+1)}}$), the moment generation function is

$$
\begin{aligned}
M(t) &= \frac{1}{(1-t2s^2)^{n/2}}\exp\left[\frac{t\sum_{i=1}^{n} m_i^2}{1-t2s^2}\right] \\
&= \frac{1}{(1-t(\frac{1}{K+1}))^{n/2}}\exp\left[\frac{t\,N_r(\frac{K}{K+1})}{1-t(\frac{1}{K+1})}\right].
\end{aligned}
\tag{1.36}
$$

This result is good for showing how performance changes as $K$ is increased to the infinity (AWGN channel).

When $K$ is approached to the infinity (AWGN channels), the probability of error becomes

$$
\begin{aligned}
P_2(SNR) &= M(t=-SNR) \\
&= \frac{1}{(1-t(\frac{1}{K+1}))^{n/2}} \exp\left[\frac{t N_r\left(\frac{K}{K+1}\right)}{1-t\left(\frac{1}{K+1}\right)}\right]\Bigg|_{t=-SNR,K=\infty} \\
&= \exp\left(-SNR \cdot N_r\right) \\
&= \exp(-SNR_t).
\end{aligned}
\tag{1.37}
$$

For $N_r = 2$, the Chi-square random variable has four degrees of freedom, i.e.,

$$a = |\alpha_1|^2 + |\alpha_2|^2 = r_1^2 + r_2^2, \tag{1.38}$$

where

$$\alpha_i = r_i e^{j\theta_i} \text{ for } i=1, 2, \ldots \tag{1.39}$$

Again, the Ricean factor is $K = \frac{m^2}{s^2}$.

The moment generating function for four degrees of freedom can be obtained from (1.7) and using the un-normalized representation (such that $s^2 = 1/2$) it is

$$M(t) = \frac{1}{(1-t)^2} \exp\left[\frac{t2K}{1-t}\right]. \tag{1.40}$$

The probability expression is again

$$
\begin{aligned}
P_2(SNR) &= \int_0^\infty Q(\sqrt{2 \cdot SNR \cdot y})\, p_Y(y) dy \\
&\le \int_0^\infty \exp(-SNR \cdot y)\, p_Y(y) dy \\
&= M(t=-SNR),
\end{aligned}
\tag{1.41}
$$

except that the moment generating function is for Chi-square random variable with four degrees of freedom. Using (1.40), we have

$$P_2(SNR) \le \frac{1}{(1+SNR)^2} \exp\left[\frac{-2SNR \cdot K}{1+SNR}\right]. \tag{1.42}$$

Setting $K = 0$ for Rayleigh fading channels, we note that

$$P_2(SNR) \le \frac{1}{(1+SNR)^2} \approx SNR^{-2}. \tag{1.43}$$

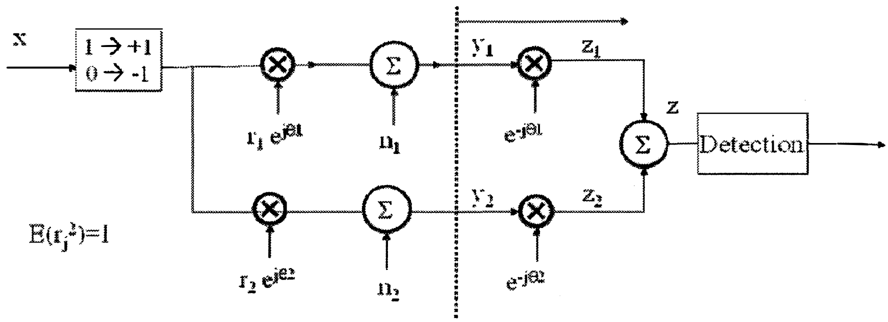
## HOMEWORK EXERCISES

1. (Why use moment generating and characteristic functions?) Let $X_1$ and $X_2$ be i.i.d. random variables. Each takes values 1, 2, 3 with a corresponding distribution [0.3 0.4 0.3].
   (a) Let $Y = X_1 + X_2$ and obtain the distribution for Y.
   (b) Obtain the characteristic function of Y.
   (c) Obtain the distribution of Y from the characteristic function of Y.
   (d) Now check if your procedure can be used to find the distribution for Y := $X_1 + \ldots + X_{100}$. If yes, describe how.
2. (P(e) for Rayleigh fading) Obtain the average error probability P(e) of BPSK transmission for Rayleigh fading channel. Assume coherent receiver. Do it for both $N_r = 1$ and $N_r = 2$.
   (a) Obtain the pdfs of the overall channel gain of the maximal ratio combining receiver.
   (b) Use the direct integration method for average P(e) and show that your results are consistent with (1.24) and (1.43).
3. Let $P = \sum_{i=1}^{2n} X_i^2$ where $X_i \sim \mathcal{N}(m = 1, s^2 = 1)$. What kind of random variable is $P$?
   (a) Write its pdf.
   (b) Find its mean, second moment, variance, moment generation function, and characteristic function of $P$.
4. Evaluate the following integration

$$P(x) = \int_0^{\infty} Q\left(\sqrt{xy}\right) p_Y(y)dy, \tag{1.44}$$

at the Ricean factor $K = 3$, where $p_Y(y)$ is the pdf of random variable $Y$. $Y = \sum_{i=1}^{3} |a_i|^2$ and $a_i = (\alpha X_1 + \beta) + j(\alpha X_2 + \beta)$. $X_1$ and $X_2$ are independent Normal distributed.

5. (Diversity receiver) Consider $N_r = 2$ case shown below. In this problem, analyze the receiver where only phase equalization is done at each branch. Obtain the expression for P(e) averaged over the fading. Explain the difference with the maximal ratio combining receiver. Are they the same in performance?




6. (MATLAB, Ricean, Craig Identity) Suppose BPSK transmission and the maximal ratio coherent detection at the receiver with $N_r$ number of receive antennas, and plot the probability of bit error for Ricean fading channel with $K$=0, 1, 5, 10 and for $N_r$ =1, 2, 3, as the function of average SNR per channel (0 dB ~ 30 dB). First use the Chernoff bound and then the Craig-identity. Compare the results from the two methods. Also, compare the result with the BPSK transmission over AWGN channel. Comments on your results.

## REFERENCES

[1] JOHN G. PROAKIS AND MASOUD SALEHI, DIGITAL COMMUNICATIONS, 5TH EDITION, MCGRAW HILL, BOSTON, USA.

# Third Possible Class Project + Class Note

Compare the method given here in this note with the new Gallager bounds for fading channel.

Reference

[1] X. Wu, H. Xiang, and C. Ling, "New Gallager Bounds in Block-Fading Channels," IEEE Trans. Information Theory, Feb. 2007.

- The Spherical bound given in Theorem 2 seems to be most simple. So, as a class project, try to compare with it.
- Make note of the complexity
- Read the conclusion of the paper. It would be interesting to pay attention to the part where the authors says their method is still too complicated since they have to evaluate the bound per distance profile base [h] and thus complexity of bound is still very high.
- See if you can incorporate the idea learned in the class and improve the bound in terms of complexity while not losing the tightness of the bound.

## Class Note begins here.

In class, we learned that the word error probability can be expressed as the following

$$P_w \le \sum_h A_h \binom{N}{h}^{-1} \underbrace{\sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^B\right)^{d_j}}_{=:\phi(h)} . \tag{1}$$

For MIMO fading case, the beta is the pairwise symbol error averaged over the fading matrix, i.e.,

$$\beta_j^B := \prod_{n=1}^{N_r}\prod_{m=1}^{N_t} \frac{1}{1+\frac{\rho_s}{4}\lambda_m^{(j)}} \exp\left( -\frac{F_r \frac{\rho_s}{4}\lambda_m^{(j)}}{1+\frac{\rho_s}{4}\lambda_m^{(j)}} \right). \tag{1.2}$$

We now define a sequence of utility coefficients $\phi(h)$, i.e.,

$$\phi(h) := \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^B\right)^{d_j} . \tag{3}$$

Then, the union bound becomes

$$P_w \le \sum_h A_h \binom{N}{h}^{-1} \phi(h). \tag{4}$$

From (4), we note that once $\phi(h)$ is given, the union bound can be obtained easily. However, calculation of $\phi(h)$ using (3) is not easy because the set partition operation according to each distance profile, i.e., $\mathbf{d}(h) \in \Omega_h$, is very complex and difficult. Recall that $\Omega_h$ is the collection of all distance profiles for a Hamming weight $h$. The cardinality of $\Omega_h$ grows prohibitively large when $h$ approaches $N/2$ (and when $N$ is large), and thus a brute-force partitioning would take large amount of time and computation.

We solve this problem by resorting to the multinomial theorem given without a proof.

*The Multinomial Theorem.* Let $m$ and $n$ be positive integers. Let $\mathcal{A}$ be the set of vectors $\mathbf{x} = (x_1, \ldots, x_n)$ such that each $x_i$ is a nonnegative integer and $\sum_{i=1}^{n} x_i = m$. Then, for any real numbers $p_1$, …, $p_n$,

$$(p_1 + \cdots + p_n)^m = \sum_{\mathbf{x} \in \mathcal{A}} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}.$$

In order to apply the multinomial theorem in our problem, we first consider the following polynomial with respect to the dummy variable z,

$$\gamma(z) := \sum_{j=1}^{J} \beta_j^B z^{w_j}. \tag{5}$$

Note that this polynomial is readily computable once we know the constellation map and the Chernoff upper bound on the PEP for each IST symbol. IST symbol here implies the internal space-time symbol. Thus, indeed the framework we have here can be extended to space-time symbol set as well. But for single antenna case, an IST symbol set is just the regular symbol set we know such as BPSK or 8-PSK. All we need to know is the mapping table between binary strings and symbols.

The transmitted IST symbol is assigned to the all-zero bit string of length $N_b$. The $j$-th IST symbol is mapped from the bit-string $b_{(j)}$ which has weight $w_j$. Now write the result of the expansion of the polynomial in the ascending power of $z$ and consider taking the $D$-th power of this polynomial. Then, the sequence of coefficients $\phi(h)$ will appear as the power expansion coefficients with respect to the dummy variable $z$, i.e.,

$$\begin{aligned}(\gamma(z))^D &= \sum_{h=0}^{N} \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^{B}\right)^{d_j} z^h \\ &= \sum_{h=0}^{N} \phi(h) z^h \end{aligned} \tag{6}$$

Making use of this result, one can obtain the expansion coefficient $\phi(h)$ from $\gamma(z)$ in the following routine:

**Step-0**: Define an IST symbol set and a constellation map.

**Step-1**: Evaluate the pairwise error probabilities $\beta_j^B$, $j$=1, 2, …, $J$, at a particular SNR $\rho_s$.

**Step-2**: Obtain the Hamming weight $w_j$ of each bit-string $b_{(j)}$ from the constellation map.

**Step-3**: Obtain the base polynomial $\gamma(z)$ by arranging the polynomial with respect to ascending power of $z$. Note that the degree of the polynomial is $N_b$. Thus, there are $N_b+1$ polynomial coefficients, i.e., $\gamma(z)=\sum_{i=0}^{N_b}\gamma_i z^i$. Let's put them in to a vector $\boldsymbol{\gamma}=\left[\gamma_0\ \gamma_1 \cdots \gamma_{N_b}\right]$.

**Step-4**: Take the $D$-fold convolution of the vector $\boldsymbol{\gamma}$, which is of length $N_b+1$. The outcome is the sequence of coefficients $\langle\phi(h), h=0,1,2,\cdots,N\rangle$ of length $DN_b+1=N+1$. Once the sequence of expansion coefficients $\langle\phi(h), h=0,1,2,\cdots,N\rangle$ is known, the union bound can be calculated readily.

Summarizing what we have so far, for the transmission of a linear ($N$, $K$) block code with a given distance spectrum $\langle A_h\rangle$ over the block Ricean fading ($N_t$, $N_r$) MIMO channel with the Ricean factor $F_r$, the union upper bound on the probability of making codeword errors is given by (4) and the sequence of expansion coefficients $\phi(h)$ is obtained from the multinomial expansion (6) and $\beta_j^B$ given.

It should be noted that the routine provides a compact and easy-to-compute method to obtain the performance of the concatenated coded modulation system. We note that the multinomial expansion (6) has been the key step to reduce the complexity of the union bound. In addition, (6) provides the following insights: The base polynomial, $\sum_{j=1}^{J}\beta_j^B z^{w_j}=\sum_{w=0}^{N_b}\gamma_j z^j$, is the union bound for a single IST channel usage. For a single block channel usage, a total of $N_b$ bits are transmitted via an IST symbol. Evaluating the base polynomial at $z=1$, a union upper bound on the probability of symbol error for the transmission of a single IST symbol is obtained. The result in (6) provides a easy-to-compute method to calculate the union upper bound on the maximum likelihood receiver for the concatenated transmission in which there are $D$-consecutive IST symbols mapped from a long block code whose block length is $DN_b$. Thus, (6) provides information as to (1) how the union bound for a single channel use is related to the union bound of the concatenated system in $D$-consecutive channel uses, and (2) how the $D$-th order time-diversity can take its effect into the overall performance.

### *A. Union Bounds for Multilevel/phase modulation over the AWGN channel*

The result in section can be used to obtain special cases such as the results for the independent fading and for the quasi-static fading channels by manipulating the pertinent system parameters (by either $T_i$ =1 or $D = 1$ respectively). In addition, it can be brought down to the union bound result for the AWGN channel.

One can use alternate pdf and MGF for the Ricean variable defined earlier in our class notes by fixing the mean of $Y$ to be equal to 1, i.e., $E(Y) = 2(s^2 + m^2) = 1$. Then, the MGF can be written as

$$M(t) = \left( \frac{1}{1 - t\left(\frac{1}{(F_r+1)}\right)} \right) \exp\left( \frac{t\left(\frac{F_r}{(F_r+1)}\right)}{1 - t\left(\frac{1}{(F_r+1)}\right)} \right). \tag{7}$$

The expression (7) is convenient for letting $F_r = \infty$ and reducing our main result to the AWGN channel case. The moment generating function becomes $M(t) = \exp(t)$ with $F_r = \infty$. Then, we obtain the union bound expression for the AWGN channel by substituting $t = -\frac{\rho_s}{4}\lambda^{(j)}$ into $M(t)$ such that we have $\beta_j^B = \exp\left(-\frac{\rho_s}{4}\lambda^{(j)}\right)$ where $\lambda^{(j)}$ is $\lambda^{(j)} := \left(s_{(*)} - s_{(j)}\right)^* \left(s_{(*)} - s_{(j)}\right)$ for a $J$-ary symbol set $\{s_{(j)}, j = 1,2,\cdots,J\}$.

### *B. 8 PSK symbol set over the AWGN channel*

Let us consider the 8-PSK symbol set, the eight equi-spaced symbols on the unit circle. The symbols are named as $s_{(1)}$, $s_{(2)}$, ..., and $s_{(8)}$ taking the counter clock wise turn starting from the first symbol at $s_{(1)} = 1$. The Gray map is used such that (000), (001), (011), (010), (110), (111), (101) and (100) are given for the eight signals respectively. Thus $J$ = 8, and $N_b = \log_2(J) = 3$

[bits/symbol]. We let $s_{(*)} = s_{(1)} = 1$ (000). The received signal is modeled as $r_d = \sqrt{\rho_s} s_d + z_d$ where $z_d$ is zero mean and unit variance complex Gaussian with independent real and imaginary part. There are four distinct symbol distances from $s_{(*)} = 1$ to the rest of 7 symbols. Thus, there are four $\lambda$ 's such as 0.5858, 2.0, 3.4142 and 4.0. There are four corresponding distinct $\beta_j^B$. Assuming $\rho_s = 10$ (10dB), they are $\exp(-\frac{5.858}{4})$, $\exp(-\frac{20}{4})$, $\exp(-\frac{34.14}{4})$, and $\exp(-\frac{40}{4})$.

## Numerical Evaluation Results Compared with Simulation

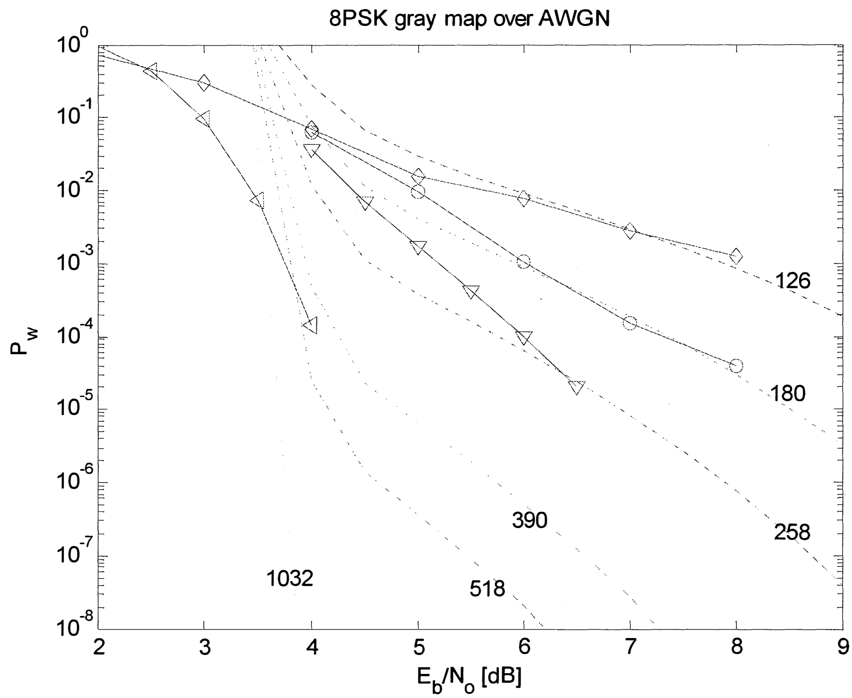



Figure 1: Union Bound on Word Error Probability for 8 PSK modulation over the AWGN channel. Numbers shown inside the figure are the block lengths of pertinent (3, 6) Gallager codes. Union bounds are shown as the dashed lines; while the simulation results for block length 126, 180, 258 and 1032 are shown as solid lines with the plotting symbols.

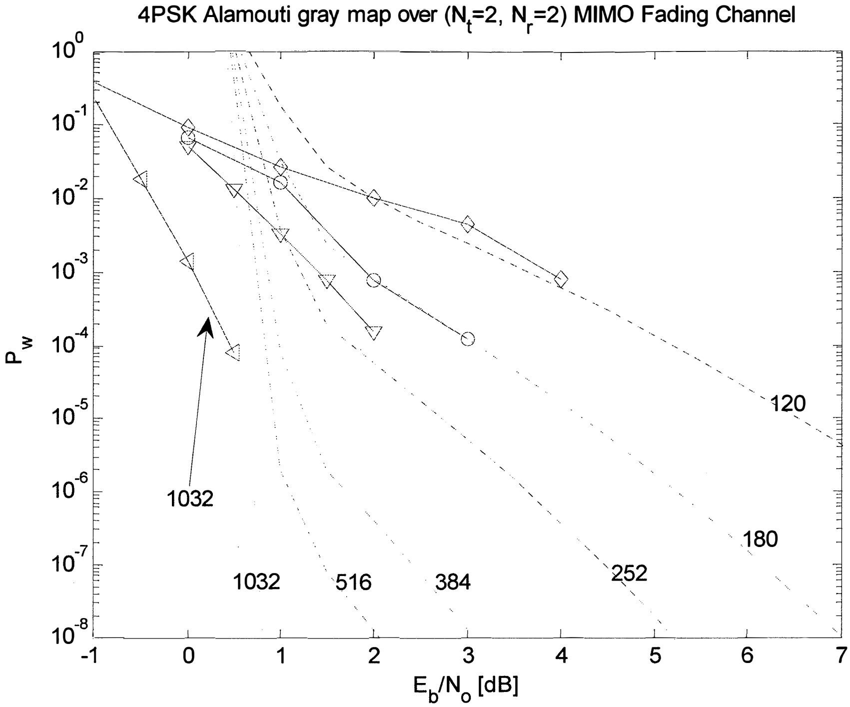



Figure 2: Union Bound on Word Error Probability for 4 PSK Alamouti block code over the (2, 2) MIMO block fading channel. Numbers shown inside the figure are the block lengths of pertinent (3, 6) Gallager codes. Union bounds are shown as the dashed lines; while the simulation results for block length 120, 180, 252 and 1032 are shown as solid lines with the plotting symbols.

# Chapter

# Union Bounds for MIMO Channels

**The use of turbo-like codes—such as the turbo codes and the low-density parity-check codes--over multi-input multi-output (MIMO) channels has recently gained significant interests from the communications and coding theory communities. The turbo-like code takes the role of the *outer* code while the space-time block code does that of the *inner* code. The outer linear block code transforms into the sequence of short internal space-time blocks according to a fixed constellation mapping rule. The soft-input soft-output decoder can be combined with a constellation de-mapper which generates *aposteriori* log-likelihood-ratios for turbo-iterations. This transceiver scheme has shown promising performance results. In this chapter, a novel performance evaluation framework based on the maximum-likelihood union-bound is proposed for the analysis of the concatenated coding scheme. The analysis is combinatorial. The codebook is decomposed into mutually exclusive subsets in such a manner that the pairwise error probability for any codeword in the subset is identically the same. The union bound is then obtained as the sum of all distinct pairwise error probabilities each of which is weighted by the cardinality of the subset. A union bound in its basic form may be useless if too difficult to be evaluated, especially for long block codes; the idea of using polynomial expansion is introduced and the union bound is simplified into a form evaluated efficiently. In addition, the random coding *error exponent* is obtained for fully random outer block codes. System simulation results and the numerical evaluations of the bounds are compared for different channel scenarios as well as for different space-time transmission schemes. The results indicate that the derived bounds are useful as benchmarking tools for the practical turbo-iterative decoding and detection receiver.**

# CHAPTER PREREQUISITE

1. The paper by Telatar on MIMO capacity
2. The paper by Tarokh et al on the calculation of pairwise error probabilities over fading channels.
3. The paper by Alamouti for transmit diversity
4. The paper by Hassibi and Hochwald on the linear dispersion code

# I. INTRODUCTION

Recently the low-density parity-check (LDPC) codes and the turbo-codes have been proposed as the outer code of a concatenated scheme which drives space-time block transmissions as the inner code for multi-input multi-output (MIMO) systems [i, ii, iii]. One of the most desirable characteristics of the LDPC or the turbo-like codes is that the complexity of the decoder grows only linearly to the length of the code, thanks to the message-passing decoding-algorithm. This advantage can be utilized to attain a capacity approaching performance in many kinds of channels by letting the block length of the code grow toward infinity. As reported in the literature [iv], the performance of the LDPC or turbo codes approaches the capacity limit within a fractional dB at the block length of more than a few hundred thousands with the binary modulation over the AWGN channels. When employed for the MIMO channels, the use of the outer block code, via the soft-input soft-output message passing decoder on the Tanner code-graph, can be used as means to exploit the diversity benefit available in space- and time-domains because a codeword, or the space-time transmitted matrix mapped from it, is spread out in all directions. As the result, if we let the length of the outer code grow to infinity, the performance of the concatenated scheme is expected to be brought very close to the MIMO channel capacity.

On the other hand, there are practical reasons to limit the length of the outer code. For example, in *IEEE* 802.11n standardization activities (see [v] for example), the length of the binary LDPC code is proposed to be about two thousand. A simple calculation would show that with the maximum clock speed of several hundred MHz (assuming state-of-the-art VLSI systems), the desired transmission speed of a few hundred mega bit-per-second is not an easy objective to be met. Due to decoding delay, it seems that the implementation of the receiver is quite a challenging task already at the length of two thousand. This implies that the performance evaluation of a moderate length concatenated block code is to be of our interest for high speed MIMO transmission applications. An accurate performance bound, therefore, by which system simulation results can be contrasted, shall serve as an invaluable tool. The performance analysis based on the constraint channel-capacity [i] can serve as the ultimate bound; but at the block length of a few thousands they are usually too loose to be useful. The maximum likelihood union bounds obtained in the chapter scales well with the block length. The other motivation for union bound is that we may be able to use them as an optimization tool for designing a better performing space-time block code or concatenated coding scheme.

In this chapter, we study a set of maximum likelihood (ML) union bound analysis techniques, tailored out for each class of MIMO fading channels. In [vi], the authors have introduced the union-bound based ML upper bounds on BPSK modulated MIMO systems over independently fading channels: the LDPC code is used as the outer code, and the *orthogonal space-time block codes* (OSTBC) [vii] and the *direct-transmission* scheme (or better known as the V-BLAST transmission scheme [viii]) are compared as the inner coding method. In this chapter, we generalize the analysis framework so that different space-time block transmission schemes, such as OSTBC, V-BLAST transmissions [viii] and Linear Dispersion (LD) codes [ ix ], can be uniformly accommodated. In addition, we extend the union bound analysis for block- and quasi-static fading channels.

Concatenated Coding Scheme for MIMO Transmission

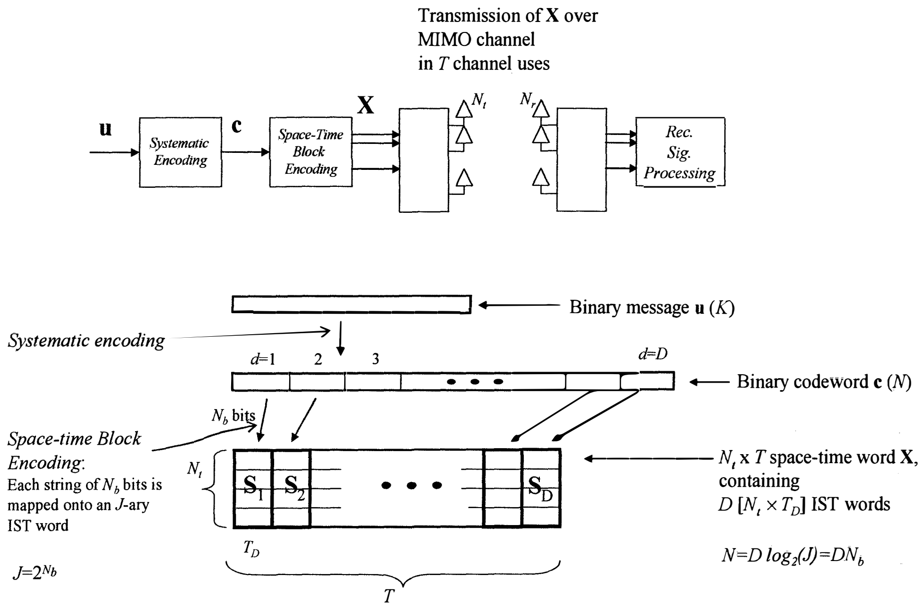



Figure 1: Transmission of concatenated code

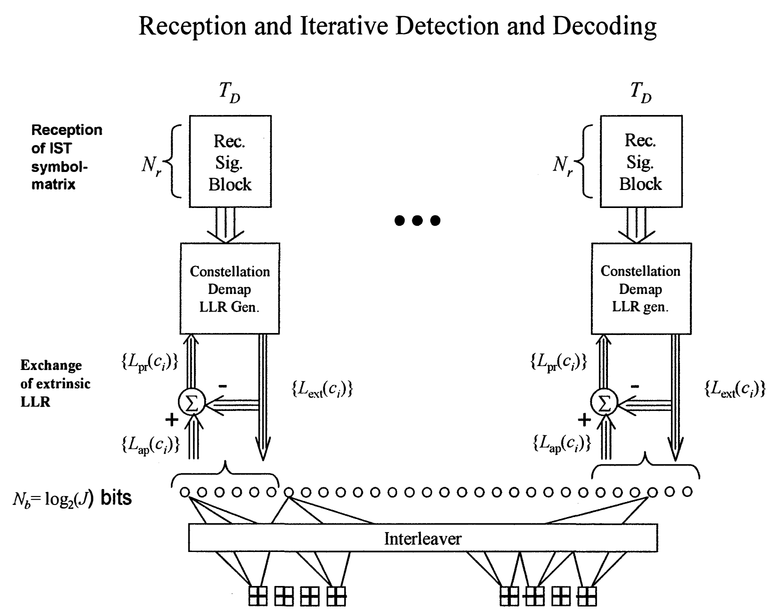



Figure 2: Reception and Iterative Detection and Decoding Scheme

The union bound in its basic form is the summation of every pairwise error probability (PEP). Using the distance spectrum [x], it can be simplified as the sum of every distinct PEP each of which is weighted by a codeword multiplicity. The multiplicity is the number of the codewords (pairwise error events) in a codebook that result in an identical PEP. For BPSK modulated coding system over AWGN channels, for example, each distinct pairwise error event can be delineated by Hamming distance between pairs of codewords. For linear codes, a codeword multiplicity is simply the number of codewords with a certain Hamming weight. Note that the straightforward summation of all pairwise error probabilities shall be computationally burdensome, even for the block length of a thousand; a distance-spectrum based union-bound would be easier to evaluate, even for a block length approaching infinity.

A similar but a careful approach should be taken to develop a computationally feasible union bound expression for the concatenated coding scheme for the MIMO system. The overall transmission scheme is depicted in Figure 1. The constellation mapping from the binary outer codeword to the space-time word is one-to-one correspondent. The key step is thus to partition the codebook into subsets of outer codewords. In each subset, the pairwise error event must lead to an identical pairwise error probability. Since the pairwise error probability for the MIMO system is due to the Euclidean distance of the pair of space-time words, knowledge of the distance spectrum on the binary code alone is not sufficient.

It proves to be that the outer codebook can be partitioned further, in addition to the one based on the Hamming weight. The cardinalities of the partitioned sets are not easily computable for a specific outer codebook. Thus, we resort to an ensemble average over all codebooks. Then, the calculation of codeword multiplicities becomes combinatorial problems by making use of the statistical property (see Theorem I) of the ensemble. An analogous partition for the space-time codebook can be conducted taking into account the one-to-one correspondence relationship between the outer code and the space-time code.

Union bound methods for fading channels have been studied extensively in the past [xi] and several have been introduced recently for space-time block codes. Some have focused on short space-time block codes (for example see the most recent one [xv]) and others do on space-time trellis codes [xii, xiii, xiv]. For trellis codes, the idea of truncation to a certain maximum number of pairwise error terms is used to obtain a computationally efficient bound. For short space-time block codes, maximum likelihood detection is manageable; thus computational efficiency needs not be of a focus either. In addition, one can exploit the availability of the geometric information of the space-time block code, and apply more sophisticated bounding techniques such as the Bonferroni-type bounds to obtain tight upper- and lower-bounds [xv].

We take the perspective of making use of a powerful outer block code to interface the MIMO block-channel-uses. Depending on the statistical information gathered on the fading channel, one may adaptively choose a certain type of space-time block code. For the block fading channel, say with a coherent time of $T_D$ channel uses, one may use a rate-optimized space-time code with dimension [$N_t$ x $T_D$]. A space-time block code which is a collection of *J-ary* [$N_t$ x $T_D$] space-time block matrices can then be designed with the objective of maximizing the mutual information theoretic channel capacity for a single block-channel-use (see [ix] for example). In this setting, the role of the outer block code, which is long enough to cover many independent block-channel-uses, can be used as a means to achieve a coding performance close to the channel capacity.

The rest of the chapter is organized as follows. In Section II, the system model is described. The ensemble of outer code and its statistical properties are developed in Section III. The union bound on the block fading channel is given in Section IV. Union bound on independent fading and slow fading cases are given in Section V. In section VI, the turbo-iterative detection and decoding algorithm is described. In Section VII, simulation results and numerical evaluation of the bounds are compared. Finally, we conclude in Section V.

## II. SYSTEM OF INTEREST

Consider the MIMO system with $N_t$-transmit and $N_r$-receive antennas, as illustrated in Figure 1. A sequence **u** of $K$ information bits is encoded into a binary codeword **c** of length $N$ by the systematic encoder. Let $R_c = K/N$ be the rate of the block code. A codeword **c** of length $N$ is mapped onto a [$N_t$ x $T$] space-time transmission matrix **X** in a one-to-one correspondent relationship. $T$ is the total number of channel uses for a codeword to be transmitted. Compared to the whole space-time codeword, we have shorter space-time symbol-matrices, named here as the Internal Space-Time (IST) codewords. We will take two slightly different approaches to describe the mapping relationship from the codeword **c** to the space-time transmission matrix **X**.

The first approach goes as follows: Taking $N_c$ coded bits at a time and mapping them to a channel-symbol out of an $M$-ary digital constellation ($M = 2^{N_c}$), we can transform the codeword into a sequence of channel-symbols of length $N/N_c$ (which is assumed an integer without loss of generality). Then, we put symbols into [$N_t$ x $T_D$] (internal space-time) IST symbol-matrices each of which carries $N_tT_D$ $M$-ary channel-symbols; and thus each IST symbol-matrix carries $N_cN_tT_D$ coded bits ($R_cN_cN_tT_D$ message bits). Thus, we can consider an IST matrix-constellation of size $J = 2^{N_tT_DN_c}$.

In the second approach, we have a set of [$N_t$ x $T_D$] IST symbol-matrices of size $J$—a signal-matrix constellation. Each [$N_t$ x $T_D$] symbol-matrix carries $N_b := \log_2(J)$ number of bits. The second approach is more simple and general.

The second approach is more suitable to model the linear dispersion codes and orthogonal space-time block codes.

We use the notation $\mathbf{s}_{(j)}$, $j$ = 1, 2, …, $J$, to denote the $J$-ary [$N_t \times T_D$] IST codeword. There are a total of $D$ IST words within a space-time codeword **X**:= [$\mathbf{S}_1$ $\mathbf{S}_2$ … $\mathbf{S}_D$ ], indexed by $d$ = 1, 2, …, $D$ = $T/T_D$ (assumed an integer WLOG). We use $\mathbf{S}_d$ as a time-parameterized value-holder for an IST word: the $d$-th IST word $\mathbf{S}_d$ takes a value $\mathbf{s}_{(j)}$ from the $J$-ary IST constellation.

Assume each ST codeword **X** is selected equi-probably and transmitted over the *flat* fading MIMO channel. Then, the [$N_r$ x $T_D$] receive signal $\mathbf{R}_d$ due to the $d^{th}$ IST block transmission is given by

$$\mathbf{R}_d = \sqrt{\rho_s}\mathbf{H}_d\mathbf{S}_d + \mathbf{Z}_d, \quad d = 1, 2, \cdots, D \tag{1}$$

where $\rho_s = \rho_b R/M$ and $\rho_b = E_b/N_o$ are the symbol-energy and the bit-energy to noise-power-spectral-density ratios respectively; $R{=}DN_bR_c/T$ is the transmission rate in bits-per-channel-use (bpcu); $E_b$ is the information bit energy; $N_0/2$ is the two-sided power-spectral-density of the white Gaussian noise present at the receiver. $\mathbf{Z}_d$ denotes the [$N_r \times T_D$] noise matrix each of whose entry is mutually-independent complex additive-white-Gaussian noise with zero mean and variance 0.5 per dimension. $\mathbf{H}_d = \{\alpha_{n,m}^d\}_{n,m}$ denotes the [$N_r \times N_t$] fading matrix. Fading coefficients $\alpha_{n,m}^d$ for different row and column indices $n$ and $m$ are mutually independent identical Ricean distributed. The channel matrix is held fixed during each block of $T_D$ channel uses. The block channels parameterized by $d$ are mutually independent.

For slow fading channel, the fading channel matrix is held fixed throughout all $d$. For fast (independent) fading channel, each and every channel use, we have independent fading.

The probability density function of any of the fading gain magnitude $\alpha := \left|\alpha_{n,m}^d\right|$ is given by

$$p_\alpha(\alpha) = 2\alpha\exp\left(-\alpha^2 - F_r\right)I_0\left(2\alpha\sqrt{F_r}\right), \quad \alpha \ge 0 \tag{2}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind and $F_r$ is the Ricean factor. The moment generation function (MGF) associated with $\left|\alpha_{n,m}^d\right|^2$ is given by

$$M(s) = \frac{1}{1-s}\exp\left(F_r\frac{s}{1-s}\right). \tag{3}$$

### III. ENSEMBLE OF OUTER CODES

The problem of codeword enumeration, for a given one-to-one correspondence relationship, from the block code to the space-time code needs to be solved for the derivation of the union bound on the concatenated code. We start with the assumption that the distance spectrum of the outer code is available (see [xvi]) and attempt to come up with something equivalent to that for the space-time code. For this, we first investigate the statistical properties of the following outer block codes:

- The ensembles of LDPC codes, each of which is distinctively specified by a set of fixed parameters--block length $L$, variable- and check-node degree distributions. Note that this set-up can be useful for defining an ensemble for both the Gallager codes [xvii] and the irregular codes [xviii, xix].
- The ensembles of concatenated turbo codes (or any other linear block code), each generated by a specific random interleaver.
- The ensemble of *fully* random block codes, defined by the fixed block length $N$ and code rate $R_c = K/N$.

Now consider an ensemble $\mathbb{C}$ of codes. Assume any code $\mathcal{C}_{sel}$ in the ensemble $\mathbb{C}$ is selectable with equal probability, i.e.

$$\Pr\left(\mathcal{C}_{sel} \text{ is selected}\right) = |\mathbb{C}|^{-1}, \text{ for } \forall\, \mathcal{C}_{sel} \in \mathbb{C} \tag{4}$$

where we use $|\bullet|$ to denote the cardinality of the set. For a fully random ($N$, $K$) block code, we note $|\mathbb{C}| = \binom{2^N}{2^K}$. Let us denote the average distance spectrum of a code in the ensemble as $\{A_h\}$. That is, $A_h$ is the number of codewords of weight $h$, averaged over all codes in the ensemble.

For the interested ensembles given above, we have the following statistical properties summarized in the following theorem:

Theorem-I: If $A_h > 0$ for a certain h, each of the $\binom{N}{h}$ distinct binary sequences of length N and weight h is a valid codeword in a certain number of codes in the ensemble. The probability of any of these binary sequences of length N appearing in any randomly selected code $\mathcal{C}_{sel}$ as an element codeword is equal.

Proof: Both LDPC and turbo codes are *linear* block codes. Hence both codes can be completely defined by their associated parity-check matrices. Instead of the ensemble $\mathbb{C}$ of codes, we may equivalently consider the corresponding ensemble $\mathcal{H}$ of parity-check matrices without loss of generality.

In case of LDPC codes, it is clear the ensemble $\mathcal{H}$ is closed under column permutation: any column permutation of a particular parity-check matrix randomly selected from $\mathcal{H}$

produces another matrix belongs to it (any column permutation does not change the variable- and the check-node degree distributions). The same statement applies to any ensemble of concatenated codes equipped with the random interleaver which plays the role of arbitrarily permuting the parity-check matrix in a column wise manner.

$A_h > 0$ means that at least one codeword, say $c_{h,1}$, of weight $h$ exists in a certain number of codes in the ensemble. Assume $\mathbf{c}_{h,2}$ is an arbitrary permutation of $\mathbf{c}_{h,1}$. That is, $\mathbf{c}_{h,2} = \pi(\mathbf{c}_{h,1})$, where $\pi(\bullet)$ is the associated column permutation pattern. Denote $\mathcal{H}_1$ and $\mathcal{H}_2$ as the sets of all parity-check matrices in $\mathcal{H}$ that $\mathbf{c}_{h,1}$ and $\mathbf{c}_{h,2}$ satisfy, respectively,

$$\mathcal{H}_1 := \left\{ H \middle| H \in \mathcal{H},\, H\mathbf{c}_{h,1}^T = \underline{0} \right\} \tag{5}$$

$$\mathcal{H}_2 := \left\{ H \middle| H \in \mathcal{H},\, H\mathbf{c}_{h,2}^T = \underline{0} \right\} \tag{6}$$

The cardinality $|\mathcal{H}_1|$ of $\mathcal{H}_1$ is a positive integer. Based on the statement given in the last paragraph, it is clear there is a one-to-one correspondence between $\mathcal{H}_1$ and $\mathcal{H}_2$, i.e.

$$\mathcal{H}_2 = \left\{ \pi(H) \middle| H \in \mathcal{H}_1 \right\} \tag{7}$$

Therefore, $N_h := |\mathcal{H}_2| = |\mathcal{H}_1| > 0$. Since each of the $\binom{N}{h}$ binary sequences of weight $h$ can be regarded as a permutation $\mathbf{c}_{h,2}$ of $\mathbf{c}_{h,1}$, the first statement of the theorem is proved. With the assumption in (4), the probability for each $\mathbf{c}_{h,2}$ of these sequences to be included in the randomly selected code $\mathcal{C}_{sel}$ is the same (i.e. equal probability) and given by

$$\Pr\left(x_{h,2} \in \mathcal{C}_{sel}\right) = N_h |\mathbb{C}|^{-1}. \tag{8}$$

END

For the ensemble of fully random codes, it can be shown that a similar proof can be given although the codes are generally not linear.

Theorem-I is essential for calculation of a “distance profile” of the ST code in the following manner.

Consider the binary sequence of length $N$ as a serial concatenation of a number $D$ of sub-sequences of length $L/D=N_b$ (Assumed integer WLOG). We denote all $J = 2^{N_b}$ distinct sub-sequences as $b_{(1)}$, $b_{(2)}$, … , $b_{(J)}$, their weights as $w_1$, $w_2$, …, $w_J$, and the numbers of their appearances in the whole sequence as $d_1$, $d_2$, …, $d_J$, respectively. Note from the definition that

$$d_j \in \{0,1,2,...,D\} \text{ and } \sum_{j=1}^{J} d_j = D. \tag{9}$$

We may collect these numbers in an array $\mathbf{d} = (d_1, d_2, ..., d_J)$ and name it as the *pairwise distance profile* (PDP). Then, we want to compute the number of distinct codewords which possess the same PDP $\mathbf{d} = (d_1, d_2, ..., d_J)$ in a code, which is a difficult task if we specify a particular code. An easier approach is to have the number averaged over the code ensemble.

We call this $A_{\mathbf{d}(h)}$ as compared to the average number of codewords $A_h$ with Hamming weight $h$.

Note that, all codewords that have the same PDP have the same Hamming weight $h = \Sigma_j d_j w_j$.

According to Theorem-I and resorting to the combinatorial techniques, the fraction of words among all possible words of weight $h$ which have the PDP $\mathbf{d}(h)$ is given by

$$\Pr(\mathbf{c} \text{ has a metric } \mathbf{d} | \mathbf{c} \text{ is of weight } h) = \binom{N}{h}^{-1} \binom{D}{\mathbf{d}(h)}, \tag{10}$$

where $\binom{\Sigma x_i}{\underline{x}} \triangleq \binom{\Sigma x_i}{x_0, x_1, \cdots x_{n-1}} = \frac{(\Sigma x_i)!}{\Pi x_i!}$ denotes the multinomial coefficient. The collection of all possible combinations of $\mathbf{d}$ contributing to a Hamming weight $h$ can be defined as

$$\Omega_h := \left\{ \mathbf{d}(h) \middle| d_j \in \{0,1,2,...,D\}, \sum_{j=1}^{J} d_j = D, \sum_{j=1}^{J} d_j w_j = h \right\}. \tag{11}$$

Note, it can be verified that $\sum_{\mathbf{d}\in\Omega_h} \binom{N}{h}^{-1} \binom{D}{\mathbf{d}} = 1$.

From (10), the average number $A_{\mathbf{d}(h)}$ is given by,

$$A_{\mathbf{d}(h)} = A_h \binom{N}{h}^{-1} \binom{D}{\mathbf{d}(h)}. \tag{12}$$

We note that $\mathbf{d}(h)$ can be used as a criterion to further divide the codebook in addition to the partition made according to the Hamming weight.

# IV. THE UNION BOUND ON THE BLOCK FADING CHANNEL

We will consider the system depicted in Figure 1. We first evaluate the union bound over the general block fading channel model. The fast (independent) and quasi-static fading channels are special cases of the block fading channel model.

## A. The Pairwise Error Probability Averaged over Channel

We assume perfect channel state information, **H**, available at the receiver. Then, the pairwise error probability between any two space-time codewords $\mathbf{X} = [\mathbf{S}_1\ \mathbf{S}_2\ \ldots\ \mathbf{S}_D]$ and $\mathbf{X}' = [\mathbf{S}'_1\ \mathbf{S}'_2\ \ldots\ \mathbf{S}'_D]$ conditioned on a channel realization is given by

$$P(\mathbf{X} \rightarrow \mathbf{X}'|\mathbf{H}) \leq \exp\left(-d^2(\mathbf{X},\mathbf{X}')\frac{\rho_s}{4}\right), \tag{13}$$

where $d(\mathbf{X},\mathbf{X}')$ is the Euclidean distance between **X** *and* **X'**. It is expressed as

$$\begin{aligned} d^2(\mathbf{X},\mathbf{X}') &= \sum_{n=1}^{N_r}\sum_{t=1}^{T}\left|\sum_{m=1}^{N_t}\alpha_{n,m}(t)\left(\mathbf{X}_{m,t}-\mathbf{X}'_{m,t}\right)\right| \\ &= \sum_{n=1}^{N_r}\sum_{d=1}^{D}\sum_{t=1}^{T_D}\left|\sum_{m=1}^{N_t}\alpha_{n,m}^{d}\left(\mathbf{S}_{d,m,t}-\mathbf{S}'_{d,m,t}\right)\right| \end{aligned}, \tag{14}$$

where in the first equation $\mathbf{X}_{m,t}$ and $\mathbf{X}'_{m,t}$ are the $m^{th}$ row and $t^{th}$ column element of **X** and **X′** respectively; $\alpha_{n,m}(t)$ is the fading gain from the $N_t$ -th transmit to the $N_r$-th receive antenna during the $t$-th channel use. In the second equation, the assumption used is that in the block fading case $\alpha_{n,m}(t)$ is held fixed during the transmission period of a single IST block, i.e. during the $T_D$ channel uses. That is, $\alpha_{n,m}(t) = \alpha_{n,m}^{d}$ for $t = (d\text{-}1)T_d + 1, (d\text{-}1)T_d + 2, \ldots, dT_d$. $\mathbf{S}_{d,m,t}$ and $\mathbf{S}'_{d,m,t}$ represent the $m$-th row and $t$-th column element of the $d$-th block $\mathbf{S}_d$ and $\mathbf{S}'_d$ in **X** and **X′** respectively.

Denoting $\Phi_{d,n} := \left(\alpha_{n,1}^{d}, \alpha_{n,2}^{d}, \cdots, \alpha_{n,N_t}^{d}\right)$, we can rewrite (14) in a matrix form,

$$d^2(\mathbf{X},\mathbf{X}') = \sum_{d=1}^{D}\sum_{n=1}^{N_r}\Phi_{d,n}\mathbf{A}_d\Phi_{d,n}^{H} \tag{15}$$

where $(\bullet)^H$ denotes the conjugate transpose, $\mathbf{A}_d$ is a $[N_t \times N_t]$ matrix in which the $p^{\text{th}}$-row and $q^{\text{th}}$-column entry is obtained by $\mathbf{A}_d = (\mathbf{S}_d - \mathbf{S}'_d)(\mathbf{S}_d - \mathbf{S}'_d)^H$.

Consider the eigenvalue decomposition $\mathbf{A}_d = \mathbf{U}_d\Lambda_d\mathbf{U}_d^H$, where $\mathbf{U}_d$ is the associated unitary matrix and $\Lambda_d$ is a diagonal matrix whose diagonal terms are the eigenvalues of

$\mathbf{A}_d$. Let's have the eigenvalues be denoted as, $\lambda_m^d$ , for $m$ = 1,2 , …, $N_t$. Denoting $\Phi_{d,n}\mathbf{U}_d =: \left(\gamma_{m,n}^d, \gamma_{m,n}^d, \cdots, \gamma_{m,n}^d\right)$, we have

$$
\begin{aligned}
d^2(\mathbf{X},\mathbf{X}') &= \sum_{d=1}^{D}\sum_{n=1}^{N_r}(\Phi_{d,n}\mathbf{U}_d)\Lambda_d(\Phi_{d,n}\mathbf{U}_d)^H \\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_r}\sum_{m=1}^{N_t}\lambda_m^d\left|\gamma_{m,n}^d\right|^2.
\end{aligned} \tag{16}
$$

Since $\mathbf{U}_d$ is a unitary matrix and $\Phi_{d,n}$'s are independent matrices with independent entries $\alpha_{m,n}^d$, $\gamma_{m,n}^d$'s have the same probability distributions as those of $\alpha_{m,n}^d$'s given in Eq. (2) and they are independent for different $m$, $n$ and $d$.

Substituting (16) into (13) and averaging it over the channel $\mathbf{H}$ (i.e. $\gamma_{m,n}^d$'s), we have the pairwise error probability,

$$
\begin{aligned}
P^{BL}(\mathbf{X}\to\mathbf{X}') &\le \prod_{d=1}^{D}\prod_{n=1}^{N_r}\prod_{m=1}^{N_t} M_R\left(-\frac{\rho_s}{4}\lambda_m^d\right) \\
&= \prod_{d=1}^{D}\underbrace{\prod_{n=1}^{N_r}\prod_{m=1}^{N_t}\frac{1}{1+\frac{\rho_s}{4}\lambda_m^d}\exp\left(-\frac{F_r\frac{\rho_s}{4}\lambda_m^d}{1+\frac{\rho_s}{4}\lambda_m^d}\right)}_{P^{BL}(\mathbf{S}_d\to\mathbf{S}'_d)} \\
&=: \prod_{d=1}^{D} P^{BL}(\mathbf{S}_d\to\mathbf{S}'_d).
\end{aligned} \tag{17}
$$

The first step is based on the MGF-based approach in [xx]. Note that $P^{BL}(\mathbf{S}_d\to\mathbf{S}'_d)$ is defined for the pairwise error probability between IST codewords $\mathbf{S}_d$ and $\mathbf{S}'_d$. Note that we use the superscript $BL$ to denote the block fading case.

## *B. The Worst Test IST Symbol and Further Upper Bound*

Our union bound is an average over an ensemble of codebooks. All codes in the ensemble are linear and thus any code contains the all-zero codeword. While the outer code may be linear, however, the space-time code is not. Thus, in general a union bound for the space-time code should be evaluated in an average sense such that each space-time codeword should be used once as the test codeword. Since such an option is too costly—there are $2^K$ different codewords, we circumvent this problem by resorting to a further upper bound which is based upon the transmission of a selected test space-time codeword composed of the worst IST symbols.

An upper bound to the IST word error probability is given by

$$P_w^{IST} \le E_{\mathbf{s}_{(i)}}\left[\sum_{j\ne i} P(\mathbf{s}_{(i)} \to \mathbf{s}_{(j)})\right]$$
$$\le \max_{\mathbf{s}_{(i)}} \sum_{j\ne i} P(\mathbf{s}_{(i)} \to \mathbf{s}_{(j)}) = \max_{\mathbf{s}_{(i)}} \sum_{j=1}^{J} P(\mathbf{s}_{(i)} \to \mathbf{s}_{(j)}) - 1 \tag{18}$$

where the first inequality is merely due to the union bound; $E_{\mathbf{s}_{(i)}}[\cdot]$ is the expectation over the equi-probable selection of one ISI codeword $\mathbf{s}_{(i)}$ as the test word, and the last step follows from $P(\mathbf{s}_{(i)} \to \mathbf{s}_{(i)}) = 1$. Let us denote the argument of the maximum operation in the last equation as $\mathbf{s}_{(*)}$, i.e.,

$$\mathbf{s}_{(*)} = \arg\max_{\mathbf{s}_{(i)}} \sum_{j=1}^{J} P^{BL}\left(\mathbf{s}_{(i)} \to \mathbf{s}_{(j)}\right). \tag{19}$$

Hence, $\mathbf{s}_{(*)}$ is the IST codeword that contributes to the worst summation of IST pairwise error probabilities.

Now, we move on to the union bound for the ST code. Making use of a specific space-time word $\mathbf{X}^*$, we can have an upper bound to the word error probability as

$$\bar{P}_w \le E_{\mathbf{X}}\left[\sum_{\mathbf{X}'\ne\mathbf{X}} P^{BL}(\mathbf{X} \to \mathbf{X}')\right] \le \sum_{\mathbf{X}\ne\mathbf{X}^*} P^{BL}(\mathbf{X}^* \to \mathbf{X}), \tag{20}$$

where the first inequality is from the union bound argument and $E_X[\bullet]$ is the expectation over the equi-probable selection of transmitted codeword $\mathbf{X}$. The second inequality holds as long as $\mathbf{X}^*$ is selected to be one of the codewords which make the sum of the pairwise error probabilities be larger than the average. It is reasonable for us to select $\mathbf{X}^* = [\mathbf{s}_{(*)}\ \mathbf{s}_{(*)}\ \dots\ \mathbf{s}_{(*)}]$ and use this for the union bound throughout this chapter.

## *C. The Pairwise Distance Profile*

Due to a large population size of the outer block code, a straightforward evaluation of (20) should be avoided. For this, we want to enumerate all pairwise error events and count the multiplicities of each and every event which leads to the same pairwise error probability. We then use a polynomial representation of the union bound and come up with the final expression for the union bound.

First, let us count the number $d_j$ of IST pairs among the total of $D$ IST pairs which is characterized by a transition from the IST symbol $\mathbf{s}_{(*)}$ to the $j$-th IST symbol $\mathbf{s}_{(j)}$. A collection of these numbers in an array, $\boldsymbol{d} = [d_1\ d_2\ \dots\ d_J]$, will be named as the Pairwise Distance Profile (PDP). It keeps the record of the number of occurrences of all possible

IST pairs, i.e. ($\mathbf{s}_{(*)}$, $\mathbf{s}_{(j)}$) for $j$=0,1,2, …, $J$-1, in a pair of space-time words ($\mathbf{X}^*$, $\mathbf{X}$). Recall that $J = 2^{N_b}$ is the size of the IST constellation.

Suppose $\mathbf{X}^*$ is mapped from the all-zero codeword. Let us denote $\chi_{\mathbf{d}(h)}$ as the set of all ST codewords $\mathbf{X}$ having the same distance profile $\mathbf{d}(h)$ from $\mathbf{X}^*$. The distance spectrum of the ST code thus can be defined as the set $\{A_{\mathbf{d}(h)}\}$ of cardinalities, i.e., $A_{\mathbf{d}(h)}:=|\chi_{\mathbf{d}(h)}|$. Denote the pre-image of $\chi_{\mathbf{d}(h)}$ in the outer code as $\mathcal{C}_{\mathbf{d}(h)}$; they have the same cardinality. $\mathcal{C}_{\mathbf{d}(h)}$ is actually the set of all binary outer codewords with the same metric $\mathbf{d}(h)$. The cardinality of $\mathcal{C}_{\mathbf{d}(h)}$ (averaged over the code ensemble) is therefore given by (12), so is that of $\chi_{\mathbf{d}(h)}$.

Let's use $\mathbf{X}_{\mathbf{d}(h)}$ to denote any erroneous word $\mathbf{X}$ which renders a particular PDP $\mathbf{d}(h)$ from $\mathbf{X}^*$.

Therefore, we can write the pairwise error probability between $\mathbf{X}^*$ and any other EST codewords $\mathbf{X}_{\mathbf{d}(h)}$ of distance metric $\mathbf{d}$ in the following way

$$
\begin{aligned}
P^{BL}\left(\mathbf{X}^* \to \mathbf{X}_{\mathbf{d}(h)}\right) &\le \prod_{d=1}^{D} P^{BL}\left(\mathbf{s}_{(*)} \to \mathbf{s}_d\right) \\
&= \prod_{j=1}^{J} P^{BL}(\mathbf{s}_{(*)} \to \mathbf{s}_{(j)})^{d_j} := \prod_{j=1}^{J} \left(\beta_j^{BL}\right)^{d_j}
\end{aligned}
\tag{21}
$$

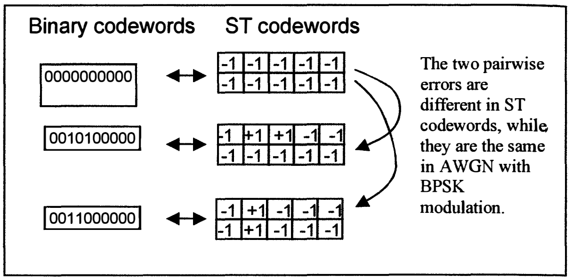
where in the second step we group like terms under power exponent $d_j$. Note using (17) we denote

$$
\begin{aligned}
\beta_j^{BL} &:= P^{BL}(\mathbf{s}_{(*)} \to \mathbf{s}_{(j)}) \\
&= \prod_{n=1}^{N_r} \prod_{m=1}^{N_t} \frac{1}{1+\frac{\rho_s}{4}\lambda_m^{(j)}} \exp\left( -\frac{F_r \frac{\rho_s}{4}\lambda_m^{(j)}}{1+\frac{\rho_s}{4}\lambda_m^{(j)}} \right)
\end{aligned}
\tag{22}
$$

where $\lambda_m^{(j)}$, $m$=1, 2, …, $N_t$, are the eigenvalues of $(\mathbf{s}_{(*)} - \mathbf{s}_{(j)})(\mathbf{s}_{(*)} - \mathbf{s}_{(j)})^H$.

**Example-1:** The pairwise error probability, conditioned up on a channel realization, between any two space-time codewords depends in general on the Euclidean distance of the two space-time cdewords at the receive signal space. Having the pairwise error probabilities averaged over a block fading channel, they depend on the so-called *column-distance profile* of the doifference matrix of the two space-time codewords. This will be illustrated in detail in this example and the notion of the distance spectrum for the concatenated coding scheme will be established.

Consider a two-transmit antenna case and assume independent fading in each channel use. Thus, the block length of inner space-time block code is 1. Assume the binary block code is of length 10. Each codeword of length 10 is transformed into a 2 x 5 space-time matrix via the binary phase shift keying (BPSK) modulation. For the transmission of an entire binary codeword, therefore, the independent MIMO fading channel should be used five times.




Suppose the one-to-one correspondent relationship between the binary codeword and the space-time codeword shown in the figure. The all-zero binary codeword is mapped to the all -1 space-time codeword matrix. Also indicated in the figure are the two space-time codewords for the two weight-2 binary codewords. Assume that the all-zero codeword is sent at the transmitter, and consider the two pairwise error events. Each event leads to a different pairwise error probability in general, while exact difference dependent upon the exact channel realization, because the Euclidean distances at the receive signal space for the two cases are different. Averaging them over the ensemble of the fading channel [30, 48, 52, 54], unconditional pairwise error probabilities are obtained. **From inspection of the pairwise error probabilities, we note that the crucial information comes from the columns of the pairwise difference matrix** $\mathbf{X}(c_h)-\mathbf{X}(c_0)$**.**

Now let us consider the five columns of the difference matrix for the two pairwise error events. We note that there are only three kinds of columns. The first is an all-zero column vector. The second is that there is a single difference in a column such that it is either $(+2, 0)^t$ or $(0, +2)^t$. The third kind is that there are two differences in a column; it is $(+2, +2)^t$. Thus, there are three kinds of per-column Euclidean distances, 0 (no difference), $2^2$ (a single difference), and $2^2 + 2^2$ (two differences).

Among all weight-2 error events, the first pairwise error shown in the figure represents all such error events that both differences occur in the same column. The second pairwise error represents all such error events that both differences occurred in a single column.

In order to systematize the approach we introduce a column count variable $q_i$ for each weight $i = 0, 1, \ldots, N_t$ (the number $N_t$ of transmit antennas is 2 in this example). The variable $q_i$ stores the count-up of the number of columns in the difference matrix for each weight $i$. We can store all column count variables into a column count vector $\mathbf{q}$ and call it *column count vector*. For example, the *column count vector* for weight-2 codewords can be written as $\mathbf{q}(2) := [q_0\ q_1\ q_2]$. The number in the parenthesis is used to denote the Hamming weight. With this convention, the *column counter vector* for the first error event is [3 2 0]. It is [4 0 1] for the second case. Note that the sum has to be 5, the total number of columns. Collectively, we can call the set of vectors as *column count profile* for weight 2. Then, the pairwise errors due to Hamming weight 2 codewords can be written as

$$\underbrace{A_2 \binom{10}{2}^{-1} 4\binom{5}{[3\,2\,0]}}_{\text{Distance Profile for the column count [3 2 0]}} \underbrace{\left(\frac{1}{1+\rho_s}\right)^{N_r}}_{\text{pairwise error due to such a column count}} + \underbrace{A_2 \binom{10}{2}^{-1} \binom{5}{[4\,0\,1]}}_{\text{Distance Profile for the column count [4 0 1]}} \left(\frac{1}{1+2\rho_s}\right)^{N_r},$$

where we note that $\binom{10}{2} = 4\binom{5}{[3\,2\,0]} + \binom{5}{[4\,0\,1]}$ (they are the binomial and multinomial coefficients) and thus all combinations for weight 2 are accounted for. The interpretation of this goes as follows: The distance profiles for each column count vector can be obtained by ensemble averaging over all codes. There are (10 choose 2) words in the set of binary words of length 10. Two positional differences can be selected randomly out of the ten positions. Then, what is the probability that both differences end up in any column? It's $\binom{10}{2}^{-1} 4\binom{5}{[3\ 2\ 0]}$ where the multinomial coefficient $\binom{5}{[3\ 2\ 0]}$ is computed as $\frac{5!}{3!2!0!}$. In addition, we note that all unconditional pairwise error probabilities belong to the same column count vector is exactly the same.

We now conclude our set-partitioning example for weight-2. This example can be extended to any other weight, and finally union bound is obtained by collecting all the terms with weights.

**End of Example-1**

**Example-2 (Set partitioning for weight 4 and beyond)**

Again, let's assume there are five columns and $N_t$=$N_r$=2. Find the set-partitioning column profiles for weight 4, 6, 8, and so on. Describe a pseudo algorithm to find them systematically.

**End of Example-2**

## *D. Union Bound in a Numerically Efficient Form*

Now the main result of this chapter is derived.

Making use of the result (21) and (17) into (20), the union bound can be written as,

$$\begin{aligned} P_w &\le \sum_{\mathbf{X}\neq\mathbf{X}^*} P^{BL}(\mathbf{X}^* \to \mathbf{X}), \\ &= \sum_h \sum_{\mathbf{d}(h)\in\Omega_h} A_{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^{BL}\right)^{d_j}. \end{aligned} \tag{24}$$

Then, making use of the combinatorial result (12), the R.H.S. of (24) can be rewritten

$$\begin{aligned} P_w &\le \sum_h A_h \binom{N}{h}^{-1} \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^{BL}\right)^{d_j} \\ &= \sum_h A_h \binom{N}{h}^{-1} \phi(h), \end{aligned} \tag{25}$$

where the second line is simply from the definition of $\phi(h)$, i.e.

$$\phi(h) := \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^{BL}\right)^{d_j}. \tag{26}$$

Recalling that $\Omega_h$ is the collection of all pairwise distance profiles with the Hamming weight $h$, the coefficient $\phi(h)$ can be efficiently calculated by making use of the following power expansion. That is, we utilize a dummy variable $z$ in the following manner

$$\begin{aligned} \left(\sum_{j=1}^{J} \beta_j^{BL} z^{w_j}\right)^D &= \sum_{h=0}^{N} \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \left(\beta_j^{BL}\right)^{d_j} z^h \\ &= \sum_{h=0}^{N} \phi_h z^h. \end{aligned} \tag{27}$$

Using (27), one can find a way to readily calculate the expansion coefficient $\phi(h)$. We suggest the following approach: First, evaluate the polynomial coefficients inside the parenthesis of the L.H.S. of (27). Second, arrange the coefficients into a sequence of increasing power index on the utility variable $z$. The length of the sequence is $N_b+1$ such that there are $z^0$, $z^1$, …, and $z^{N_b}$ because $w_j$ is the Hamming weight of a binary string of length $N_b$. Third, take the $D$-fold convolution of the ordered sequence of length $N_b+1$. The outcome of the $D$-fold convolution is the coefficient sequence $\langle \phi(h), h=0,1,2,\cdots,N\rangle$ of length $DN_b+1=N+1$.

Now we can summarize our main result in the following theorem.

*Theorem-II: For the transmission of a linear* ($N$, $K$) *block code with a given distance spectrum* $\langle A_h\rangle$ *over the block Ricean fading* ($N_t$, $N_r$) *MIMO channel with the Ricean factor* $F_r$, *the union upper bound on the probability of making codeword errors is given by*

$$P_w \le \sum_{h=1}^{N} \binom{N}{h}^{-1} A_h \phi(h) \tag{28}$$

*where* $\phi(h) := \sum_{\mathbf{d}\in\Omega_h} \binom{D}{\mathbf{d}} \prod_{j=1}^{J} \left(\beta_j^{BL}\right)^{d_j}$ *and* $\beta_j^{BL}$ *is given by*

$$\beta_j^{BL} = \prod_{n=1}^{N_r}\prod_{m=1}^{N_t} \frac{1}{1+\frac{\rho_s}{4}\lambda_m^{(j)}} \exp\left(-\frac{F_r \frac{\rho_s}{4}\lambda_m^{(j)}}{1+\frac{\rho_s}{4}\lambda_m^{(j)}}\right). \tag{29}$$

$F_r$ *is the Ricean factor and* $\lambda_m^{(j)}$ , *for* $j$=1,2,…,$J$, *are the eigenvalues of the matrix* $\mathbf{A}_{(j)} = \left(\mathbf{s}_{(*)} - \mathbf{s}_{(j)}\right)\left(\mathbf{s}_{(*)} - \mathbf{s}_{(j)}\right)^H$.

The bound on the codeword error (28) can be extended to the bound on the bit error probability by replacing $A_h$ with $A_h'$,

$$A_h' = \sum_{\omega=1}^{K} \frac{\omega}{K} A_{\omega,h} \tag{30}$$

where $A_{\omega,h}$ is the number of the codewords with input weight $\omega$ and output weight $h$. For the ensembles interested in this chapter, (30) can be simplified as (see Appendix for proof)

$$A_h' = \frac{h}{L} A_h . \tag{31}$$

# V. UNION BOUNDS ON INDEPENDENT, QUASI-STATIC FADING, AND RANDOM CODING CASES

In this section, we extend the main result to independent and quasi-static fading channels, as well as for the random coding case.

## *A. Fast (Independent) Fading Case*

We now focus on the concatenated code over the fast (independent) fading case. The channel model given in (1) is modified a little bit to accommodate the independent fading even within the transmission of an IST symbol-matrix, $t$ = 1, 2, ..., $T_D$. Stacking the column of the matrix $\mathbf{R}_d$, and similarly for those of others, $\mathbf{S}_d$ and $\mathbf{Z}_d$, into a single column, we first obtain $\mathbf{r}_d$, $\mathbf{s}_d$ and $\mathbf{z}_d$ with dimension $[N_r T_D \text{ x } 1]$, $[N_t T_D \text{ x } 1]$ and $[N_r T_D \text{ x } 1]$ respectively. Then, we can transform the input-output relationship in to the following equation:

$$\mathbf{r}_d = \sqrt{\rho_s}\tilde{\mathbf{H}}_d \mathbf{s}_d + \mathbf{z}_d, \tag{32}$$

where we define the $[N_r T_D \text{ x } N_r T_D]$ matrix $\tilde{\mathbf{H}}_d$ as a block diagonal matrix, i.e.,

$$\tilde{\mathbf{H}}_d := \begin{pmatrix} \mathbf{H}_{d,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{d,2} & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \\ \mathbf{0} & & \cdots & \mathbf{H}_{d,T_D} \end{pmatrix}. \tag{33}$$

Each matrix $\mathbf{H}_{d,t}$ along the diagonal is the channel at the $t$-th time epoch, for $t$=1, 2, ..., $T_D$; they are mutually independent; $\mathbf{0}$ denotes the $[N_r \text{ x } T_D]$ matrix with the elements of zeros. This input-output relationship can be used for the derivation of the union bound for independent fading case, as well as in the description of MIMO detection and decoding operation given in section VI.B.

Each column of an IST symbol-matrix undergoes an independent MIMO fading channel. Thus, we note that any $J$-ary $[N_t \text{ x } T_D]$ symbol-matrix is composed of $T_D$ columns, i.e.

$$\mathbf{s}_{(j)} = [\mathbf{s}_{(j),1}\ \mathbf{s}_{(j),2} \cdots \mathbf{s}_{(j),T_D}], \tag{34}$$

for $j$=1,2,..., $J$.

We first consider a single transmission of IST symbol-matrix. The pairwise error probability from the symbol-matrix $\mathbf{s}_{(*)}$ to $\mathbf{s}_{(j)}$ is given by,

$$\underbrace{P^F(\mathbf{s}_{(*)} \to \mathbf{s}_{(j)}) \le \prod_{t=1}^{T_D}\left(1+\frac{\rho_s}{4}\left|\mathbf{s}_{(*),t}-\mathbf{s}_{(j),t}\right|^2\right)^{-N_r} \exp\left(-N\frac{F_r\frac{\rho_s}{4}\left|\mathbf{s}_{(*),t}-\mathbf{s}_{(j),t}\right|^2}{1+\frac{\rho_s}{4}\left|\mathbf{s}_{(*),t}-\mathbf{s}_{(j),t}\right|^2}\right)}_{=:\beta_j^F}, \tag{35}$$

where $\mathbf{s}_{(*)}$ is obtained from (19). As usual, we assume each of the *J-ary* IST symbols, $\mathbf{s}_{(1)}$, $\mathbf{s}_{(2)}$, …, $\mathbf{s}_{(J)}$, is equally likely selectable for transmission. Then, the pairwise error probability between any two ST codewords $\mathbf{X}^* = [\mathbf{s}_{(*)}\,\mathbf{s}_{(*)} \ldots \mathbf{s}_{(*)}]$ and $\mathbf{X} = [\mathbf{S}_1\,\mathbf{S}_2 \ldots \mathbf{S}_D]$ is given by,

$$P^F(\mathbf{X}^* \to \mathbf{X}) \le \prod_{d=1}^{D} P^F(\mathbf{s}_{(*)} \to \mathbf{S}_d)\,. \tag{36}$$

As was done in previous section, we can decompose the R.H.S. of (36) with respect to the pairwise distance profile $\mathbf{d}$. Then, the pairwise error probability between $\mathbf{X}^*$ and any word $\mathbf{X}_{\mathbf{d}(h)}$ with PDP $\mathbf{d}$ is,

$$\sum_{X_{\mathbf{d}(h)} \ne X^*} P(\mathbf{X}^* \to \mathbf{X}_{\mathbf{d}(h)}) \le \prod_{j=1}^{J} P(\mathbf{s}_{(*)} \to \mathbf{s}_{(j)})^{d_j} = \prod_{j=1}^{J} \beta_j^{Fd_j} \tag{37}$$

by grouping the like terms in (36) under power exponent $d_j$.

Now the fast fading result can be summarized in the following corollary:

*Corollary-1: For the transmission of a linear* $(N, K)$ *block code with distance spectrum* $A_h$ *over the fast (independent) Ricean fading* $(N_t, N_r)$ *MIMO channel with the Ricean factor* $F_r$, *the union upper bound on the probability of making codeword errors is given by*

$$P_w \le \sum_{h=1}^{N} \binom{N}{h}^{-1} A_h \phi_h \tag{38}$$

*where each* $\phi_h$ *is the coefficient of the polynomial expansion,*

$\left(\sum_{j=1}^{J} \beta_j^F z^{w_j}\right)^D = \sum_{h=0}^{L} \phi_h z^h$. *That is,* $\phi_h := \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} \prod_{j=1}^{J} \beta_j^{F\delta_j}$.

## *B. The Quasi-Static Fading case*

In the case of quasi-static Ricean fading, the channel model in (1) can be used with a modification that the channel matrix $\mathbf{H}_1$ is held fixed throughout the entire transmission period of a space-time codeword, i.e., $\mathbf{H}_1 = \mathbf{H}_2 = \ldots = \mathbf{H}_D$. For this channel, the pairwise error probability between $\mathbf{X}^* = [\mathbf{s}_{(*)}\, \mathbf{s}_{(*)} \;\ldots\; \mathbf{s}_{(*)}]$ and other ST codeword can be obtained as (see [5] for details),

$$P^{SL}\left(\mathbf{X}^* \to \mathbf{X}_{\mathbf{d}(h)}\right) \le \prod_{n=1}^{N_r}\prod_{m=1}^{N_t} \frac{1}{1+\frac{\rho_s}{4}\lambda_m} \exp\left( -\frac{F_r \frac{\rho_s}{4}\lambda_m}{1+\frac{\rho_s}{4}\lambda_m} \right) \tag{39}$$

where $\lambda_m$'s are the eigenvalues of $\mathbf{A} = \mathbf{B}\mathbf{B}^H$ where $\mathbf{B}$ is the difference matrix, i.e., $\mathbf{B} := \mathbf{X}^* - \mathbf{X}_{\mathbf{d}(h)}$. The R.H.S. of (39) satisfies the following property.

*Lemma-1: The matrix* **A** *are determined completely by the pairwise distance profile* **d** of $X_{\mathbf{d}(h)}$.

Proof: It is sufficient to prove that $\mathbf{A}_1 = (\mathbf{X}^* - \mathbf{X}_1)(\mathbf{X}^* - \mathbf{X}_1)^H$ is equal to $\mathbf{A}_2 = (\mathbf{X}^* - \mathbf{X}_2)(\mathbf{X}^* - \mathbf{X}_2)^H$ where $\mathbf{X}_1$ and $\mathbf{X}_2$ are any two matrices which produce the same PDP **d** when compared with $\mathbf{X}^*$. In such a case, $\mathbf{X}_1$ and $\mathbf{X}_2$ can be regarded as a block permutation of each other. By the block permutation, we mean the permutation of the IST blocks within a ST codeword. Analogously, the difference matrix $(\mathbf{X}^* - \mathbf{X}_1)$ is a block permutation of $(\mathbf{X}^* - \mathbf{X}_2)$. Therefore, $\mathbf{A}_1 = \mathbf{A}_2$. END

As an immediate result of Lemma-1, we note that the sets of eigenvalues of $\mathbf{A}_1$ and $\mathbf{A}_2$ are exactly the same, and thus the pairwise error probabilities with respect to $\mathbf{A}_1$ and $\mathbf{A}_2$ are exactly the same. From this, we note that the pairwise error probability is determined completely by the *pairwise distance profile* **d** of $X_{\mathbf{d}(h)}$. This implies that for each distinct PDF $\mathbf{d}(h)$ there is a distinct set of eigenvalues of matrix $\left(\mathbf{X}^* - \mathbf{X}_{\mathbf{d}(h)}\right)\left(\mathbf{X}^* - \mathbf{X}_{\mathbf{d}(h)}\right)^H$.

Similarly to the block fading case, therefore, we can write the upper bound in the following way:

*Corollary-2: For the transmission of a linear* $(N, K)$ *block code with distance spectrum* $A_h$ *over the quasi-static Ricean fading* $(N_t, N_r)$ *MIMO channel with the Ricean factor* $F_r$, *the union upper bound on the probability of making the codeword error is given by,*

$$P_w^{SL} \le \sum_{h=1}^{N} A_h \binom{N}{h}^{-1} \sum_{\mathbf{d}(h)\in\Omega_h} \binom{D}{\mathbf{d}(h)} P^{SL}(\mathbf{X}^* \to \mathbf{X}_{\mathbf{d}(h)}) \tag{40}$$

*and an upper bound on the ST bit error probability is given by replacing $A_h$ in with $A'_h$* in (31).

Unlike fast and block fading cases, the idea of using polynomial expansion cannot be applied here and thus (40) cannot be simplified any further. The straightforward and direct way is to enumerate all elements in the set $\Omega_h$'s and evaluate the bound given in (40). This approach is not in desirable form because of high computational complexity. Thus, it remains to be seen as a future task to find an efficient way to proceed with the current expression (40).

In addition, the union bounds for quasi-static fading case may produce a loose bound in its current form, especially for a small number of receiver antennas. The pairwise error probability for a particular PDP $\mathbf{d}(h)$ would be dominated by the term of the largest eigenvalue, such as

$$\prod_{n=1}^{N_r}\prod_{m=1}^{N_t} \frac{1}{1+\frac{\rho_s}{4}\lambda_m} \exp\left( -\frac{F_r \frac{\rho_s}{4}\lambda_m}{1+\frac{\rho_s}{4}\lambda_m} \right) \approx \prod_{n=1}^{N_r} \frac{1}{1+\frac{\rho_s}{4}\lambda_*} \exp\left( -\frac{F_r \frac{\rho_s}{4}\lambda_*}{1+\frac{\rho_s}{4}\lambda_*} \right), \tag{41}$$

where $\lambda_* := \max\{\lambda_m, m = 1,2,\dots,N_t\}$. For Rayleigh fading cases, it can be approximated as $(1+\frac{\rho_s}{4}\lambda_*)^{-N_r}$; note that the order of the power-law decreasing PEP is only $N_r$. This rate of decrease may not be fast enough to suppress the influence of the spectral component $A_h$ in the product. Thus, another interesting research issue remained as a future task is to come up with a tight bounding technique for the qausi-static fading channels. One feasible idea is to apply the Fano-Gallager's tight bounding technique to the instantaneous signal-to-noise ratio, as explored in [xxi].

## *C. Error Exponent for Random Code*

The error exponent can be obtained by assuming random code. The advantage is that a meaningful bound can be obtained quickly (in closed forms) and used to ascertain the performance even without the knowledge of the distance spectrum which requires a numerical evaluation of its own or sometimes is not even available.

The random block code is not linear in general since they are selected randomly out of the binary $N$-tuple space. Therefore, we need to provide a little bit of adjustment to the lines of analysis given for the linear code case. Namely, the analysis has been based on the assumption that $\mathbf{X}^*$ is mapped from the all-zero codeword that exists in each and every code in an ensemble. For random codes, however, we cannot find such a codeword. The all-zero codeword does not exist in every code in an ensemble. This problem can be dealt with the following procedure.

Consider the ensemble of random codes defined by the fixed block length $N$ and code rate $R_c$. For any code $C$ in the ensemble, we randomly select any one of its codewords as the test codeword. Define $A_h^C$ as the number of codewords in the code that have the same Hamming distance $h$ from the test codeword. The average of $A_h^C$ over the ensemble is denoted by $A_h$ and can be obtained as

$$A_h = \begin{cases} \tilde{A}_h & h > 0 \\ \tilde{A}_h + 1 & h = 0 \end{cases} \tag{42}$$

where

$$\tilde{A}_h := \frac{2^{NR_c} - 1}{2^N} \binom{N}{h}.$$

It can be shown that (42) is obtained by considering the facts that there are in total $\binom{N}{h}$ distinct binary sequences of distance $h$ from the test codeword, and that the odd for each of these sequences (other than the test codeword) to be included in a particular codebook is $2^{NR_c} - 1$ out of $2^N$. Since the test codeword is actually a codeword with distance zero from itself, we add the additional one to $A_0$. It should be noticed that (42) holds without regard to the specific selection of a test codeword from each codebook.

Note that this definition of distance spectrum is more precise than what was used in previous sections where the weights (the Hamming distances from the all-zero sequence) of codewords determines the spectrum. Under such a definition, the distance spectrum would have been given as

$$A_h = \frac{2^{NR_c}}{2^N} \binom{N}{h}. \tag{43}$$

Assume $\mathbf{X}^*$ is mapped from the corresponding $RC$ in each code. We therefore have the following corollary of Theorem II:

*Corollary-3: For the transmission of a random (N, K) block code over the block Ricean fading ($N_t$, $N_r$) MIMO channel with the Ricean factor $F_r$, the union upper bound on the probability of making the block codeword error is given by*

$$P_w \le 2^{-N\cdot E(R_c,\rho_b)} \tag{44}$$

*where the error exponent* $E(R,\rho_b)$ *is*

$$E(R_c,\rho_b) := 1 - R_c - \frac{\log_2(\sum_{j=1}^{J}\beta_j^{BL})}{N_t N_c T}. \tag{45}$$

Proof: Based on the distance spectrum in (42) and the upper bound on the block fading channel (25), an upper bound is obtained as follows

$$\begin{aligned}
P_w &\le \sum_{X\neq X^*} P(X^* \to X) \\
&= \sum_{h=0}^{N}\binom{N}{h}^{-1} \tilde{A}_h \sum_{\mathbf{d}(h)\in\Omega_h}\binom{D}{\mathbf{d}(h)}\prod_{j=1}^{J}\beta_j^{BL\delta_j} \\
&\le 2^{-N(1-R_c)}\sum_{h=0}^{N}\sum_{\mathbf{d}(h)\in\Omega_h}\binom{D}{\mathbf{d}(h)}\prod_{j=1}^{J}\beta_j^{BL\delta_j} \\
&= 2^{-N(1-R_c)}\left[\sum_{j=1}^{J}\beta_j^{BL}\right]^D \\
&= 2^{-N[1-R_c]}2^{D\log_2\left(\sum_{j=1}^{J}\beta_j^{BL}\right)}
\end{aligned} \tag{46}$$

END

We may define $\rho_b^*$ as the cut-off SNR such that $\rho_b^* = \inf_{\rho_b} E(R_c,\rho_b) > 0$. With minor manipulations on the system parameters the bound and exponent can be also written as

$$P_w \le 2^{-D\cdot E(R,\rho_b)} \tag{47}$$

where the error exponent $E(R,\rho_b)$ is

$$E(R,\rho_b) := N_b - \underbrace{N_b R_c}_{=:R_{IST}} - \log_2(\textstyle\sum_{j=1}^{J}\beta_j^{BL}).$$

Note that $R_{IST} := N_b R_c$ denotes the transmission rate per IST symbol which is taken over the duration of $T_D$ channel uses. Thus, the transmission rate is obtained as $R = \frac{R_{IST}}{T_D}$ bit-per-channel use.

# VI. THE LDPC CODE AND ITERATIVE DETECTION/DECODING ALGORITHM

We now discuss the LDPC code, encoding and decoding operation, and the super-iterative MIMO detection and decoding. The encoding and decoding operations are mainly from the Gallager's thesis [xvii]. The super-iterative MIMO detection and decoding algorithm straightforwardly follows the maximum aposteriori (MAP) log-likelihood radio (LLR) generation methods. Thus, the algorithm is standard but we will briefly discuss them for completeness.

## *A. The ($N$, $d_b$, $d_c$) Gallager code*

In this chapter, we assume the ($N$, $d_b$, $d_c$) Gallager codes. There are $d_b$ 1s in each column and $d_c$ 1s in each row in the parity check matrix of a code. The parity-check matrix is generated randomly by using the Gallager's approach [xxii]. Once a parity check matrix is obtained, it is systematized by going through the Gaussian elimination process. From the systematic parity check matrix, the systematic generator matrix is obtained. Making use of the systematic format, the decoding operation is simplified.

## *B. The super-iteration algorithm*

The super-iterative detection and decoding algorithm is based on the MAP LLR generation algorithm. For the description of the iterative algorithm, we introduce an index which describes the relevant position of the bits in the codeword vector **c**. Referring to Figure 1, the bits in a codeword are categorized into $D$ consecutive groups (each group with $N_b$ bits), i.e.,

$$\mathbf{c} = [c_{1,1}\ c_{1,2} \cdots c_{1,N_b}; c_{2,1}\ c_{2,2} \cdots c_{2,N_b}; \cdots ; c_{D,1}\ c_{D,2} \cdots c_{D,N_b}].$$

Focusing on one of the IST blocks, without loss of generality, we will describe the steps to generate the *extrinsic* LLR on bits. Without loss of generality, let's focus on the first IST block, and bits $c_{1,1}\ c_{1,2} \cdots c_{1,N_b}$. For simplicity, we will omit the block index and refer to them simply as $c_1\ c_2 \cdots c_{N_b}$.

Making use of the general linear dispersion code and the corresponding procedure given in [ix], one can transform the input-output relationship of the IST symbol-matrices (1) into the following linear systems of equation with the real-valued [$2Q$ x 1] input vector **s** and [$2N_rT$ x 1] output vector **r**, i.e.

$$\mathbf{r} = \sqrt{\rho_s}\mathcal{H}\mathbf{s} + \mathbf{w}, \tag{48}$$

where $\mathcal{H}$ is the equivalent [$2N_r\, T_D$ x $2Q$] real-valued channel matrix which subsumes the effect of the original channel matrix **H** and the set of $Q$ linear dispersion matrices $\{\mathbf{A}_q, \mathbf{B}_q\}_{q=1,\dots,Q}$; and **w** denotes the noise vector each element of which is independent identically distributed real-valued additive-white-Gaussian noise with zero mean and variance 0.5. This model will be useful for the block fading channel model. For fast (independent) fading model, one can use the input-output relationship given in (32).

With the input-output model (48), explanation on the routines of the MIMO log-likelihood ratio (LLR) generation and the super-iteration will be given (while it is analogous to use the input-output relationship given in equation (32) for the independent fading case). Note that one can simply use the input-output relationship (1) as well; but using (48) might have an advantage for implementation point of view since all values in (48) are real valued.

First, note that each symbol-vector **s** carries $N_b$ coded bits. Thus, upon receiving an IST symbol, the LLR Generator should produce the *extrinsic* LLRs on $N_b$ coded bits. Following the standard procedure, one can generate a maximum likelihood ratio on a particular bit $c_k$ in the following manner:

$$\begin{aligned} L_{pos}(c_k) &:= \log\frac{\Pr\{c_k = 1 \mid \mathbf{r}, \mathcal{H}\}}{\Pr\{c_k = 0 \mid \mathbf{r}, \mathcal{H}\}} = \log\frac{\sum_{\mathbf{s}:c_k=1}\Pr\{\mathbf{r}\mid\mathbf{s},\mathcal{H}\}\Pr\{\mathbf{s}\}}{\sum_{\mathbf{s}:c_k=0}\Pr\{\mathbf{r}\mid\mathbf{s},\mathcal{H}\}\Pr\{\mathbf{s}\}} \\ &= L_{pr}(c_k) + \underbrace{\log\frac{\sum_{\mathbf{s}:c_k=1}\exp\left[-\left\|\mathbf{r}-\sqrt{\rho_s}\mathcal{H}\mathbf{s}\right\|^2 + \sum_{i\neq k:c_i=1}^{N_b} L_{pr}(c_i)\right]}{\sum_{\mathbf{s}:c_k=0}\exp\left[-\left\|\mathbf{r}-\sqrt{\rho_s}\mathcal{H}\mathbf{s}\right\|^2 + \sum_{i\neq k:c_i=1}^{N_b} L_{pr}(c_i)\right]}}_{L_{ext}(c_k)}, \end{aligned} \tag{49}$$

for $k$ =1, 2,3, …, $N_b$, which can be feed-forwarded to the decoder. Note that $L_{pr}(c_k)$'s are the extrinsic messages forwarded from the LDPC decoder. Referring to the illustration given in Figure 2, the *extrinsic* messages, $L_{ext}(c_k)$'s, are the ones that need to be forwarded to the LDPC decoder. This exchange of extrinsic messages between the decoder and the MIMO detection unit will be referred to as the *super-iteration.*

## *C. The internal iteration for LDPC decoding*

The message-passing decoding algorithm follows the development of Gallager described in [xvii]. The message-passing algorithm runs on the bipartite graph. A number of decoding iterations are run and the *aposteriori* LLRs on each and every coded bit is obtained for final decision at the end of a fixed number of iteration. In Figure 2, we name this *aposteriori* LLRs as $L_{ap}(c_k)$. The prior messages (or the extrinsic message fedback to the MIMO detection unit) can be generated by removing the message forwarded to the decoder, i.e. $L_{pr}(c_k) = L_{ap}(c_k) - L_{ext}(c_k)$. As compared to the *super-iteration* between the decoder and the MIMO detection unit, we will call this loop the *internal* LDPC decoding iteration.

# VII. NUMERICAL EVALUATION AND SYSTEM SIMULATION

In this section, we will consider two kinds of experiment. One is a non-ensemble based approach and the other is an ensemble based approach. A non ensemble based approach is to select an LDPC code randomly from an ensemble and simulate. Usually, the probability of selecting a bad code is small, thus the selected code does not exhibit any error floor behavior. The probability of such event is much higher than selection of a bad code and exhibiting an error floor. Since the error floor of the union bound is always caused by the weight 2 component of the distance spectrum, the union bound for LDPC codes almost always shows an eminent error floor. This may make one wonder if there is an inconsistency. To remedy this problem, we consider simulation based on an empirical ensemble. This is a randomly selected ensemble of codes. How big the size of this ensemble should be? The right size for this ensemble can be quantitatively calculated from the distance spectrum.

We will first discuss our simulation results based on the ensemble approach, then the non ensemble based approach.

## *A. Ensemble Based Simulation Approach*

In this section, we compare the derived upper bounds with simulation results for fast Rayleigh fading channels. The bounds are evaluated for the ensemble of Gallager's LDPC codes, whose distance spectrum is calculated according to [xxiii]. In simulation, the receiver is assumed to have perfect channel state information and deploys the standard turbo-iterative decoding algorithm. While interested readers are referred to [xxiv] for details on turbo-iterative procedure, we briefly sketch the algorithm of the turbo-iterative receiver shown in Fig. 2 used in our simulation. The detector takes both the channel observations and the *a priori* information $L_{A1}$ from the decoder to compute the new *a posteriori* information $L_{D1}$ on each coded bit. The calculation is based on the standard maximum a posteriori (MAP) algorithm. The difference between $L_{D1}$ and $L_{A1}$ is referred to as the *extrinsic* messages which are used as the *a priori* input, $L_{A2}$, to the LDPC decoder. Then, the decoder generates the *a posteriori* information $L_{D2}$, and feedbacks the corresponding extrinsic messages as *a priori* information to the detector, thus completing a single cycle of an iteration. There are two kinds of iteration—the *super* and the *internal* iteration. The *super* iteration refers to the iteration between the detector and the decoder. The *internal* iteration refers to the message passing iteration between the check and the bit nodes within the LDPC code graph.

To obtain the average performance of a given code ensemble, we use each randomly generated LDPC code for ten codeword transmissions. There are a total of 10,000 LDPC codes in an empirical ensemble (the choice of the size will be discussed shortly later). The simulation is terminated at each SNR point if either 1,000 word error events have occurred or 100,000 transmitted codewords have been transmitted. For fair comparison, the error performance is plotted with respect to the normalized SNR, $E_b/N_0 = MN\rho_s / R_t$, where $\rho_s$, as defined before, is the average symbol energy at each transmit antenna, $R_t = RMK_b$ is the transmission rate of the system in information bits per channel-use and $R = K/L$ is the rate of the LDPC code.

The turbo receiver operates on the combined graph of the detector (or the constellation demapper) and the decoder. Union bound is an upper bound on the maximum likelihood receiver. While the two receivers are different, union bounds have been successfully used to gauge the performance of the iterative algorithms. In this section, we compare the performance of turbo receivers with the union upper bounds for LDPC codes with different block lengths, the results of which is illustrated in Fig. 3. In simulation, the detector and the decoder exchange the extrinsic information over 10 *super*-iterations, while the LDPC decoder runs its own message-passing decoding operation over 20 *internal* iterations. This means there are a total of 200 iterations per decision on a codeword.

We settle on these iteration numbers after repeated simulation experiments, and have a confidence that the decoding results will not be improved any significantly further without orders of magnitude increase in the number of iterations. Thus, it is likely that the simulation results reported in this paper are close to the best performance that the turbo-iterative receiver can provide for each scenario. The ultimate limit for the turbo-receiver would be the threshold value computable by the EXIT chart analysis or the density evolution method. Since these analyses are meant for infinite length and infinite number of iterations, one may not be assured 100% anyhow on how close the performance is to the limit at a finite block length and a finite number of iterations. On the other hand, we note from simulation that smaller iteration options, say 2 super and 5 internal iterations, renders performance curves which go well above the upper bound curves.

The foregoing discussion implies that the union bound results can be used as benchmarking references for turbo receiver designs. One can get a sense on how many iterations should be done to come close to, or even surpass, the maximum likelihood union upper bound, for a given constellation, at a given block length, and for a given coding scheme.

As shown in Fig.3, the performance of LDPC coded MIMO systems improves as the block length is increased; but still stays around the cutoff-rate limit [xxv] for block lengths up to few thousands. It is known that union bounds work well within this cutoff rate region. This is verified in the figure where the derived upper bound provides close prediction on the waterfall position (less than 0.5dB SNR gap) and on the error floor. Note that this is meaningful observation as LDPC codes with moderate block lengths up to few thousands are suitable for high speed communication applications with stringent requirements on decoding complexity and delay.

As the block length approaches infinity (e.g., more than a million), the turbo-iterative receiver will likely perform at a level very close to the *threshold* value predicted by density evolution [xxvi]. The threshold would be located at a point beyond the cut-off rate and come near the capacity. For a large block length at which the union bound becomes loose in describing the water-fall behavior, one can resort to some tighter bounding techniques such as the first and second Gallager bounding techniques and their recent variations. These tight bounding techniques are yet to come for MIMO systems, even though we envision that a tighter bounding technique can be built upon the combinatorial union bound developed in this paper.

Another observation is that the error floor predicted by the union bound is consistent with the simulation result, except that it is shifted upward by 5 to 10 dB. This upward shift in the error curve is mainly due to the Chernoff bound, $Q(x) \le \exp(-x^2/2)$, used in obtaining the pairwise error probability in (36). To see this effect, let us consider the case of (3000, 3, 6) LDPC code in Fig. 3 and suppose $E_b/N_0$ = 5dB or $\rho_s$ = 1.5832. The minimum possible Hamming distance of these codes is 2 because only codewords with even weights are valid. For simplicity, let us roughly assume the fading gain is equal to 1. In this case, it is easy to verify that the pairwise error probably between any two codewords of distance 2 is always equal to $Q(\sqrt{4\rho_s})$. The ratio of this $Q$ function and its Chernoff bound is 0.14, or −8.52 dB, which explains the result shown in Fig. 3. In addition, it is clear that the error floor is lowered as the block length is increased.

Our simulation result is consistent with our expectation. The error floor of the ensemble is the manifestation of one of well known Gallager's results (See Theorem 2.4 [Gallager's thesis]): as the block length is increased, an ensemble of codes tends to contain fewer and fewer bad codes which contain the codewords of "small" weights but these few codes dominate the performance over the ensemble. Small here is meant to indicate a number less than the *minimum distance* of the ensemble of codes. Thus, a small number of few bad codes dominate the error floor behavior of the ensemble. In reality, it is difficult to choose such a bad code in random selection, especially for a code with a large block length. This is the underlying reason why we choose to take the approach of ensemble average in simulation to determine the performance of LDPC codes, rather than the approach to simulate a particular code randomly chosen from the ensemble.

The appropriate size of this empirical ensemble of codes can be determined from examining the distance spectrum of the ensemble, and thus depends up on the choice of a block length. For a quantitative discussion, let us take the ensemble of (3000, 3, 6) LDPC codes. Only words with even weights are candidates for valid codewords. The poorest codes are thus the ones with weight 2 codeword. The error floor due to weight 2 looks almost flat, as shown in our union bound figures. The distance spectrum of this ensemble at Hamming weight 2 is $A_2 = 0.0018$. This number roughly implies that in the worst case there are as much as 18 bad codes in 10,000 selections each of which has a single weight 2 codeword. In the best case scenario, there is a single code out of 10,000 which has 18 weight 2 codewords. From this argument, we note that our choice of ensemble size 10,000 makes sense. It is highly likely that the ensemble of 10,000 codes includes at least one bad code. In turn, this implies that one can easily avoid the selection of a poor code in practice: just draw another code upon noticing an error floor in simulation.

It is worth noting that the error floor behavior predicted in union bound is somewhat different for the turbo code case. Some features in turbo code cases include that the slopes of error floors are usually stiffer, and the simulation result of a randomly selected single code coincides well with the prediction made by the union bound. Why? The short answer for this question is that there is no fractional component in the distance spectrum of the turbo code ensemble. The constituent convolutional code in turbo codes is usually chosen with a decent minimum (*free*) distance, usually much larger than 2. Thus, the error floor shows a stiffer slope. The distance spectrum of the ensemble of turbo codes is obtained through the random scramble and normalization of two distance spectra (or weight enumerating functions in [xxvii]) of the two components codes via an abstract device called the *uniform interleaver*. Since the component code has no fractional component, the distance spectrum of the turbo code does not have fractional components either. The distance spectral component of the turbo code is simply zero for all those weights smaller than the minimum distance of the constituent code. This means with probability close to 1, no code in the turbo code ensemble has any codeword whose weight is less than the minimum distance. With probability approaching 1 asymptotically as the block length increased, therefore, a turbo code with a randomly selected interleaver will exhibit the typical performance predictable by the error floor of the union bound.

We next compare the upper bound and the simulation results in different channel and modulation scenarios. In this case, the iterative algorithm runs 5 super iterations and 20 LDPC internal iterations. Illustrated in Fig. 4 and 5 are the comparisons of the bounds with the two different bit error rate simulation results: The first bit error rate, denoted as $P_{b|0}$, is obtained by simulations where the all-zero codeword is transmitted all the times. The second, denoted as $P_b$, is obtained in simulations where all codewords are transmitted equiprobably. This has been done to measure the tightness of the further upper bound with the all-zero codeword mapped to the worst space-time symbols in the constellation. As expected, the two error probabilities $P_b$ and $P_{b|0}$ are very close to each other for PSK modulation systems because of the symmetry in the $Q$-ary base constellation and thus in its hyper-constellation of the space-time words. In the case of 8QAM modulation, however, the all-zero codeword is mapped onto a space-time word closest to the mass-center of the hyper-constellation, and thus simulations with the all-zero codeword only transmissions leads to a performance result worse than those with all codewords transmitted equiprobably. From the results, we can at least expect that the upper bound developed can be useful for PSK modulation systems, while for constellation with unequal energy symbols the bound should be improved. Nevertheless, in all scenarios, the bounds are shown useful to predict the error floors.

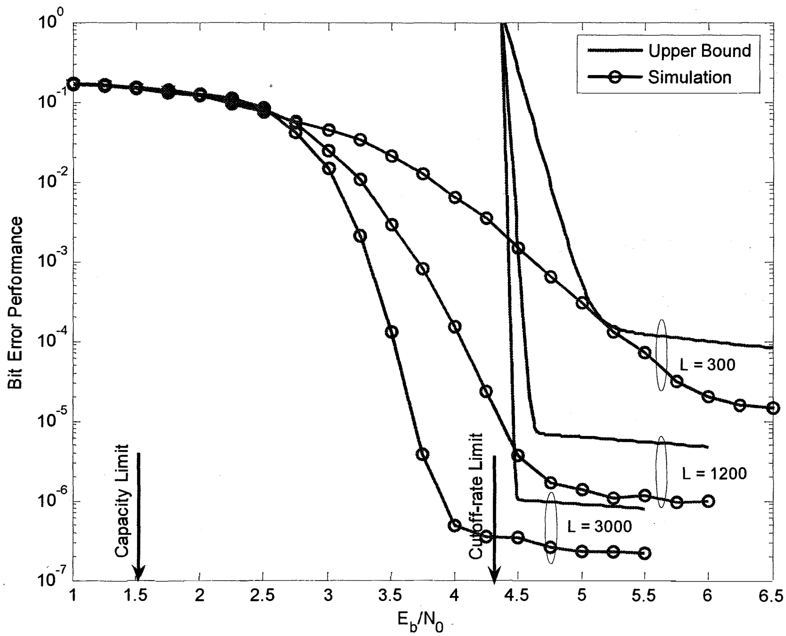



Fig. 3: Simulation results v.s. upper bounds for LDPC codes with different block lengths $L$. (4PSK, $M = N = 2$)

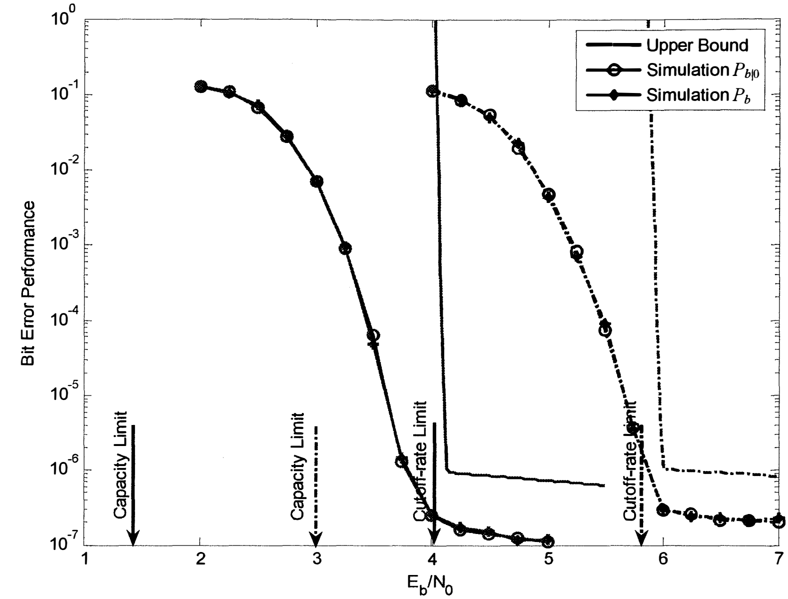



Fig. 4: Simulation results v.s. upper bounds for LDPC (3000,3,6) coded MIMO systems. Solid curves are the results for 4PSK modulated 4 × 4 MIMO systems, while dot-dashed curves 8PSK modulated 2 × 2 MIMO systems. Gray mapping is adopted in the PSK modulation.

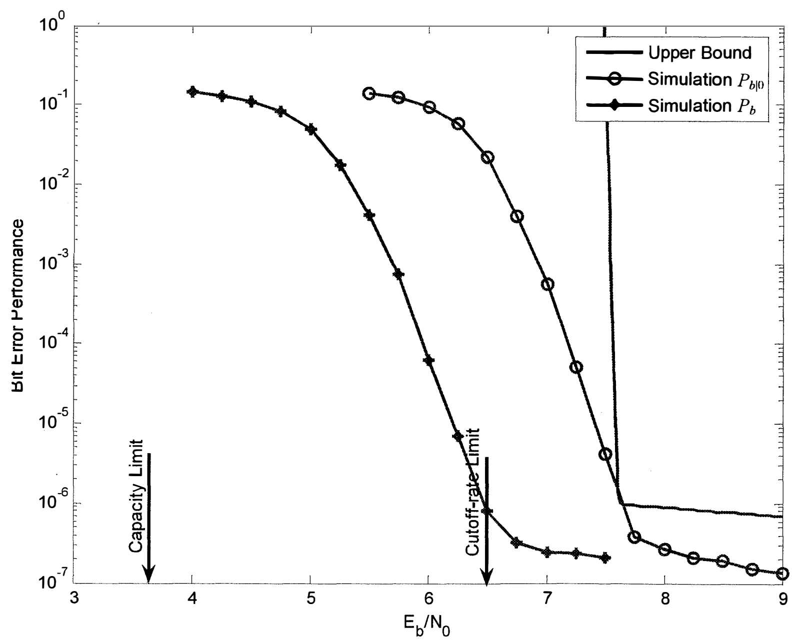



Fig. 5: Simulation results v.s. upper bounds for LDPC (3000,3,6) coded MIMO systems with 8QAM modulation. Gray mapping is adopted which maps the bit strings such as 000, 001, ..., 111 to constellation points (+1, +1), (+1, −1), (+2, +1), (+2, −1), (−1, +1), (−1, −1), (−2, +1) and (−2, −1), respectively.

## Reference for Ensemble Based Simulation Approach

[1] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *IEEE Commun. Theory Workshop*, Aptos, CA, 1999.

[2] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections and applications," IEEE Trans. Info. Theory, vol. 48, pp. 3029-51, Dec. 2002.

[3] I. Sason, S. Shamai and D. Divsalar, "Tight exponential upper bounds on the ML decoding error probability of block codes over fully interleaved fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp.1296 - 1305, Aug. 2003.

[4] H. Bouzekri and S. L. Miller, "An upper bound on turbo codes performance over quasi-static fading channels," *IEEE Trans. Inform. Theory*, vol. 7, no. 7, pp. 302-304, Jul. 2003.

[5] Jingqiao Zhang and Heung-No. Lee, "Performance analysis on coded systems over quasi-static (MIMO) fading channels", *Proc. Int. Conf. Communications*, Seoul, Korea, May 2005.

[6] Jingqiao Zhang and Heung-No Lee, "An upper bound on the error performance of coded system over Rayleigh fading MIMO channels," *IEEE Communications Letter*, vol. 9, no. 9, Step. 2005.
[7] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," IEEE Trans. Commun., vol.52, pp. 670-8, Apr. 2004.
[8] M. Schwartz and A. Vardy, "On the stopping distance and the stopping redundancy of codes," *IEEE Trans. on Inform. Theory*, vol. 52, no. 3, pp. 922-32, Mar. 2006.
[9] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47,no. 2 , pp. 619 – 637, Feb. 2001.
[10] R. G. Gallager, *Low Density Parity Check Codes*, MIT Press, 1963.
[11] V. Tarokh, N. Seshadri, and A.R.Calderbank, "Space–Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction", *IEEE Trans. Inform. Theory*, vol.44, no.2, pp.744 - 765, Mar. 1998.
[12] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: asymptotic distance distributions," *IEEE Trans. on Inform. Theory*, vol. 48, no. 4, pp.887 – 908, Apr. 2002.
[13] B.M. Hochwald, and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Commun.*, vol. 51, no. 3, pp. 389 – 399, 2003.
[14] J. Zhang and Heung-No Lee, "A Performance Bound on Random-Coded MIMO Systems," *IEEE Commun. Letters*, vol 10, no.3, pp.168-170, March 2006.
[15] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation", *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 657 – 670, Feb. 2001.
[16] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. On Inform. Theory*, vol.42, no.2, pp. 409-28, Mar, 1996.

## *B. Non-ensemble based simulations*

For Non-ensemble based approach, we will consider a few space-time block code examples for numerical evaluation of the bounds and for Monte Carlo system simulation of the turbo-iterative receiver described in Section VI. We first consider our choices of two different constellation maps. Then, discussion and comparison on the bounds and simulation results will be given.

### *1) Constellation Map, the All-Zero Outer Codeword and The Matched Masking*

In this section, we shall address the constellation maps.

A specific constellation map between the outer code and the space-time block code is needed for both numerical evaluation of union bounds, as well as for system simulations.

Recall from section IV.B that we obtained a further upper bound by resorting to a test codeword composed of the worst IST symbols, i.e. $\mathbf{X}^* = [\mathbf{s}_{(*)}\ \mathbf{s}_{(*)} \cdots \mathbf{s}_{(*)}]$. This helps us to avoid the average operation in which one has to choose each and every space-time codeword as a test word.

We have taken the following steps with regard to the mapping rules in our bounds and simulation results:

- Step-I: On the base constellation, such as 4-PSK or 8-PSK, regardless of the IST coding scheme, we select and use a Gray-map. This is a basic set up which will be called 'Scheme I.'
- Step-II: Find the worst IST symbol $\mathbf{s}_{(*)}$ from (19), and form the test space-time codeword $\mathbf{X}^* = [\mathbf{s}_{(*)}\ \mathbf{s}_{(*)} \cdots \mathbf{s}_{(*)}]$.
- Step-III: From the selected map in Step-I, obtain the bit pattern for $\mathbf{X}^* = [\mathbf{s}_{(*)}\ \mathbf{s}_{(*)} \cdots \mathbf{s}_{(*)}]$. Let's use $\mathbf{b}^*$ to designate the bit-pattern for $\mathbf{s}_{(*)}$. The length of $\mathbf{b}^*$ is $N_b$. Thus, $[\mathbf{b}^*\ \mathbf{b}^* \cdots \mathbf{b}^*]$ is for $\mathbf{X}^*$.
- Step-IV: Designate $\mathbf{b}^*$ as the mask and apply this mask to obtain a new mapping rule for all bit patterns for the $J$-ary IST symbols. The masking operation is to take the bit-by-bit XOR operation to each and every map under Scheme I with

the mask $\mathbf{b}^*$. One consequence of this masking operation is to have the pre-image of the space-time word $\mathbf{X}^*$ be the all-zero codeword. Let's call this 'Scheme II.'

Both mapping schemes can be implemented easily for system simulations. In simulations, therefore, we attempt to examine if there should be any performance difference between the two mapping schemes. From intuition, we expect no performance. A verification of no simulative difference between the two schemes will give us the assurance that the union bound, obtained based on Scheme II as the mapping rule and the transmission of all-zero outer codeword as the test word, is well defined and will provide a useful bound.

In table I, we tabularize the Almouti code example for both mapping schemes. Note that the basic map for the 4-PSK constellation is given as 00, 01, 10, 11 for $-j$, $-1$, 1, and $j$. Scheme I is simply the concatenation of these for the two symbols $s_a$ and $s_b$. Since the PSK constellation is symmetric, any IST symbol can be selected as the test word. Choose $(-j, -j)$ as the test word. Then, the bit pattern (1 0 1 0) is selected as the mask. XOR'ing the bit strings with the mask, one can obtain Scheme II, as shown in Table I. A table for the linear dispersion code can be generated similarly.

*2) Examples*

The upper bound analysis conducted below is done under the Scheme II mapping rule.

However, it is also expected to serve as a good performance benchmark for Scheme I as well. In fact, system simulation results indicate that both schemes yield almost the same performance (See Figure 5 and Figure 6).

*Example-1: (*The *Alamouti code)* Let' us consider the ($N_t$=2, $N_r$) Rayleigh fading channel. We suppose the use of 4-PSK ($M$=4) base constellation and the Alamouti code [xxviii]). Alamouti code takes two channel symbols and maps them to a two-by-two space-time matrix in the following way,

$$\begin{bmatrix} s_a \\ s_b \end{bmatrix} \Rightarrow \mathbf{S} = \begin{bmatrix} s_a & -s_b^* \\ s_b & s_a^* \end{bmatrix}. \tag{50}$$

There are thus $J = 16$ different space-time matrix-symbols (Alamouti codewords) in the constellation; each symbol is transmitted over $T_D = 2$ channel uses. We have

$$\sum_{j=1}^{J} \beta_j z^{w_j} = 1 + 4\left(1+\frac{\rho_s}{2}\right)^{-2N_r} z + 6\left(1+\rho_s\right)^{-2N_r} z^2 + 4\left(1+\frac{3\rho_s}{2}\right)^{-2N_r} z^3 + \left(1+2\rho_s\right)^{-2N_r} z^4. \tag{51}$$

Note that Eq. (51) is the key step to obtain the union bound (47). For the union bound calculations (for (28) or (38)), we first collect the coefficients of the utility variable $z$ into a sequence, i.e., $[1 \;\; 4(1+\rho_s/2)^{-2N_r} \;\; 6(1+\rho_s)^{-2N_r} \;\; 4(1+3\rho_s/2)^{-2N_r} \; (1+2\rho_s)^{-2N_r}]$ and take the $D$-fold *convolution* of the sequence. The output of the convolution is the sequence $\langle \phi(h) \rangle$, from $h$=0, 1, 2, …, $N$.

*Example-2:* (The *Linear Dispersion* code [ix, Eq. (35), (36)]) Hassibi and Hochwald proposed a new class of space-time block codes in [ix] which is claimed to subsume many previously proposed space-time schemes including the orthogonal space-time block codes, Alamouti codes, V-BLAST, and others. A linear dispersion code is a collection of $J$ [$N_t$ x $T_D$] dispersion matrices. Each dispersion matrix is composed of $Q$ basic dispersion matrices, $\mathbf{A}_q$ and $\mathbf{B}_q$, for $q$ =1, 2, …, $Q$. Each $\mathbf{A}_q$, or $\mathbf{B}_q$, are real valued matrices found from an optimization which is aimed to maximize the so-called *perfect-knowledge* channel capacity (ensemble averaged over the fading channel). By modulating these base matrices with regular complex-valued constellation points, $a_r + j\, a_i$, one can obtain a space-time transmission symbol **S**, i.e.

$$\mathbf{S} = \sum_{q=1}^{Q} a_{r,q}\mathbf{A}_q + a_{i,q}\mathbf{B}_q \; . \tag{52}$$

By ensuring, through a design in the optimization process, the dispersion matrices, $\mathbf{A}_q$ and $\mathbf{B}_q$, q=1, 2, …, $J$, disperse the energy of the symbols $a_{r,q}$ and $a_{i,q}$ equally in all spatial and temporal directions. One of the contributions of the chapter is to be able to explicitly show the benefit of the rate optimized space-time block code in the form of favorable error performance. For Alamouti code as an example, they show that when $N_r > 1$, the use of Alamouti code is deficient in terms of capacity. That is, it does achieve the order of diversity two times $N_r$; but it falls short in terms of the transmission rate. Using the linear dispersion code, one can achieve the same maximum order of diversity while significantly reducing, or almost recovering in some cases, the sacrifice in transmission rate.

In this chapter, we consider two example linear dispersion (LD) codes given in [ix] for ($N_t$ =3, $N_r$=1, $T_D$=4, $Q$ =3). The LD code in [ix, see Eq. (35)] is given by

$$\mathbf{S}=\sqrt{\tfrac{4}{3}}\begin{bmatrix} a_{r,1}+ja_{i,1} & a_{r,2}+ja_{i,2} & a_{r,3}+ja_{i,3} \\ -a_{r,2}+ja_{i,2} & a_{r,1}-ja_{i,1} & 0 \\ -a_{r,3}+ja_{i,3} & 0 & a_{r,1}-ja_{i,1} \\ 0 & -a_{r,3}+ja_{i,3} & a_{r,2}-ja_{i,2} \end{bmatrix}. \tag{53}$$

The other linear dispersion code is from Eq. (36) in [ix] for ($N_t$ = 3, $N_r$ =1, $T_D$ = 4):

$$\mathbf{S}=\sqrt{\tfrac{4}{3}}\begin{bmatrix} a_{r,1}+a_{r,3}+j\left[\frac{a_{i,2}+a_{i,3}}{\sqrt{2}}+a_{i,4}\right] & \frac{a_{r,2}-a_{r,4}}{\sqrt{2}}-j\left[\frac{a_{i,1}}{\sqrt{2}}+\frac{a_{i,2}-a_{i,3}}{2}\right] & 0 \\ \frac{-a_{r,2}+a_{r,4}}{\sqrt{2}}-j\left[\frac{a_{i,1}}{\sqrt{2}}+\frac{a_{i,2}-a_{i,3}}{2}\right] & a_{r,1}-j\frac{a_{i,2}+a_{i,3}}{2} & -\frac{a_{r,2}+a_{r,4}}{\sqrt{2}}+j\left[\frac{a_{i,1}}{\sqrt{2}}-\frac{a_{i,2}-a_{i,3}}{2}\right] \\ 0 & \frac{a_{r,2}+a_{r,4}}{\sqrt{2}}+j\left[\frac{a_{i,1}}{\sqrt{2}}-\frac{a_{i,2}-a_{i,3}}{2}\right] & a_{r,1}-a_{r,3}+j\left[\frac{a_{i,2}+a_{i,3}}{\sqrt{2}}-a_{i,4}\right] \\ \frac{a_{r,2}-a_{r,4}}{\sqrt{2}}+j\left[\frac{a_{i,1}}{\sqrt{2}}+\frac{a_{i,2}-a_{i,3}}{2}\right] & -a_{r,3}+ja_{i,4} & -\frac{a_{r,2}+a_{r,4}}{\sqrt{2}}+j\left[\frac{a_{i,1}}{\sqrt{2}}-\frac{a_{i,2}-a_{i,3}}{2}\right] \end{bmatrix}. \tag{54}$$

The LD code given in (53) in fact is the orthogonal block code with $Q$ = 3, and achieves a mutual information 5.13 bits/channel-use at 20 dB SNR; while the LD code (54) has been obtained from a gradient search with $Q$ = 4, and has mutual information 6.25 bit-per-channel-use at the 20 dB SNR (As compared to the capacity 6.41 bpcu). In this chapter, the performance of these codes will be compared in both the union bound analysis as well as the system simulation with the turbo-iterative detection algorithm.

One thing to notice is that the optimized LD code gives about 2-3 dB SNR performance advantage over the orthogonal space-time block code (see Fig. 6 in [ix]) in all SNRs when both are compared at a fixed transmission rate for fairness. For this, we use 8-PSK as the base constellation for the LD code (53) and 4-PSK for the LD code (54). Once all the symbol-matrices are given, one can find the reference symbol-matrix $\mathbf{s}_{(*)}$ according to (19), and evaluate the sum of the pairwise metrics as done in example (51) for any specific channel model.

*3) Comparison of the Union Bounds and System Simulation Results*

We have three scenarios. In Scenario 1, the IST code is the linear dispersion code given in Eq. (54). We use 4-PSK base constellation. Thus, the number of coded bit each IST symbol carries is $N_b = Q\log_2(M) = 4\log_2(4) = 8$. In Scenario 2, the IST code is the linear dispersion code (which is in fact an orthogonal space-time block code) given in

(53). The number of coded bits each IST symbol carries is $3\log_2(8)=9$. In both scenarios, an IST symbol consumes $T_D=4$ channel uses. In Scenario 3, the Alamouti code with 4-PSK is used. Thus, each Alamouti codeword carries $2\log_2(4)=4$ coded bits in $T_D=2$ channel uses. The rate of outer LDPC code is 1/2. In order to make *N* to be a multiple of $N_b$, Gallager's ($N$=2700, $d_b$=3, $d_c$=6) LDPC code [xvii] for Scenario 2 and (3000, 3, 6) for Scenario 1 and 3 are used. With these schemes, the transmission rates for Scenario 1, Scenario 2 and Scenario 3 are calculated respectively as $8\times 1/2\times 1/4=1$, $9\times 1/2\times 1/4=9/8$, and $4\times 1/2\times 1/2=1$ [info-bit-per-channel-use]. Thus, in all scenarios, around 1 information bit per channel use is transferred. The first two scenarios use $(N_t=3, N_r=3)$ MIMO, while the third scenario uses $(N_t=2, N_r=2)$ MIMO, Rayleigh flat-fading channels.

The detector and the decoder exchange the extrinsic log-likelihood ratios in super turbo-iterations, while the decoder runs its own internal LDPC decoding operation.

For the Scheme II mapping rule, we select the mask vectors to be [1 0 1 1 1 0 0 0 1] and [1 0 1 0] for Scenario 1 and 3 respectively because with 4-PSK the function in used to obtain $\mathbf{s}_{(*)}$ (19) yields the same value for each and every $\mathbf{s}_{(j)}$. In Scenario 2, there exist several maximizing IST symbols $\mathbf{s}_{(*)}$. We select the one with the smallest index *j*; the corresponding mask vector is [0 0 0 1 0 0 0 1].

We first compare the bounds and simulation results for Scenario 2 over the independent fading ($N_t$=3, $N_r$=3) channel. For Figure 3, we use five super-iterations and ten internal LDPC iterations, while for Figure 4 we use fifty LDPC internal iterations. From the figures, we see the impact of super-iteration vs. the internal LDPC iterations. We see that the impact of super-iteration becomes very small after three iterations. But note that the impact of internal LDPC iteration is relatively large. Increased from 10 to 50, the internal LDPC iteration has given us about 0.5 dB SNR advantage.

For the rest of simulations, we use three *super* iterations and ten *internal* LDPC iterations. Therefore, the decoder runs thirty iterations in total.

The derived bounds for the independent fading case are compared with system simulation results in Figure 6. First note that, as expected, the performances of Scheme I and Scheme II are almost the same in all investigated scenarios. The figures generally show less than a 0.5 dB SNR difference between the union-bound upper bound and the simulation result. This is in contrast to the case of binary coded AWGN systems, where about 2-3 dB SNR gap usually exists between the union bound and turbo-iterative simulation result.

The same coding schemes and scenarios are used for block fading cases, shown in Figure 5. In block fading cases, the channel stays fixed once randomly chosen during the transmission of an IST symbol—during $T_D$ channel uses. The figure again shows very small SNR differences between Scheme I and Scheme II and they are less than 0.5 dB away from the water-fall SNR of the union-upper bound.

From the two channel cases, we note that the upper bounds derived in this chapter shall serve as a good performance prediction tool. We also note that the performance is almost the same for the two different channel cases. Basically, it is because the amount of time-selective diversity is very large in both cases. Even in block fading channels, the duration of fixed channel fading ($T_D$ channel uses) is much smaller than the total number of channel uses $T$. For Scenario 2, $T_D$ =4 while $T$ =1200, and $D$=2700/9=300; thus the order of time-diversity is 300.

# VIII. DISCUSSION FOR THE BOUNDS ON THE QUASI-STATIC FADING CHANNELS

During the past several years, much research effort has been spent on the prediction of the error performance for turbo- and low-density parity-check (LDPC) coded systems with the maximum likelihood (ML) decoding assumption. This interest has been motivated by the splendid error correction performance of turbo-like codes which comes very close to the theoretical limit at a large block size. In a region close to the capacity limit, it has been known, that the usual union bound is loose. Thus, the demand for finding tight performance bounds that would continuously be useful in this region has been very high. Fulfilling this need, there has been a series of substantial recent progress such as 0,0,0,0,0,0. Namely, they are variations on the so called Fano-Gallager's tight bounding methods (also called as the *limit-before-averaging* bounding technique), which were originally introduced by Fano 0 and further developed by Gallager for tight bounds on the error performance of LDPC codes operating on AWGN channels in 1963 0. There is a semi-tutorial paper by Shamai and Sason 0 for further reading to understand recent developments and how the recently developed bounds are inter-related with each other. Included in this chapter are the Duman-Salehi bound, the Divsalar bound, the Viterbi-Viterbi bound, and the Suman-Feder bound.

In this chapter, we are interested in developing a tight performance bounding technique for the space-time transmission of low-density parity-check code over the multiple input multiple output channel. One of objectives is to strike a balance between the performance and the complexity of the evaluation technique.

Under the assumption that the all-zero codeword $\underline{c_0}$ is transmitted, the Fano-Gallager bounding technique can be started with the following decomposition:

$$\Pr(\text{error}) = \Pr(f \in \Re)\Pr(\text{error} \mid f \in \Re) + \Pr(f \in \overline{\Re})\Pr(\text{error} \mid f \in \overline{\Re}) \qquad (55$$

where $f$ is an utility function, called the Fano-Gallager tilting measures, of performance-related random variables such as the additive white Gaussian noise $w$ and the channel fading gain $\alpha$. $\Re$ and $\overline{\Re}$ are disjoint geometrical regions. They can be found sought to minimize the right hand side of (55). On the first error term, the union bound is applied; on the second term, a trivial bound $\Pr(\text{error} \mid f \in \overline{\Re}) \le 1$ is used.

Depending on how tight a bound we want to have, a rather complicated definitions on the geometrical region and the utility function $f$ can be made, as we examine from the tight bound results obtained for single input single output channels. Considering the received signal over a AWGN channel,

$$y = \alpha x_0 + w, \qquad (56)$$

where $\underline{x_0}$ is the modulated signal for the all-zero binary codeword $\underline{c_0}$, $\alpha$ is the channel gain of 1, and $w$ is the AWGN noise, for example, Divsalar 0 defines the region to be a hyper-dimensional sphere, $\Re$, and takes the approach of optimizing the radius and the position of the sphere for a bound. Since the region is a rather simple sphere, the bound is obtained in a closed form. But we note that the bound is not tight and the bound is greater than 1 in low SNR region.

A tighter bounding technique for single-input single-output fading channel takes in general the more complicated form. For the discussion of fading channel, let's model the channel as having in equation (56) with an independent fading gain $\alpha$ for each channel symbol transmission. In this case, the utility function $f$ and the region depends both on the noise $w$ and the fading gain $\alpha$, and a full blown application of the Fano-Gallager tilting measures becomes a very complex task and the obtained bounds are cumbersome to be evaluated. For example, the result obtained in 0 has three parameters in the final expression of the bound to be optimized numerically.

In this chapter, we aim to find a simpler but effective approach that does not leave any parameter to be optimized numerically. This will help us proceed with the more complicated transmission cases which is the space-time transmission of binary block codes over multiple transmit and multiple receive antenna cases.

In our approach, we select the regions $\mathfrak{R}$ and $\overline{\mathfrak{R}}$ in order to make the distinction between the "high" and "low" instantaneous SNR events. By adopting the union bound for the high instantaneous SNR region, i.e. $f \in \mathfrak{R}$, while the trivial bound is used for the small instantaneous SNR region, we have

$$\Pr\{\text{error}\} \le \Pr(f \in \mathfrak{R}) \sum_{c' \neq c_{\underline{0}}} P(c_{\underline{0}} \to c' \,|\, f \in \mathfrak{R}) + \Pr(f \in \overline{\mathfrak{R}}). \tag{57}$$

The summand in the inequality is the pairwise error probability from $c_{\underline{0}}$ to any other codeword $c'$ conditioned on $f \in \mathfrak{R}$. Namely, we take the event to be defined as an "outage" event such that the fading gain is smaller than a certain optimal threshold value. We note that a similar approach has been independently adopted by Bouzekri and Miller in 0 to find tight union bounds for turbo-coded modulation signals over quasi-static fading channels.

Our analysis framework can be used to subsume the materials in 0 as a special case. The work in 0 focuses on binary coded modulation over Rayleigh channels, while our work, see Proposition 2, formulates the idea of Fano-Gallager's bounding technique in a general form so that the bounds can be obtained for any arbitrarily-sized constellations and for any fading channels with non-degenerate distribution (such as a distribution with a point mass), including Rayleigh, Ricean and Nakagami channels. Second, making use of the classical ideas of performance averaging over an ensemble of codes, we develop the notion of a distance spectrum for space-time transmissions, see Proposition 1. We develop this for the ensembles of LDPC codes but it can be readily extended to other linear block codes, including the turbo codes. Third, we make use of the fact that an equivalent SISO channel model can be derived for space-time block coded MIMO systems and have the upper bound extended to MIMO systems. Fourth, Bouzekri and Miller use the Chernoff bound, $Q(x) \le \exp(-x^2/2)$, to upper bound the Gaussian $Q$-function. This is rather loose and can be improved by making use of the Craig Identity $Q(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(\frac{-x^2}{2\sin^2\theta}\right) d\theta$. Note that it is not difficult to numerically evaluate the integral of a smooth function over a finite interval. In this chapter, we evaluate both approaches and illustrate the amount of benefit of our approach for both SISO and MIMO channels. Roughly, using the Craig identity gives us a tighter result, about 1dB in SNR.

**This part is not complete yet.**

REFERENCES FOR THIS SECTION


[1] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inform. Theory*, no. 40, pp.1284–1292, Jul. 1994.

[2] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2011 – 2018, Dec. 2003.

[3] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes", *IEEE Commun. Theory Workshop*, Aptos, CA, 1999.

[4] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems over AWGN and fading channels," *Global Telecommunications Conference*, vol. 3, pp. 1605 – 1610, Nov. 2000.

[5] Igal Sason, S. Shamai, and D. Divsalar, "Tight exponential upper bounds on the ML decoding error probability of block codes over fully interleaved fading channels," IEEE Trans. on Commun., vol 51, no. 8, pp. 1296-1305, August, 2003.

[6] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*, MIT Press, 1963.

[7] R. G. Galager, *Low Density Parity Check Codes*, MIT Press, 1963.

[8] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections and applications," IEEE Trans. Info. Theory, vol. 48, pp. 3029-51, Dec. 2002.

[9] H. Bouzekri and S.L. Miller, "An upper bound on turbo coded performance over quasi-static fading channels," *IEEE Commun. Letters*, vol. 7, no. 7, July 2003.

# IX. CHAPTER SUMMARY

Union bounds for the concatenated coding scheme have been obtained in this chapter for different classes of fading channels. The outer block codes assumed in this chapter are liner codes, except for the random block codes, with a known ensemble-averaged distance spectrum. The union bounds are then obtained by carefully considering a number of combinatorial codeword-enumeration problems arising from the use of the long block code deriving a sequence of short space-time block codes. We took the approach of taking the space-time transmission of size $[N_t \text{ x } T_D]$ as the unit of block-channel-use to accommodate the most general space-time transmission scheme over the block fading channel. Each inner space-time (IST) block matrix can carry $N_b$ number of coded bits and is transmitted during $T_D$ channel uses.

Through a number of specific examples explored in this chapter as to the selection of inner space-time block transmission schemes, including the linear dispersion codes, the usefulness of the concatenated scheme has been verified. With a length that is sufficiently long to cover a sizable number of the independent block-channel-uses, the use of a long outer block code shall bring the operation region of the concatenated scheme to a region (in terms of both SNR or transmission rate) very close to the ergodic channel capacity.

The usefulness of the proposed union bounding technique has been verified since the bounds faithfully predict the simulated performance of the iterative MIMO detection and decoding receiver. The verification was done at the block length of about 3000.

It is well known for the additive-white Gaussian channels and the single-input single-output fading channels that the use of a union bound for turbo-like codes is loose for the region of around and beyond the cut-off rate. A number of tight union bounds over the single-input and single-output channels with binary modulation schemes have been obtained in the past by a number of researchers (see, for example, a recent publication [xxix] and the references therein). Their basic methods are rooted on the so-called Fano-Gallager's tilting measure techniques [xxx]. For the concatenated MIMO transmission scheme considered in this chapter, we took the combinatorial union bound approach first.

This was motivated to see how the union bound will perform as compared to the simulated performance of the message-passing algorithm. On the foundation of our combinatorial union bound techniques developed in this chapter, more sophisticated performance analysis techniques, such as using the Fano-Gallager's techniques and the Bonferroni type bounds, can be applied to tighten the bounds. It is remained to be seen how these techniques can be used to tighten the gaps observed in this chapter.

## APPENDIX

Consider the ensemble of LDPC codes or concatenated codes in section III. Each of these codes map $K$ information bits into codeword of length $L$. Denote the generating matrix of the code as $G$; we can find its systematic form $G_s = (P\ I_K)$ by Gauss-Jordan elimination, where $I_K$ is the $[K\times K]$ identity matrix. Therefore, the last $K$ bits of the codeword are the information bits. For any codeword with input weight $\omega$ and output weight $h$, the weights of its first $L$ - $K$ bits and last $K$ bits are $\omega$ and $h$ - $\omega$ respectively. For simplicity, we denote this ($\omega$, $h$ - $\omega$) as a metric of the codeword.

Resorting to Theorem-I, we obtain the probability that any codeword **c** of weight $h$ has a metric ($\omega$, $h$-$\omega$) as

$$\begin{aligned}&\Pr\left(\mathbf{c} \text{ has a metric } (\omega, h-\omega)\middle|\mathbf{c} \text{ is of weight } h\right)\\ &=\binom{L}{h}^{-1}\binom{K}{\omega}\binom{L-K}{h-\omega}=:P_{(\omega,h-\omega)|h}\end{aligned} \quad (58)$$

The average number $A_{\omega,h}$ of the codeword with metric ($\omega$, $h$-$\omega$) in one code is therefore given by

$$A_{\omega,h} = A_h P_{(\omega,h-\omega)|h} = A_h\binom{L}{h}^{-1}\binom{K}{\omega}\binom{L-K}{h-\omega}. \quad (59)$$

By simple manipulation, we arrive to the final result:

$$\begin{aligned}A_h' &= \sum_{\omega=1}^{K}\frac{w}{K}A_{\omega,h} = A_h\sum_{\omega=1}^{K}\frac{\omega}{K}\binom{L}{h}^{-1}\binom{K}{\omega}\binom{L-K}{h-\omega}\\ &=\frac{hA_h}{L}\binom{L-1}{h-1}^{-1}\sum_{\omega=1}^{K}\binom{K-1}{\omega-1}\binom{L-K}{h-\omega}=\frac{h}{L}A_h.\end{aligned}$$

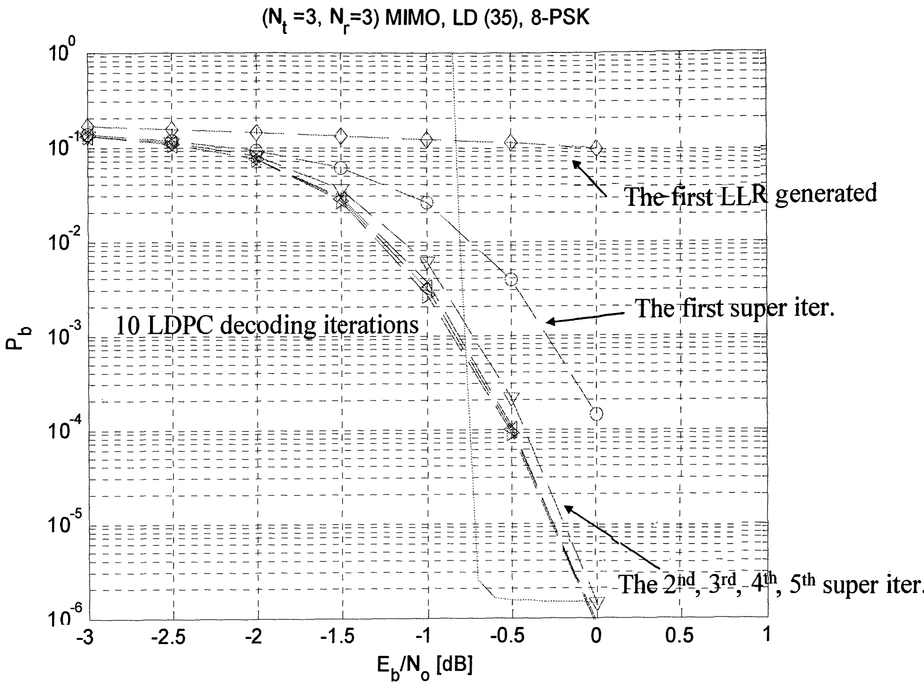



Figure 3: Scenario 2: ($N_t$ = 3, $N_r$=3) MIMO with Linear Dispersion code given in (53). The base constellation is the 8-PSK. The bounds compared with the simulated bit-error rate at each stage of super-iterations. There are five super-iterations. In each super-iteration, there are ***ten internal LDPC decoding iterations***.

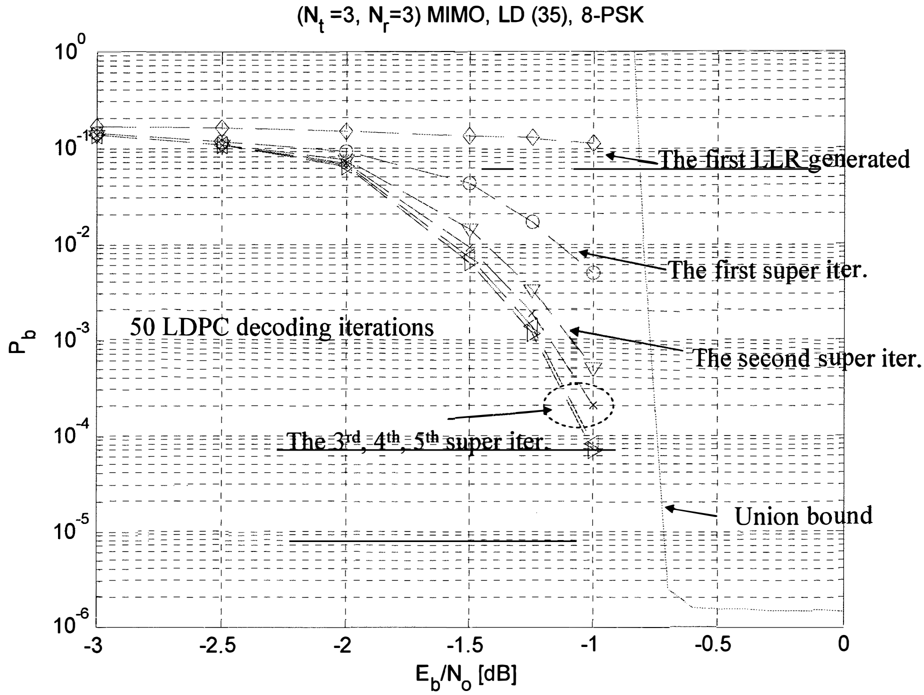



Figure 4: Scenario 2: ($N_t$= 3, $N_r$=3) MIMO with Linear Dispersion code given in (53). The base constellation is the 8-PSK. The bounds compared with the simulated bit-error rate at each stage of super-iterations. There are five super-iterations. In each super-iteration, there are **50 *internal LDPC decoding iterations***.

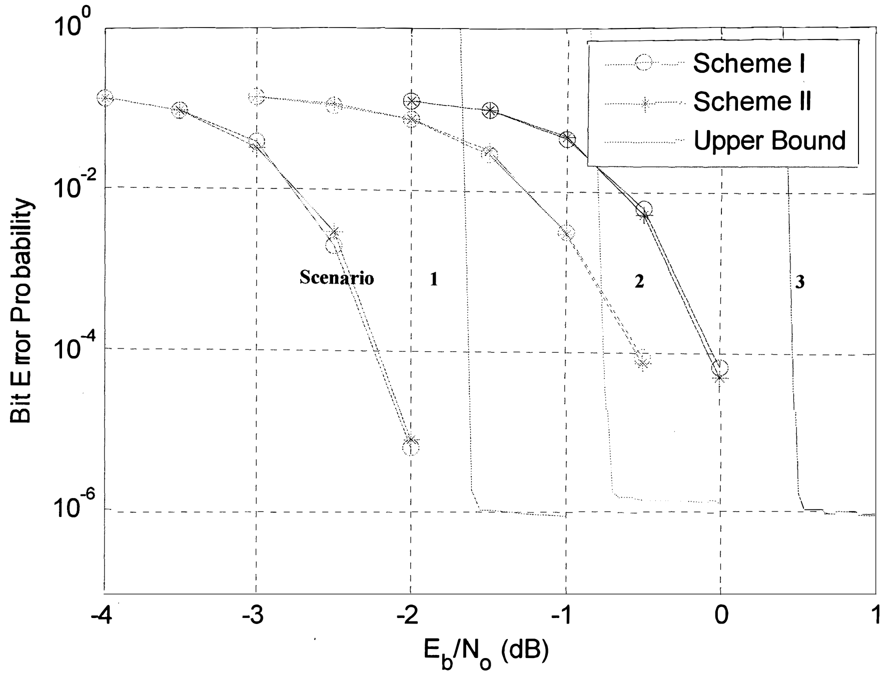



Figure 5: Comparison of simulation results with the union upper bounds for independent fading channels (10 LDPC decoding iterations and 3 Super Iterations)
Scenario 1: $N_t$=$N_r$ =3, 4PSK, LDPC(3000, 3, 6) + Linear Dispersion code in Eq. (54);
Scenario 2: $N_t$=$N_r$ =3, 8PSK, LDPC(2700, 3, 6) + Linear Dispersion code in Eq. (53);
Scenario 3: $N_t$=$N_r$ =2, 4PSK, LDPC(3000, 3, 6) + Alamouti Code.

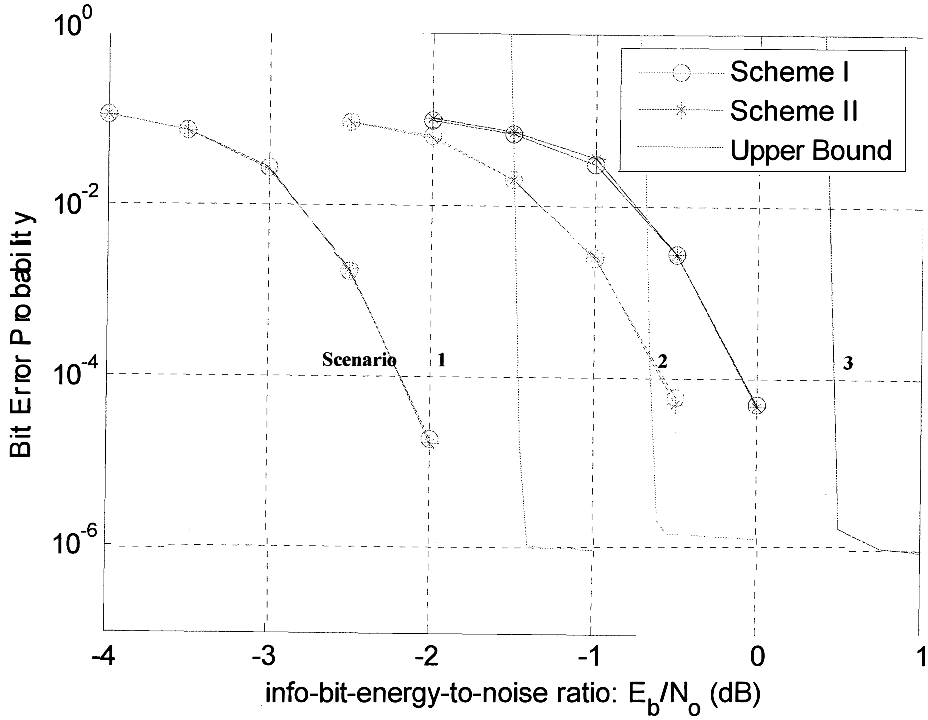



Figure 6: Comparison of simulation results with the union upper bounds for block fading channels (
LDPC decoding iterations and 3 Super Iterations).
Scenario 1: $N_t$=$N_r$=3, 4PSK, LDPC(3000, 3, 6) + Linear Dispersion code in Eq. (54);
Scenario 2: $N_t$=$N_r$=3, 8PSK, LDPC(2700, 3, 6) + Linear Dispersion code in Eq. (53);
Scenario 3: $N_t$=$N_r$=2, 4PSK, LDPC(3000, 3, 6) + Alamouti Code.

Table 1: $N_t$=2 with 4-PSK modulation, Gray mapping
(Scheme II with a mask vector [1 0 1 0])

| Binary string before & after masking | $(s_a, s_b)$ | $\mathbf{s}_{(j)}$ | Binary string before & after masker | $(s_a, s_b)$ | $\mathbf{s}_{(j)}$ |
|---|---|---|---|---|---|
| (0 0 0 0)→ (1 0 1 0) | (-j, -j) | $\begin{pmatrix} -j & -j \\ -j & j \end{pmatrix}$ | (1 0 0 0)→ (0 0 1 0) | (1, -j) | $\begin{pmatrix} 1 & -j \\ -j & 1 \end{pmatrix}$ |
| (0 0 0 1)→ (1 0 1 1) | (-j, -1) | $\begin{pmatrix} -j & 1 \\ -1 & j \end{pmatrix}$ | (1 0 0 1)→ (0 0 1 1) | (1, -1) | $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ |
| (0 0 1 0)→ (1 0 0 0) | (-j, 1) | $\begin{pmatrix} -j & -1 \\ 1 & j \end{pmatrix}$ | (1 0 1 0)→ (0 0 0 0) | (1, 1) | $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ |
| (0 0 1 1)→ (1 0 0 1) | ( -j, j) | $\begin{pmatrix} -j & j \\ j & j \end{pmatrix}$ | (1 0 1 1)→ (0 0 0 1) | ( 1, j) | $\begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix}$ |
| (0 1 0 0)→ (1 1 1 0) | (-1, -j) | $\begin{pmatrix} -1 & -j \\ -j & -1 \end{pmatrix}$ | (1 1 0 0)→ (0 1 1 0) | (j, -j) | $\begin{pmatrix} j & -j \\ -j & -j \end{pmatrix}$ |
| (0 1 0 1)→ (1 1 1 1) | (-1, -1) | $\begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}$ | (1 1 0 1)→ (0 1 1 1) | (j, -1) | $\begin{pmatrix} j & 1 \\ -1 & -j \end{pmatrix}$ |
| (0 1 1 0) → (1 1 0 0) | (-1, 1) | $\begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}$ | (1 1 1 0)→ (0 1 0 0) | (j, 1) | $\begin{pmatrix} j & -1 \\ 1 & -j \end{pmatrix}$ |
| (0 1 1 1) → (1 1 0 1) | (-1, j) | $\begin{pmatrix} -1 & j \\ j & -1 \end{pmatrix}$ | (1 1 1 1)→ (0 1 0 1) | (j, j) | $\begin{pmatrix} j & j \\ j & -j \end{pmatrix}$ |

References


[i] B.M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Commun.*, vol. 51, no. 3, pp. 389 – 399, Mar. 2003.

[ii] S. ten Brink, G. Kramer and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. on Commun.*, vol. 52, no. 4, pp. 670 – 678, Apr. 2004.

[iii] Heung-No Lee, "LDPC coded modulation MIMO OFDM Transceiver: Performance Comparison with MAP Equalization," *Proc. of IEEE VTC*, vol.2, pp. 1178-81, Jeju , Korea , Apr. 2003.

[iv] S. Y. Chung, G. D. Forney, Jr., T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, pp. 58–60, Feb. 2001.

[v] The WWiSE* web-site for an alliance of companies and entities developing a proposal for the IEEE 802.11n Wireless LAN standard at http://www.wwise.org/.

[vi] Jingqiao Zhang and Heung-No Lee, "An upper bound to the error performance of coded system over Rayleigh fading MIMO channels," To appear in *IEEE Commun. Letter.*

[vii] V. Tarokh, H. Jafarkhani and A.R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp.1456 – 1467, July 1999.

[viii] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41-59, 1996.

[ix] B. Hassibi; B.M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. on Info.Theory*, vol. 48, no. 7, pp. 1804-24, July 2002.

[x] L.C. Perez, J. Seghers and D.J. Costello, Jr., "A distance spectrum interpretation of turbo codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1698-1709, Nov. 1996.

[xi] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded ModulationWith Applications*. New York: McMillan, 1991.

[xii] D. Aktas and M. P. Fitz, "Computing the distance spectrum of space-time trellis codes," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Chicago, IL, Sept. 2000.

[xiii] D. Aktas and Fitz, M.P., "Distance spectrum analysis of space-time trellis-coded Modulations in quasi-static Rayleigh-fading channels," *IEEE Trans. on Info. Theory*, vol. 49, Issue 12, pp. 3335 –44, Dec. 2003.

[xiv] M. Byun, D. Park, and B. Lee, "Performance analysis of space–time trellis-coded modulation in quasi-static Rayleigh fading channels," in *Proc. IEEE Int. Conf. Communications*, vol. 3, New York, NY, Apr. 2002, pp. 1596–1600.

[xv] F. Behnamfar, F. Alajaji, and T. Linder, "Tight error bounds for space-time orthogonal block codes under slow Rayleigh flat fading," *IEEE Trans. on Commun.*, vol. 53, no. 6, pp.952-6, June, 2005.

[xvi] S. Litsyn and V. Shevelev, "Distance distributions in ensembles of irregular low-density parity-check codes", *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3140 – 3159, Dec. 2003.

[xvii] R. G. Gallager, *Low Density Parity Check Codes*, MIT Press, 1963.

[xviii] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Infor. Theory*, vol. 47, no. 2 , pp. 619 – 637, Feb. 2001.

[xix] Michael G. Luby, Michael Mitzenmacher, M. Amin Shokrollahi, and Daniel A. Spielman, "Improved Low Density Parity Check Codes Using Irregular Graphs," 47(2), pp. 585-598, Feb. 2001.

[xx] M.K. Simon and M. S. Alouini, *Digital communication over fading channels, a unified approach to performance analysis,* Wiley, 2000.

[xxi] Jingqiao Zhang and Heung-No Lee, "Performance Analysis on Coded System over Quasi-Static (MIMO) Fading Channels," To appear in *Proc. of ICC 2005.*

[xxii] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations," TDA Progress Report 42-122, August 15, 1995.

[xxiii] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: asymptotic distance distributions," *IEEE Trans. on Inform. Theory*, vol. 48, no. 4, pp.887 – 908, Apr. 2002.

[xxiv] B.M. Hochwald, and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Commun.*, vol. 51, no. 3, pp. 389 – 399, 2003.

[xxv] J. Zhang and Heung-No Lee, "A Performance Bound on Random-Coded MIMO Systems," *IEEE Commun. Letters*, vol 10, no.3, pp.168-170, March 2006.

[xxvi] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation", *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 657 – 670, Feb. 2001.

[xxvii] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. On Inform. Theory*, vol.42, no.2, pp. 409-28, Mar, 1996.

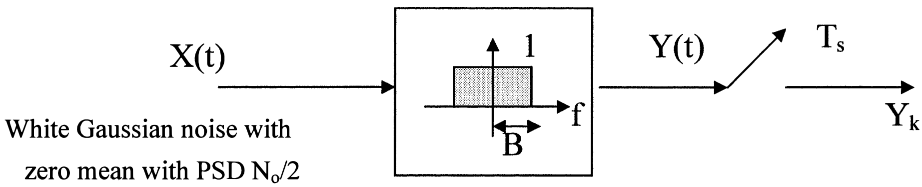
[xxviii] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol.16, no. 8, pp:1451 – 1458, Oct. 1998.

[xxix] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections and applications," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3029 - 3051, December 2002.

[xxx] S. Shamai and I. Sason, "Variations on the Gallager bounds with some applications," *Physica A Journal*, vol. 302, pp. 22-34, December 2001.

HW#2

P.1 (Noise power spectral density, what is it?) Power spectral density of filtered and sampled noise.




As shown above, a continuous random process X(t) is filtered and sampled to produce a discrete random process $Y_k$. X(t) is a white Gaussian process with zero mean and the power spectral density No/2. No is measured. It is $10^{-6}$ Watts/Hz. The filter is a brick wall whose bandwidth is B = 10 kHz. The sampling rate is 40 kilo samples per second. Do this problem twice, once for $T_s = 1/(2B)$ and the other for 1/(4B).

a. Obtain and sketch the PSD and the autocorrelation function of Y(t)

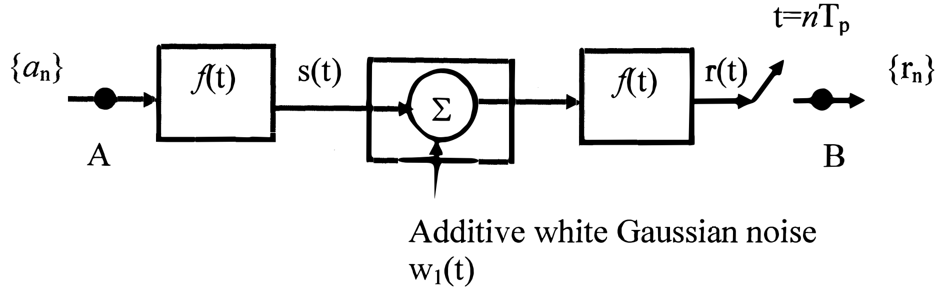
b. Obtain the PSD and sketch the autocorrelation function of $Y_k$.

P.2 (The Nyquist rate of a channel with bandwidth *W* Hz) First define the following terminology (in your own words and in one or two sentences maximum).

- Channel symbol
- Channel symbol alphabet
- Baud
- Symbol energy (average)
- Bit energy (average)
- Wide sense stationary random processes
- Cyclostationary random processes
- Baseband
- Passband
- Roll off factor and Square root raised cosine filter

Now, suppose a channel with *W* Hz bandwidth. What is the maximum baud that can be supported by the channel? What happens when a baud is chosen larger than the Nyquist rate 2*W*?

P.3 (Transmission over AWGN channel) Consider the following block diagram. Suppose the channel is brick wall filter with bandwidth *W* Hz. At the receiver, AWGN is added to the received signal. A transmit shaping filter *f*(t) is used at the transmitter to modulate the channel symbol sequence $\{a_n\}$. The same filter is used at the receiver to demodulate the signal. Let's suppose the shape of the filter is given by a square root raised cosine filter (see Proakis/Salehi) $f(t)$ whose roll off factor is 20%. The energy of the filter is 1.




(a). Suppose the channel-symbol sequence is {1 -1 -1  1  1  -1}. Sketch the corresponding signal s(t) for this sequence. For sketching, use the filter shape $f(t)$ which is truncated for four symbol periods around the origin. Ignore the noise for now.
(b). Do the same for the received signal r(t). Sketch the signal r(t).
(c). Now consider the sampled sequence $\{r_n = r(t = nT_p)\}$ at the receiver. While still ignoring the contribution of the noise, obtain the input-output relationship from point A to point B. At this point, comment on our choice of shaping filter $f$(t).
(d). Now let's include the AWGN, and re-obtain the input-output relationship from point A to point B (the discrete-time channel).
(e). $N_o$ has been measured. It is $10^{-6}$ Watts/Hz. Obtain the statistical characteristics of the noise sample at the output of the A/D sampler. Assume *wide-sense stationary* noise, and obtain the mean and the auto-correlation function of the noise samples.
(f). Now, let's suppose the channel-symbol $a_n$ is drawn randomly from an alphabet. Every symbol is independent from each other. Let $E\{a_n\} = 0$ and $E\{a_n^2\} = E_s$. Note $E$ is the avg. energy of channel symbol at the receiver. Now, find $E_s$ such that the ratio $E_s/N_0 = 30$dB.
(g). Now further suppose $a_n$ is Gaussian distributed with mean zero and variance $E_s$. Use the value you found in (f), and obtain the capacity of the discrete-time channel of (d). That is, how many bits per channel-use can be transferred to the receiver with very small P(e)?
(h). Suppose we have a channel with bandwidth $W = 10$ KHz. Obtain the signal-to-noise ratio. What is the maximum transmission rate [bits/sec] that is achievable over this channel?
(i). This time, let's suppose that the channel-symbol $a_n$ is not Gaussian distributed. Instead, it is a binary uniform random variable over the alphabet $\{+\sqrt{E}, -\sqrt{E}\}$. Use the same $E$ and $N_0$ obtained previously. What would be the consequence of making this change to the achievable transmission rate? More specifically, what would be the maximum transmission rate for this case when SNR is very large, say 100dB? Justify your answer.

P.4 (Matched filter) Suppose there is a signal whose waveform is given by

$$f(t) = \begin{cases} A_c \cos(2\pi f_c t), & 0 \le t \le T \\ 0, & o.w. \end{cases}.$$

We want to design an optimal receiver and sampler system which observes the signal under the additive white Gaussian noise (zero mean and PSD $\frac{N_o}{2}$) and makes a decision variable at time $t = 5.5T$. That is, the received signal is

$$y(t) = f(t) + z(t)$$

where z(t) is the AWGN noise.

(a). Determine the impulse response of the optimal filter used by the optimal receiver, and obtain the expression of your decision variable which achieves the maximum possible signal-to-noise ratio.

(b). Obtain the expression for the signal-to-noise (SNR) ratio of your decision variable and show that it attains the maximum possible SNR.

P.5 (P(e) calculation) Recall the Time Division example (Example 1) given in the lecture. Derive the probability of symbol error P(e) as a function of $E_s/N_o$. In this case, what is the relation between $E_b$ and $E_s$ where $E_s$ is the symbol energy while $E_b$ is the energy required to transmit a single information bit. You can use the Q-function (or the *erfc* function) as the final answer. Note there are four signals. Suppose we give a binary map to the four signals in the following way: 00, 01, 10, 11 for $x_1(t)$, $x_2(t)$, $x_3(t)$, and $x_4(t)$ respectively. Now for this map, derive the bit error probability $P_b$ expression as a function of $E_b/N_o$. Evaluate the two expressions in MATLAB and draw graphs. You may refer to Proakis/Salehi Figure 4.3-8.

P.6 (MATLAB simulation) Verify the symbol error probability and the bit error probability of P.5 in MATLAB simulation. For simulation, you need to obtain at least 100 errors for each simulation point. In each graph, show both theoretical and simulation results, and make sure your results match.

**HW#3**

For each MATLAB part, please turn in your MATLAB code. Reading the Shannon's 1948 paper will help. The 2nd lecture note here implies my note on "Information Theory and Gallager's Random Coding bounds."

P.1 (Geometric View of Signals and Noise) In this part, we design an 8-QAM system. (a). Design a rectangular-shaped 8-QAM constellation whose average energy of the constellation is 5. Give all x- and y-labels for each constellation point.

(b). What is the dimension of your signal constellation? That is, how many orthogonal signals for your signal set are needed? Obtain the set of basis signals and normalize them. Show how each signal in the constellation can be represented by your basis signals.

(c). Draw transmitter and receiver systems based on your normalized basis. Include the sampler and the signal detector (the decision maker) in your system.

(d). Consider an AWGN noise (zero mean and PSD $\frac{N_o}{2}$) present at the front end of the receiver. What is the variance and mean of the noise at the sampled output of your system?

(e). Obtain an upper and a lower bound on P(e) as tight as possible.

P.2 (M-PSK, M-QAM) Obtain P(e) for 8-PSK and simulate 8-PSK system for symbol error rate. Compare the simulation results with the theoretical P(e). Repeat for 8-QAM. Comment on your 8-PSK and 8-QAM results.

P.3 (Entropy calculation exercises)

a. Entropies of two random variables, i.e. H(X), H(Y), H(X|Y), H(Y|X): Do P6.5 Proakis/Salehi

P.4 (An example for Shannon's key idea) Consider a BSC with parameter $p$ which was considered in the class. Let $\mathbf{X}_1$ denote an input 1-0 binary vector of length $n$. Say this input vector has five 1s in it. Suppose we send this input vector over the BSC channel and obtain an output vector $\mathbf{Y}$. Use $p = 1/16$ and block length $n = 32$.

a. Describe the typical set for $\mathbf{Y}$ in your own words (Two or three sentences maximum).
b. Give the size of typical set in a. Compare it with the size of the complete set.
c. Suppose we have obtained a typical output $\mathbf{Y}$. Say this vector contains only two differences compared with the input $\mathbf{X}$. Now describe the typical set for the input $\mathbf{X}$ which could have generated the output $\mathbf{Y}$. How big is this typical set? Compare it with the complete set size for the input.
d. At this point, let's consider selecting another input word $\mathbf{X}_2$ randomly. Obtain the probability that a newly selected word $\mathbf{X}_2$ belongs to the typical input set obtained in c. Discuss the consequence of selecting $\mathbf{X}_2$ which belongs to the typical input set of c. In particular, describe the decoding error when both X1 and X2 are used in a codebook as codeword elements.

P.5 (Upper/lower bound on P(e)) Use P2.11 Proakis/Salehi in HW#1. Obtain the upper and lower bounds on P(e) for the constellation as function of $N_o$. Be specific to the distance calculations. Now do MATLAB simulation and verify the upper/lower bound expressions.

P.6 (Reading Shannon's 1948 paper)

1. Obtain the entropy of a Bernoulli random variable with parameter $p$. Draw the entropy as function of $p$ as $p$ is varied from 0 to 1. Use MATLAB.
2. Let X and Y be mutually independent binary random variables with the alphabet {0, 1}. Let X be a Bernoulli with parameter p = 0.1 (i.e., Pr{X = 1}), and Y be a Bernoulli with q

= 0.2 (i.e., Pr{Y=1}). Let Z = 0 when X = 0, 1 when X = 1 and Y = 0, and 2 when X = 1 and Y = 1. Prove/disprove H(Z) = H(X) + (1-p)H(Y).

P.7 (Shannon's Entropy) Read page 12 of Shannon's paper and design two binary random variables X and Y which give the following results.

a) The joint entropy H(X, Y) is smaller than the sum of individual entropies H(X) + H(Y).
b) The joint entropy H(X, Y) is equal to H(X) + H(Y).
c) The conditional entropy H(X|Y) is smaller than H(X).

P.8 (Equivocation and channel capacity) Provide definitions/comments for the following items (one or two sentences of your own at maximum)

a. Equivocation.
b. Relation of equivocation to conditional entropy
c. Relation of equivocation to channel capacity
d. Comment on Theorem 11. Why does it make sense?

P.9 (Capacity of a channel) Suppose there is a channel which transfers three letters *a*, *b* and *c*. The channel transfers the letter *a* without error. The channel however makes errors while transferring *b* and *c*. With probability *p*, it changes *b* to *c*; with equal probability, *c* to *b*. Find the capacity of this channel with respect to the given parameter *p*.

HW#4

P.1 Proakis/Salehi P6.42 (Mutual Information for Gaussian Noise Channel)

P.2 Proakis/Salehi P6.43 (Capacity of 3 input 2 output channel, non symmetric case)

P.3 (Derivation of capacity for the erasure channel) Find the capacity of the erasure channel with parameter p. This time, use I(X; Y) = H(Y) – H(Y|X) to prove that the capacity is 1 – p.

P.4 (Proakis/Salehi Section 6.5-2 Channel Capacity) Read Section 6.5-2. Suppose there is a channel which can be described by Y = X + E mod 2, where E is Bernoulli with parameter $p$. X is Bernoulli with parameter $q$. Let's use $q = 0.5$ and $p = 0.01$.
(a). Find H(Y|X).
(b). Suppose using the same channel $n = 1000$ times. Suppose that the input vector **X** is the all-zero word. Describe the set of typical output sequences in English (Three sentences maximum).
(c). What is the size of the typical set Proakis/Salehi talk about in Section 6.5-2?
(d). Use the Stirling's approximation and show there is a nice relationship between H(Y|X) and the size of typical set.

P.5 (Proakis/Salehi P6.48, Coding vs. No Coding)

P.6 (Proakis/Salehi P6.58, Capacity)

P.7 (Proakis/Salehi P6.59, Rate over BSC with p)

HW#5

P.5 (P(e) for $M$-ary orthogonal signals) Consider P(e) for $M$-ary orthogonal signals over infinite bandwidth AWGN channel, and show that P(e) can be upper bounded by

$$P(e) \le \exp\left(-k\left(\frac{E_b}{2N_o} - \ln(2)\right)\right),$$

where $k = \log_2(M)$. Draw the upper bounds as $M$ is increased, once with respect to $E_s/N_o$ and once more with respect to $E_b/N_o$. Explain your results. According to this bound, what is the minimum required $E_b/N_o$ for reliable communications?

P.6 Prove that (4) in lecture note is minimum of (1) at s = 1/(1+ρ) (of the 2nd lecture note). Use the Holder's inequality given in (2).

P.7 (Derivative of the exponent) Find the expression of the derivative of $E_o(\rho, Q)$ with respect to $\rho$ at $\rho = 0$ (Read Chapter 5 Gallager)

P.8 (Gallager bound for binary input AWGN channel) Let's consider a binary input {+1, -1} AWGN channel. The PSD of the noise is $N_o/2$.

(a) Obtain a tight union bound, similar to eq. (4) of the 2nd lecture note, for this channel.

(b) Obtain the error exponent expression of this channel $E_o(\rho, Q)$. What type of Q is desirable? Why?

(c) Obtain the expression for error exponent E(R) and sketch it.