Digital Communications Fundamentals

Carrson C. Fung

Dept. of Electronics Engineering

National Chiao Tung University

Signal Retrieval and Communication

- Theory of systems for the conveyance of information
- Characteristics of communication systems
  - Uncertainty
    - Noise and "information" (deterministic vs. probabilistic)
  - Keep in mind: Signal retrieval problem
    - Communication (only particular type of signal retrieval problem)
  - Optimal design is crucial
    - Many "optimal" designs are not optimal – depends on objective
  - How do we do it? (We are engineers, this is important!)
    - Statistical signal detection and estimation theory
      - Weiner optimum filter, matched filter, adaptive filter, and many more…
    - Information theory and coding
      - Shannon says it can be done, but didn't tell us how: block, iterative coding, …
- Usually two resources to consider
  - Bandwidth vs. Power

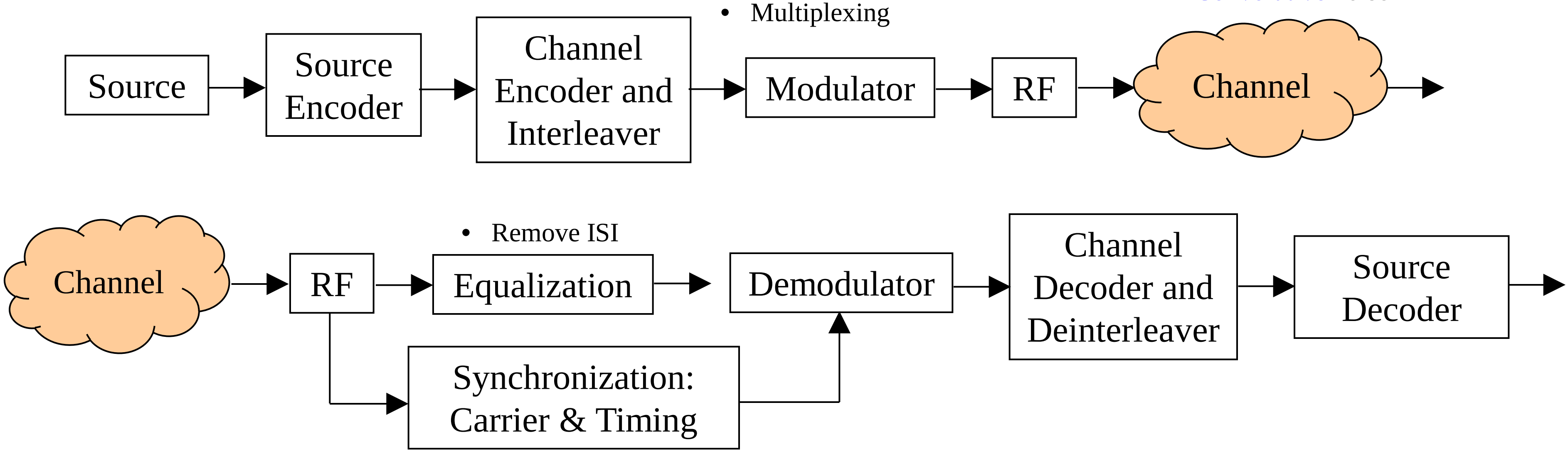
Block Diagram of a Narrowband Digital Communication System

- Source coding: remove redundancy → increase efficiency
- Channel coding: increase redundancy → protect information

- Ease of radiation
- Reduce noise and interference
- Increase BW efficiency: R_b/B
- Channel assignment
- Multiplexing

- Internal and external additive noise
- Convolutive noise

- Remove ISI



- Carrier: Coherent modulation requires carrier
- Timing: Need to know when to sample to recover digital signal

Keep in mind that this is only a **model**!

Can we make it simpler? More complicated? Consequences?

Bandwidth and Power Efficiency

- Channel bandwidth and transmit power are two primary communication resources and have to be used as efficient as possible
  - Spectrum utilization efficiency (bandwidth efficiency)
    - Measured by the achievable data rate per unit bandwidth R_b/B
  - Power utilization efficiency (energy efficiency)
    - Measured by the required E_b/N_0 to achieve a certain bit error probability
- It is always desirable to maximize bandwidth at a minimal required E_b/N_0
  - However, in certain scenario, such as space communications, it is important to achieve high energy efficiency as bandwidth is abundant, but power is scarce
- Discussion shall be restricted to uncoded system

M-ary Signaling for Bandwidth- vs. Power-Limited System

- Bandwidth-limited system
  - Spectrally-efficient modulation techniques can be used to save bandwidth at the expense of power, i.e. E_b/N_0, e.g. MPSK
- Power-limited system
  - Power-efficient modulation techniques can be used to save power at the expense of bandwidth, e.g. MFSK
- A symbol in an M-ary alphabet is related to a unique sequence of k bits: $M = 2^k$ ➔ $k = \log_2 M$, M is the alphabet size
- Symbol refers to the member of the M-ary alphabet that is transmitted during each symbol duration T_s
- Symbols are then mapped to a voltage of waveform
- Example: $M = 16$: 1 0 1 1 1 0 0 1 ➔ (1011), (1001) 🡐 M-tuple
  - Each M-tuple is a symbol with length T_s

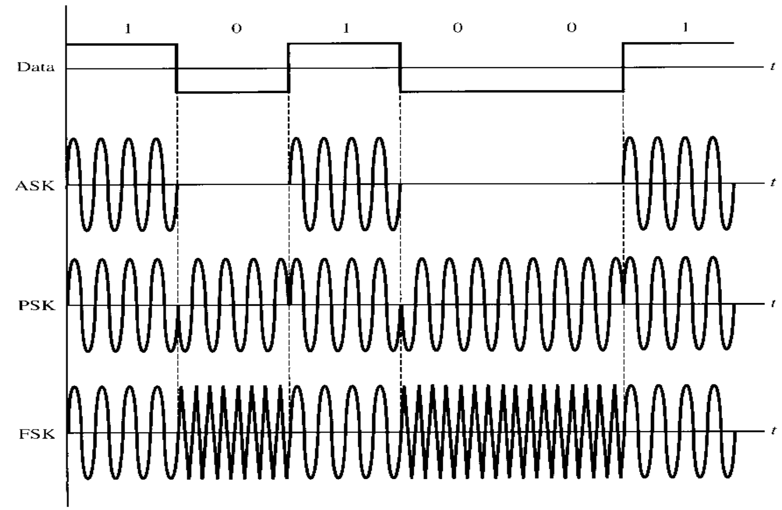


Figure 4.19
Examples of digital modulation schemes.

ASK* (Analogous to AM):

$$x_{ASK}(t) = A_c\left[1 + d(t)\right]\cos(2\pi f_c t)$$

PSK* (Analogous to PM):

$$x_{PSK}(t) = A_c \cos\left(2\pi f_c t + \frac{\pi}{2} d(t)\right)$$

FSK* (Analogous to FM):

$$x_{FSK}(t) = A_c \cos\left(2\pi f_c t + k_f \int^t d(\alpha)\, d\alpha\right)$$

* $d(t)$ is a line code, e.g. NRZ

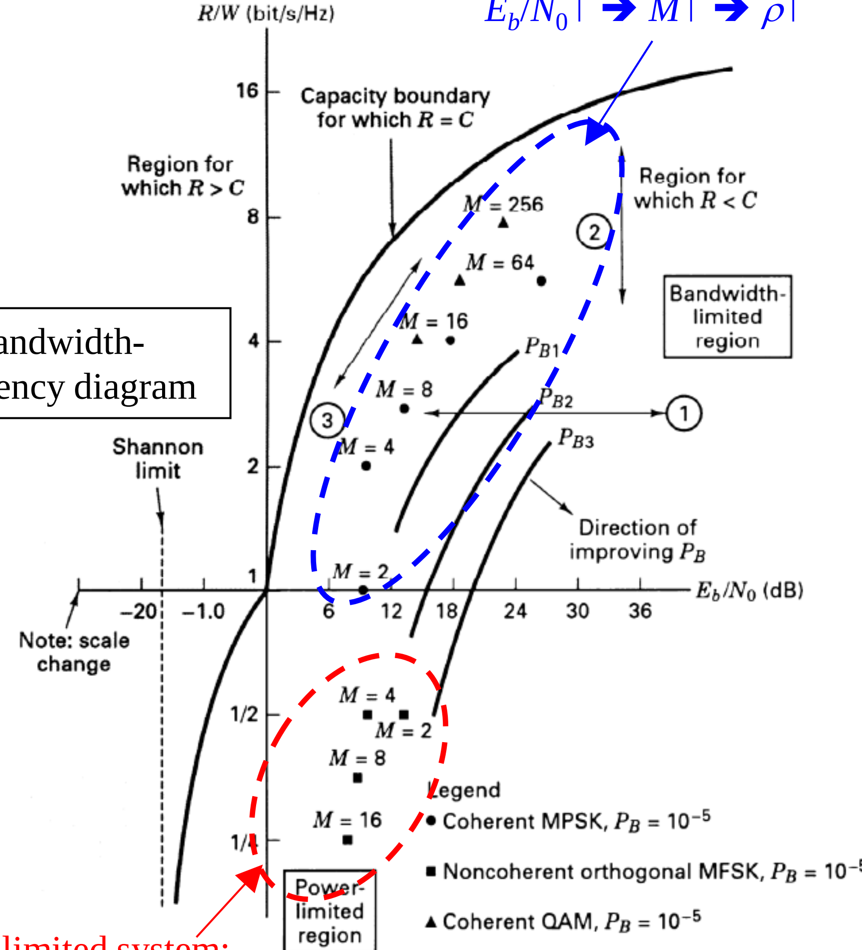
M-ary Signaling for Bandwidth- and Power-Limited System

- Channel bandwidth required to pass M-ary signals (symbols) is $B = \frac{2}{T_s}$
- Note that symbol duration $T_s = T_b \log_2 M$
  - T_b: bit duration
- Also date rate $R_b = (\log_2 M)/T_s = (\log_2 M)/(T_b \log_2 M) = 1/T_b$
- Hence $B = \frac{2}{T_b \log_2 M} = \frac{2R_b}{\log_2 M}$
- And bandwidth efficiency

$$\rho \triangleq \frac{R_b}{B} = \frac{\log_2 M}{2} = \frac{\log_2 M}{BT_s} \text{bits/s/Hz}$$

BW-limited system: Fixed P_B and B, $E_b/N_0\uparrow \rightarrow M\uparrow \rightarrow \rho\uparrow$

Bandwidth-efficiency diagram



Pwr-limited system: Fixed P_B, $B\uparrow \rightarrow \rho\downarrow \rightarrow E_b/N_0\downarrow$

Signal-Space Analysis

Block diagram of a generic digital communication system

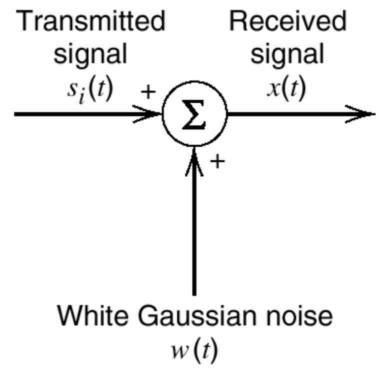
- Source symbols m_i from alphabet of M symbols denoted as m_1, m_2, …, m_M
- $p_i \triangleq Pr(m_i) = \frac{1}{M}$, for $i = 1, 2, \ldots, M$.
- Transmitter codes m_i into a distinct signal $s_i(t)$ suitable for transmission over channel
- $s_i(t)$ occupies for T duration and has finite energy

$$E_i = \int_0^T s_i^2(t)$$

Signal-Space Analysis

- Assumptions
  - Channel is linear and bandwidth is wide enough to accommodate the transmit signal $s_i(t)$ with little or no distortion
  - Channel noise, $w(t)$, is sample function of a zero mean white Gaussian noise process ➔ makes receiver calculation tractable
- Then, the channel is referred to as *additive white Gaussian noise* (AWGN) channel, where the output is modeled as

$$x(t) = s_i(t) + w(t), \quad \begin{cases} 0 \le t \le T, \\ i = 1, 2, \ldots, M \end{cases}$$



Geometric Representation of Signals

- Represent any set of energy signal $\{s_i(t)\}$ as a linear combination of N orthonormal basis functions, where $N \leq M$, i.e.

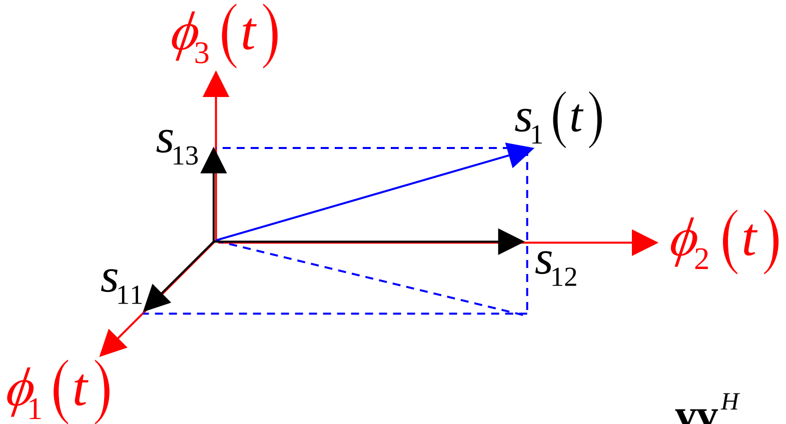
$$s_i(t) = \sum_{j=1}^{N} s_{ij}\phi_j(t), \quad \begin{cases} 0 \leq t \leq T, \\ i = 1, 2, \ldots, M \end{cases},$$

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)dt, \quad \begin{cases} i = 1, 2, \ldots, M, \\ j = 1, 2, \ldots, N \end{cases}$$

- The real-valued basis functions $\phi_1(t)$, $\phi_2(t)$, …, , $\phi_N(t)$, are orthonormal, i.e.

$$\int_0^T \phi_i(t)\phi_j(t)dt = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases},$$

where δ_{ij} is the Kronecker delta function



Recall the projection matrix: $\mathbf{P} = \dfrac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}$.

If $\mathbf{v}$ is orthonormal(rewrite as $\mathbf{q}_i$), then

$$\mathbf{a} = \underbrace{\mathbf{q}_1 \underbrace{\mathbf{q}_1^H\mathbf{a}}_{\text{analysis}}}_{\text{synthesis}} + \underbrace{\mathbf{q}_2 \underbrace{\mathbf{q}_2^H\mathbf{a}}_{\text{analysis}}}_{\text{synthesis}} + \underbrace{\mathbf{q}_3 \underbrace{\mathbf{q}_3^H\mathbf{a}}_{\text{analysis}}}_{\text{synthesis}}$$

$\{s_{ij}\}_{j=1}^{N}$ may naturally be viewed as an N-dimensional *signal vector*, denoted by $\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}$, for $i = 1, 2, \ldots, M$

Example: Binary Phase-Shift Keying (BPSK)

- Pair of signals used to represent binary symbols 1 and 0

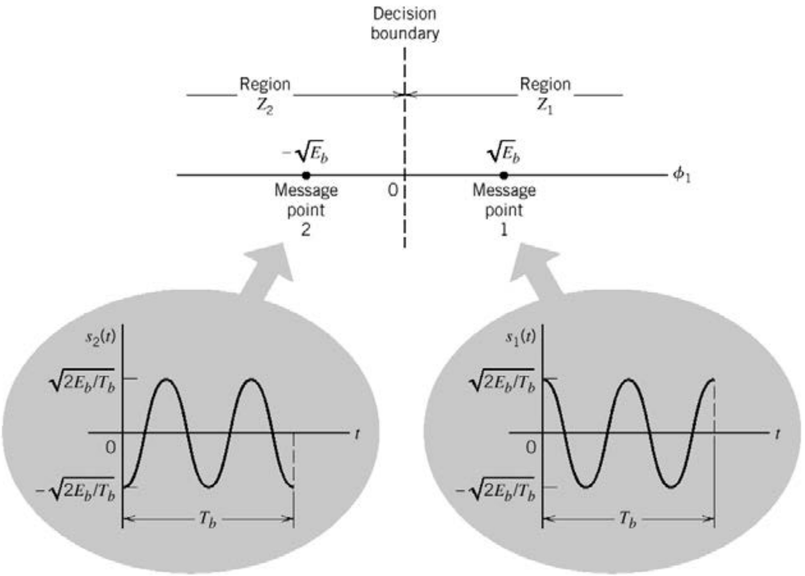
$$s_1(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t)$$
$$s_2(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t)$$

- Note the orthonormal basis functions is

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t), \text{ for } 0 \le t < T_b$$

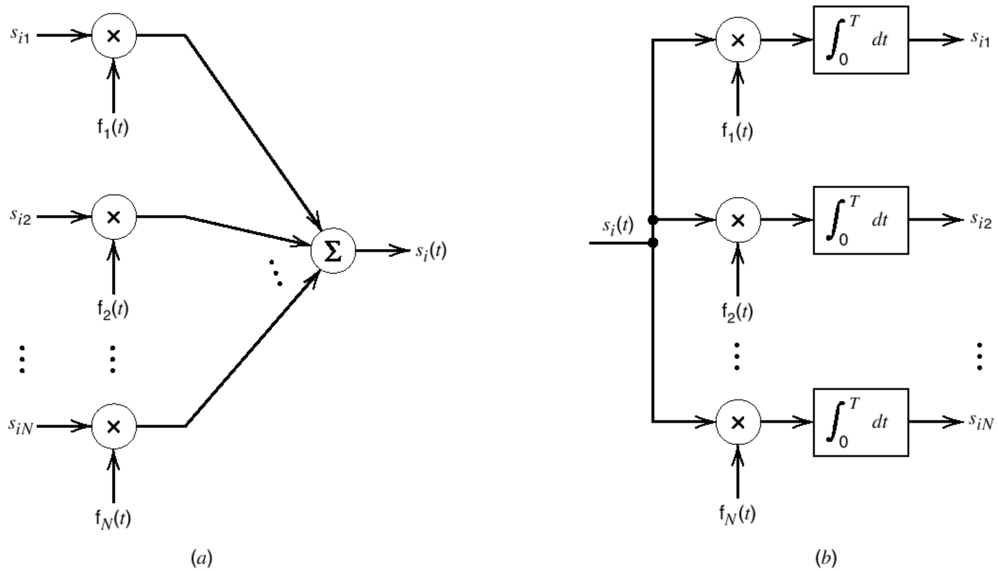
$$\Longrightarrow \begin{cases} s_1(t) = \sqrt{E_b}\phi_1(t), & 0 \le t < T_b \\ s_2(t) = -\sqrt{E_b}\phi_1(t), & 0 \le t < T_b \end{cases}$$

$$\Longrightarrow \begin{cases} s_{11} = \int_0^{T_b} s_1(t)\phi_1(t)dt = +\sqrt{E_b} \\ s_{21} = \int_0^{T_b} s_2(t)\phi_1(t)dt = -\sqrt{E_b} \end{cases}$$

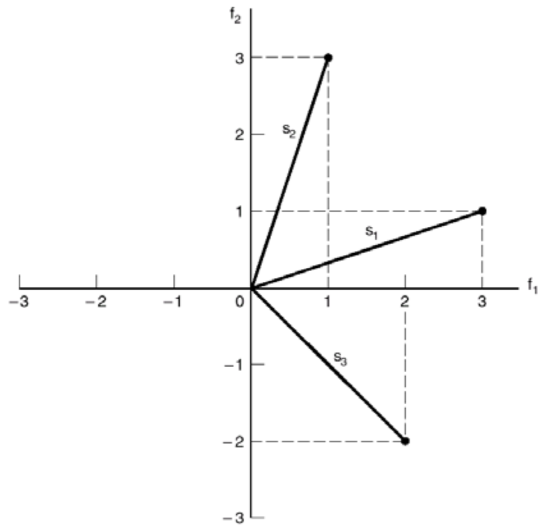


Signal-space diagram for coherent binary PSK system. The waveforms depicting the transmitted signals $s_1(t)$ and $s_2(t)$, displayed in the inserts, assume $n_c = 2$.

Geometric Representation of Signals



(a) Synthesizer for generating the signal $s_i(t)$. (b) Analyzer for generating the set of signal vectors $\{s_i\}$.



Illustrating the geometric representation of signals. $N = 2$, $M = 3$.

- Given N elements of the vectors $\mathbf{s}_i$, (i.e. $s_{i1}, s_{i2}, \ldots, s_{iN}$) operating as input, can use the synthesizer shown to generate $s_i(t)$
- Given $s_i(t)$, $i = 1, 2, \ldots, M$, can use the analyzer in (b) to obtain $s_{i1}, s_{i2}, \ldots, s_{iN}$. This consists of a bank of N product-integrators or correlators

Geometric Representation of Signals

- Induced norm of $\mathbf{s}_i$: $\|\mathbf{s}_i\| = \left(\mathbf{s}_i^T \mathbf{s}_i\right)^{1/2} = \left(\sum_{j=1}^{N} s_{ij}^2\right)^{1/2}, \quad fori = 1, 2, \ldots, M$
- Energy of $s_i(t)$ can be computed as

$$E_i = \int_f^T s_i^2(t)dt = \int_0^T \left[\sum_{j=1}^{N} s_{ij}\phi_j(t)\right]\left[\sum_{k=1}^{N} s_{ik}\phi_k(t)\right] dt$$
$$= \sum_{j=1}^{N}\sum_{k=1}^{N} s_{ij}s_{ik}\int_0^T \phi_j(t)\phi_k(t)dt$$
$$= \sum_{j=1}^{N}\sum_{k=1}^{N} s_{ij}s_{ik}\delta_{jk} = \sum_{j=1}^{N} s_{ij}^2 = \|\mathbf{s}_i\|^2$$

- Inner product: $\langle \mathbf{s}_i, \mathbf{s}_k \rangle \triangleq \int_0^T s_i(t)\mathbf{s}_k(t)dt = \mathbf{s}_i^T \mathbf{s}_k$
- Euclidean distance: $\|\mathbf{s}_i - \mathbf{s}_k\|^2 = \sum_{j=1}^{N} (s_{ij} - s_{kj})^2 = \int_0^T (s_i(t) - s_k(t))^2 dt$
- Cosine of angle θ_{ik}: $\cos\theta_{ik} = \frac{\mathbf{s}_i^T \mathbf{s}_k}{\|\mathbf{s}_i\|\|\mathbf{s}_k\|}$

Gram-Schmidt Orthogonalization/ Orthonormalization

Step 1: Orthonormalize $s_1(t)$ to obtain $\phi_1(t)$

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$$

Step 2: Project $s_2(t)$ onto the space spanned by $\phi_1(t)$, then subtract from $s_2(t)$ and normalize. The result will be orthonormal to $\phi_1(t)$

$$\text{Projection: } s_{21} = \langle \phi_1(t), s_2(t) \rangle = \int_0^T s_2(t)\phi_1(t)dt$$

$$\text{Subtraction: } g_2(t) = s_2(t) - s_{21}\phi_1(t)$$

$$\text{Normalization: } \phi_2(t) = \frac{g_2(t)}{\|g_2(t)\|} = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)dt}} = \frac{s_2(t) - s_{21}\phi_1(t)}{\sqrt{\int_0^T [s_2(t) - s_{21}\phi_1(t)]\,[s_2(t) - s_{21}\phi_1(t)]\,dt}}$$

$$= \frac{s_2(t) - s_{21}\phi_1(t)}{\sqrt{E_2 - s_{21}^2}} \quad (\text{cross terms of denominator} = -2s_{21}^2)$$

Gram-Schmidt Orthogonalization/ Orthonormalization

- In general, $\phi_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t)dt}}$, for $i = 1, 2, \ldots, N$
- Remarks
  - The signals $s_1(t)$, $s_2(t)$, …, $s_M(t)$ form a linearly independent set, i.e. $k_1 s_1(t) + \ldots + k_M s_M(t) = 0$ iff $k_1, \ldots, k_M$ equal 0. In that case, $N = M$
  - The signals $s_1(t)$, $s_2(t)$, …, $s_M(t)$ do not form a linearly independent set, then $N < M$, and $g_i(t) = 0$, for $i > N$

Statistical Characterization of Correlator Outputs

- Denote X_j as the random variable whose sample value is represented by the correlator output x_j, for $j = 1, 2, \ldots, N$
- From AWGN channel model, $X(t)$ is a Gaussian process (since $W(t)$ is AWGN)

$$\begin{aligned} \mu_{X_j} &= E\left[X_j\right] = E\left[s_{ij} + W_j\right] \\ &= s_{ij} + E\left[W_j\right] = s_{ij} \\ \sigma^2_{X_j} &= var\left[X_j\right] = E\left[\left(X_j - s_{ij}\right)^2\right] \\ &= E\left[W_j^2\right], \end{aligned}$$

$$\text{with } W_j \triangleq \int_0^T W(t)\phi_j(t)dt$$

$$\Rightarrow$$

$$\begin{aligned} \sigma^2_{X_j} &= E\left[\int_0^T W(t)\phi_j(t)dt \int_0^T W(u)\phi_j(u)du\right] \\ &= E\left[\int_0^T \int_0^T \phi_j(t)\phi_j(u)W(t)W(u)dtdu\right] \\ &= \int_0^T \int_0^T \phi_j(t)\phi_j(u)E[W(t)W(u)]dtdu \\ &= \int_0^T \int_0^T \phi_j(t)\phi_j(u)R_W(t,u)dtdu \\ &= \frac{N_0}{2}\int_0^T \int_0^T \phi_j(t)\phi_j(u)\delta(t-u)dtdu \\ &= \frac{N_0}{2}\int_0^T \phi_j^2(t)dt = \frac{N_0}{2}, \ \forall j \end{aligned}$$

where the expression for $R_W(t,u)$ is obtained as the noise is assumed to be WSS and has a constant PSD $N_0/2$. The last equality is obtained because $\phi_j(t)$ is orthonormal

Statistical Characterization of Correlator Outputs

- X_j are mutually uncorrelated because $\phi_j(t)$ form an orthogonal set:

$$
\begin{aligned}
cov\left[X_j X_k\right] &= E\left[\left(X_j - \mu_{X_j}\right)\left(X_k - \mu_{X_k}\right)\right] \\
&= E\left[\left(X_j - s_{ij}\right)\left(X_k - s_{ik}\right)\right] \\
&= E\left[W_j W_k\right] \\
&= E\left[\int_0^T W(t)\phi_j(t)dt \int_0^T W(u)\phi_k(u)du\right] \\
&= \int_0^T \int_0^T \phi_j(t)\phi_k(u)R_W(t,u)dtdu \\
&= \frac{N_0}{2}\int_0^T \int_0^T \phi_j(t)\phi_k(u)\delta(t-u)dtdu \\
&= \frac{N_0}{2}\int_0^T \phi_j(t)\phi_k(t)dt = 0, \quad \text{for } j \neq k
\end{aligned}
$$

- Since X_j is Gaussian r.v., hence, they are also statistically independent

Statistical Characterization of Correlator Outputs

$$\mathbf{X} \triangleq \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$$

elements are independent Gaussian rv's with mean s_{ij} and variance $N_0/2$

- Hence, the conditional pdf of X, given that $s_i(t)$ (or corresponding m_i) was transmitted, can be expressed as the product of the conditional probability density functions of its individual elements

$$f_X(\mathbf{x}|m_i) = \prod_{j=1}^{N} f_{X_j}(x_j|m_i)\,, \text{ for } i = 1, 2, \ldots, M$$

x and x_j are samples values of **X** and X_j

- Channel that satisfies the above equation is called memoryless channel

Statistical Characterization of Correlator Outputs

Since X_j is Gaussian rv with mean s_{ij} and variance $N_0/2$, we have

$$f_{X_j}(x_j|m_i) = \frac{1}{(\pi N_0)^{1/2}} \exp\left[-\frac{1}{N_0}(x_j - s_{ij})^2\right], \quad \text{for} \quad \begin{array}{l} j = 1, \ldots, N, \\ i = 1, \ldots, M \end{array}$$

So

$$f_{\mathbf{X}}(\mathbf{x}|m_i) = (\pi N_0)^{-N/2} \exp\left[-\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2\right], \text{for } i = 1, \ldots, M$$

Note that the AWGN channel is equivalent to an N-dimensional vector channel modeled as

$$\mathbf{x} = \mathbf{s}_i + \mathbf{w}, \quad \text{for } i = 1, 2, \ldots, M$$

Why do we care about this?

In the context of AWGN channel, the optimal receiver is the ML detector

Design objective: Expected cost $E(C)$ or Bayes' Risk R

A generalization of the minimum P_e criterion assigns costs to each type of error. Let C_{ij} be the cost if we decide H_i but H_j is true. The expected cost or Bayes risk is

$$R \triangleq E(C) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij} P\left(H_i \middle| H_j\right) P\left(H_j\right)$$

$$= C_{00} P\left(H_0 \middle| H_0\right) P\left(H_0\right) + C_{11} P\left(H_1 \middle| H_1\right) P\left(H_1\right)$$

$$+ C_{10} P\left(H_1 \middle| H_0\right) P\left(H_0\right) + C_{01} P\left(H_0 \middle| H_1\right) P\left(H_1\right)$$

If $C_{00} = C_{11} = 0,\ C_{10} = C_{01} = 1,$ then $R = P_e$.

Bayes' Risk

Let $R_1 = \{\mathbf{x} : \text{decide } H_1\}$ be the critical region and $R_0 = \{\mathbf{x} : \text{decide } H_0\}$

$$R = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij} P\left(H_i \middle| H_j\right) P\left(H_j\right)$$

$$= C_{00} P\left(H_0 \middle| H_0\right) P\left(H_0\right) + C_{11} P\left(H_1 \middle| H_1\right) P\left(H_1\right)$$

$$+ C_{10} P\left(H_1 \middle| H_0\right) P\left(H_0\right) + C_{01} P\left(H_0 \middle| H_1\right) P\left(H_1\right)$$

$$= C_{00} P\left(H_0\right) \int_{R_0} p\left(\mathbf{x} \middle| H_0\right) d\mathbf{x} + C_{01} P\left(H_1\right) \int_{R_0} p\left(\mathbf{x} \middle| H_1\right) d\mathbf{x}$$

$$+ C_{10} P\left(H_0\right) \int_{R_1} p\left(\mathbf{x} \middle| H_0\right) d\mathbf{x} + C_{11} P\left(H_1\right) \int_{R_1} p\left(\mathbf{x} \middle| H_1\right) d\mathbf{x}.$$

Result: MAP Detector

Assume that $C_{10} > C_{00}$, and $C_{01} > C_{11}$, the detector which minimizes the Bayes risk is to decide H_1 if

$$\frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} > \frac{(C_{10} - C_{00})P(H_0)}{(C_{01} - C_{11})P(H_1)} = \gamma$$

Since $p(\mathbf{x}|H_i)P(H_i) \propto p(H_i|\mathbf{x}) \Rightarrow$ the optimal MAP detector is

decide H_i if $\quad (C_{01} - C_{11})\, p(H_1|\mathbf{x}) > (C_{10} - C_{00})\, p(H_0|\mathbf{x})$

ML Detector

$$P_e = \Pr\{\text{decide } H_0, H_1 \text{ true}\} + \Pr\{\text{decide } H_1, H_0 \text{ true}\}$$

$$= P(H_0|H_1)P(H_1) + P(H_1|H_0)P(H_0)$$

If $P(H_0) = P(H_1) = p_i$

$$\Rightarrow R = P_e = p_i\left[P(H_0|H_1) + P(H_1|H_0)\right]$$

So the detector that minimizes the P_e is the optimal ML detector

Decide H_1 if

$$\frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} >= \gamma = 1$$

or decide H_1 if $\quad p(\mathbf{x}|H_1) > p(\mathbf{x}|H_0)$

Example

We have the detection problem

$$H_0 : x[n] = w[n], \qquad n = 0,1,\ldots,N-1$$

$$H_1 : x[n] = A + w[n], \quad n = 0,1,\ldots,N-1,$$

where $A > 0$ and $w[n]$ is WGN with variance σ^2. Assuming $p(H_0)$ $= p(H_1) = 1/2 \Rightarrow \gamma = 1$

Decide H_1 if

$$\frac{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > 1$$

Example

Taking log

$$-\frac{1}{2\sigma^2}\left(-2A\sum_{n=0}^{N-1}x[n]+NA^2\right)>0,$$

or we decide H_1 if $\overline{x} > A/2$.

This has the same form as the NP criterion except for the threshold.

To determine the P_e, note that

$$\overline{x} \sim \begin{cases} N\left(0,\dfrac{\sigma^2}{N}\right), & \text{conditioned on } H_0 \\ N\left(A,\dfrac{\sigma^2}{N}\right), & \text{conditioned on } H_1 \end{cases}$$

Example

What happens when $A = 1$ and $N = 1$

$$H_0 : x[0] = w[0], \qquad n = 0,1,\ldots,N-1$$

$$H_1 : x[0] = 1 + w[0], \quad n = 0,1,\ldots,N-1$$

$\Rightarrow$ Decide H_1 if $x[0] > 1/2$

$$x[0] \sim \begin{cases} N\left(0,\sigma^2\right), & \text{conditioned on } H_0 \\ N\left(1,\sigma^2\right), & \text{conditioned on } H_1 \end{cases}$$

Likelihood Functions

- The likelihood and log-likelihood (LL) function are defined as

$$L(m_i) \triangleq f_{\mathbf{X}}(\mathbf{x}|m_i), \quad \text{for } i = 1, 2, \ldots, M$$
$$l(m_i) \triangleq \log L(m_i), \quad \text{for } i = 1, 2, \ldots, M$$

- The LL function for an AWGN channel is

$$l(m_i) = -\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2, \quad i = 1, 2, \ldots, M$$

where the constant term $-(N/2)\log(N_0)$ is ignored

- Receiver will use the LL function to detect the presence of the transmitted symbol

Coherent Detection of Signals in Noise: MAP Decoding

- Signal detection problem
  - Given the observation vector **x**, perform a mapping from **x** to an estimate $\hat{m}$ of the transmitted symbol, m_i, in a way that would minimize the probability of error in the decision making process
- Decision making criterion: minimize probability of error: $P_e(m_i|x) = Pr(m_i \text{ not sent}|\mathbf{x}) = 1 - Pr(m_i \text{ sent}|\mathbf{x})$
- Can be shown that the optimum decision rule is

$$\text{Set } \widehat{m} = m_i \text{ if}$$

$$Pr(m_i \text{ sent}|\mathbf{x}) \geq Pr(m_k \text{ sent}|\mathbf{x}), \forall k \neq i$$

for $k = 1, \ldots, M$. This is known as maximum a posteriori probability (MAP) rule

Coherent Detection of Signals in Noise: ML Decoding

- Applying Bayes' rule, the MAP decision rule becomes

$$\text{Set } \widehat{m} = m_i \text{ if}$$

$$\frac{p_k f_{\mathbf{X}}(\mathbf{x}|m_k)}{f_{\mathbf{X}}(\mathbf{x})} \text{ is maximum for } k = i$$

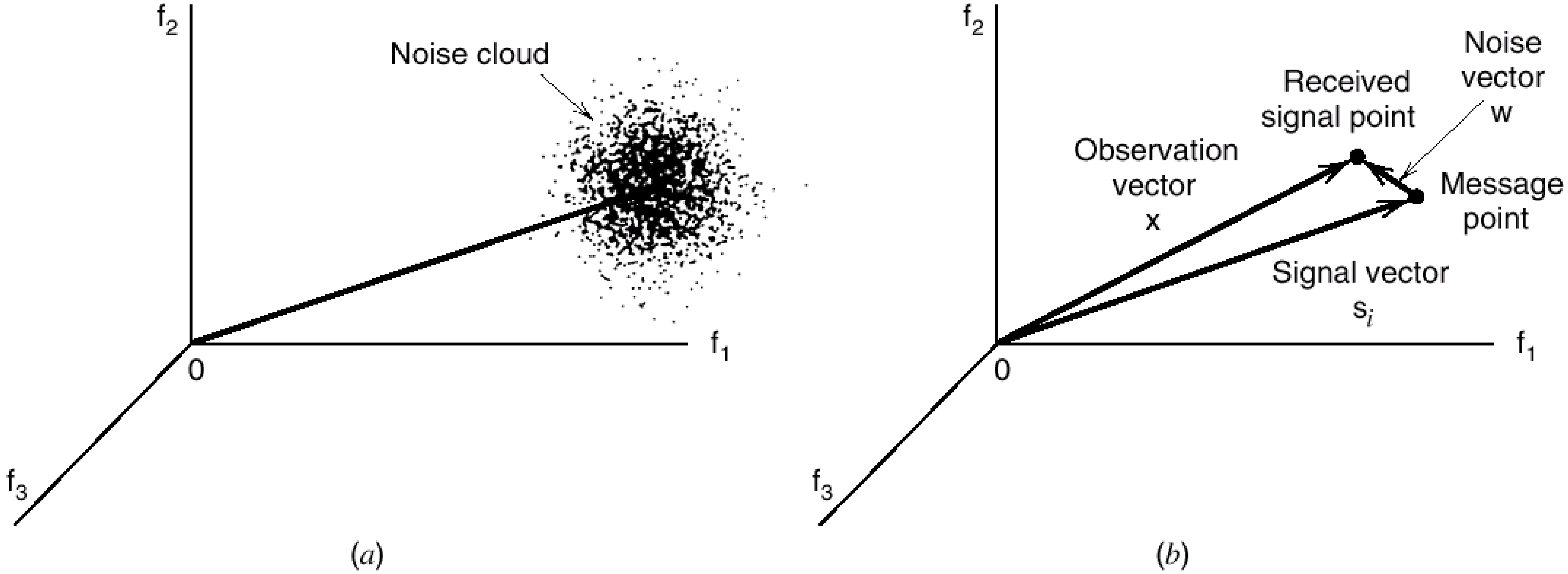
  where p_k is the a priori probability of transmitting symbol m_k
- Assuming $p_k = p_i$, for all i, the decision rule becomes the maximum likelihood rule

$$\text{Set } \widehat{m} = m_i \text{ if}$$

$$l(m_k) \text{ is maximum for } k = i$$

- The ML decoder can be used at the receiver to decode the transmitting symbol
- Note that the ML rule applies for all additive noise
  - No assumption about its statistical property is made

Graphical Illustration of Signal Constellation



Illustrating the effect of noise perturbation, depicted in (*a*) on the location of the received signal point, depicted in (*b*).

Graphical Interpretation of ML Rule

- Let Z denote the N-dimensional space of all possible observation vectors $\mathbf{x}$. This is known as observation space. Since the ML rule says $\widehat{m} = m_i$, where $i = 1, \ldots, M$, the Z is partitioned into M-decision regions, $Z_1, Z_2, .., Z_M$.
- ML rule can be restated as

Observation vector $\mathbf{x}$ lies in region Z_i if

$l(m_k)$ is maximum for $k = i$

ML Rule for AWGN

- Recall the LL function for AWGN channel is

$$l(m_k) = -\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{kj})^2, \quad k = 1, 2, \ldots, M$$

- So maximum value of $l(m_k)$ is attained when the term in the sum is minimized by choosing $k = i$. So ML rule for AWGN channel becomes

Observation vector $\mathbf{x}$ lies in region Z_i if

$$\sum_{j=1}^{N}(x_j - s_{kj})^2 = \|\mathbf{x} - \mathbf{s}_k\|^2 \text{ is minimum for } k = i$$

- Note that $\|\mathbf{x} - \mathbf{s}_k\|$ is the Euclidean distance between received signal point and message point, represented by $\mathbf{x}$ and $\mathbf{s}_k$, respectively

ML Rule for AWGN

- Note that

$$\|\mathbf{x}-\mathbf{s}_k\|^2 = \sum_{j=1}^{N}(x_j - s_{kj})^2 = \sum_{j=1}^{N} x_j^2 - 2\sum_{j=1}^{N} x_j s_{kj} + \underbrace{\sum_{j=1}^{N} s_{kj}^2}_{E_k}$$

- ML rule becomes

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if}$$

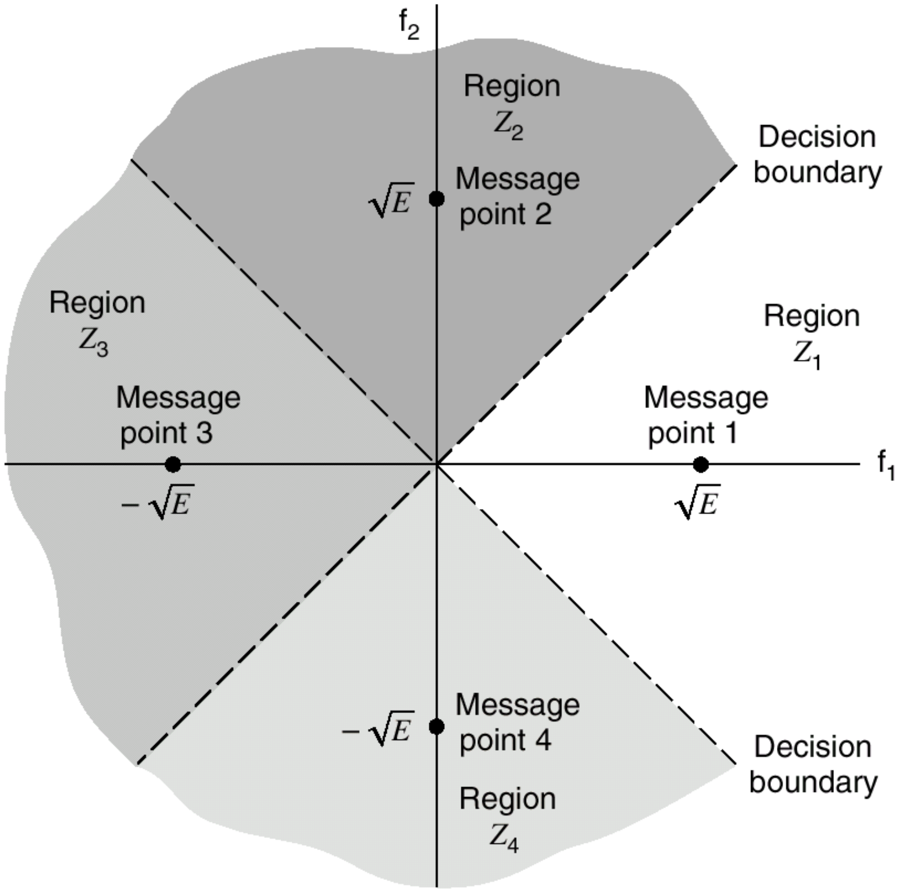
$$l(m_k) = \sum_{j=1}^{N} x_j s_{kj} - \frac{1}{2}E_k \text{ is maximum for } k = i$$

$l(m_k)$ is called the likelihood function (LL)

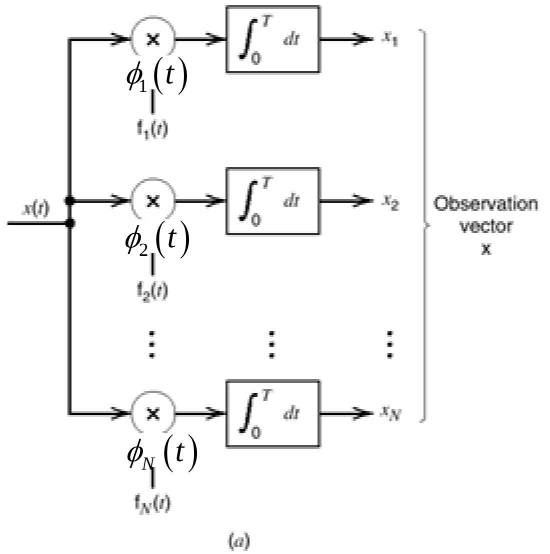
- Can deduce from this rule that the decision regions are regions of the *N*-dimensional observation space *Z*, bounded by linear [(*N*-1)-dimensional hyperplane] boundaries

Partition of Observation Space Into Decision Regions

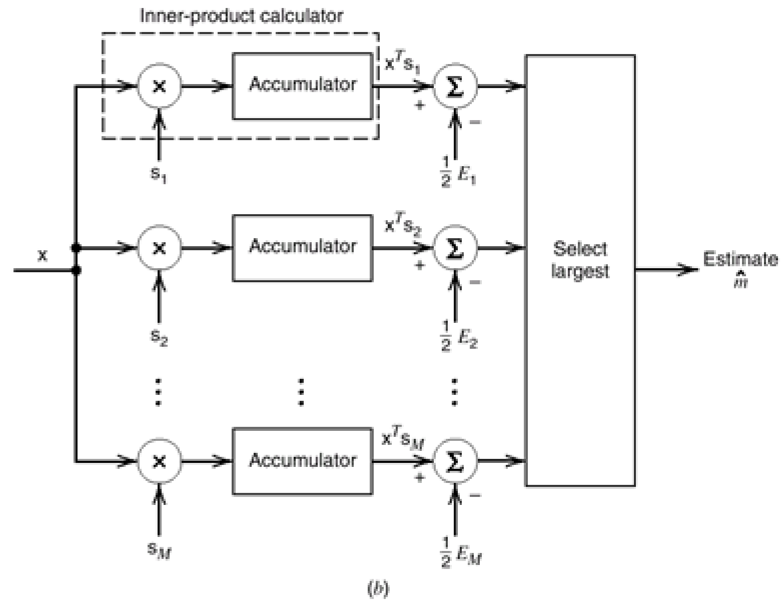
Illustrating the partitioning of the observation space into decision regions for the case when $N = 2$ and $M = 4$; it is assumed that the M transmitted symbols are equally likely.



Correlation Receiver



(a) Detector or demodulator.



(b) Signal transmission decoder.

$$x_j = \int_0^T x(t)\phi_j(t)dt, \quad \text{for } j = 1, 2, \ldots, N$$

Observation vector $\mathbf{x}$ lies in region Z_i if

$$\sum_{j=1}^{N} x_j s_{kj} - \frac{1}{2}E_k \text{ is maximum for } k = i$$

Probability of Symbol Error

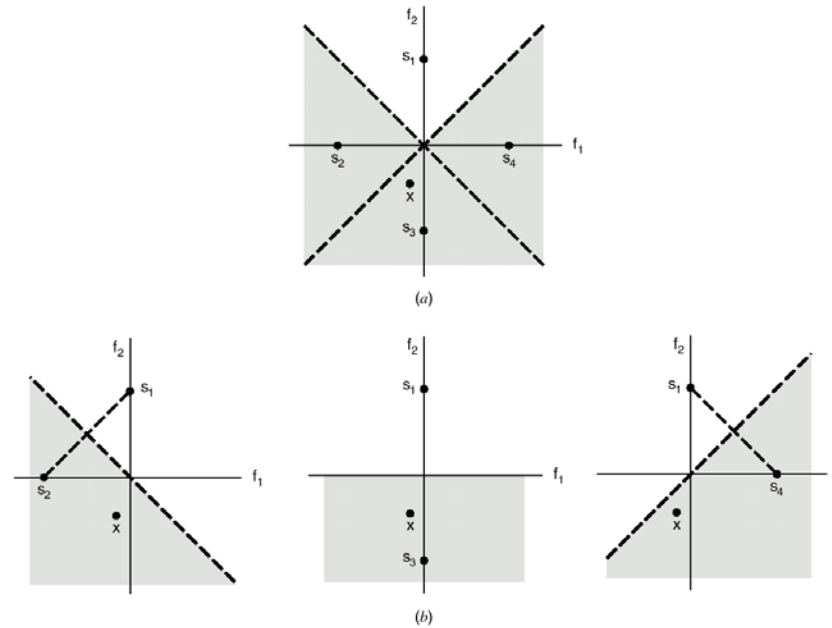
- P_e is a way to evaluate noise performance of receiver
- Use idea of observation space partitioned into M Z_i partitions, then an error occurs when **x** does not fall inside Z_i
- Use the idea of union bound to compute the pairwise probability
  - If a data transmission system uses only a pair of signals, $\mathbf{s}_i$ and $\mathbf{s}_k$, then the pairwise probability $\Pr(A_{ik}) = \Pr_2(\mathbf{s}_i,\mathbf{s}_k)$ is the probability of the receiver mistaking $\mathbf{s}_k$ for $\mathbf{s}_i$ (only *two* signal vectors are compared)
    - A_{ik} denotes the event that the observation vector **x** is closer to the signal vector $\mathbf{s}_k$ than to $\mathbf{s}_i$ when the symbol m_i (or vector $\mathbf{s}_\mathrm{i}$) is sent
  - depends on only two signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$

Union Bound on the P_e

- Conditional probability of symbol error when symbol m_i is sent, $P_e(m_i)$, is equal to the probability of the union of events, $A_{i1}, A_{i2}, \ldots, A_{i,i-1}, A_{i,i+1}, \ldots, A_{i,M}$
  - A_{ik} denotes the event that the observation vector **x** is closer to the signal vector $\mathbf{s}_k$ than to $\mathbf{s}_i$ when the symbol m_i (or vector $\mathbf{s}_i$) is sent
  - This is overbounded by the sum of the probabilities of the constituent events, i.e.

$$P_e(m_i) \le \sum_{\substack{i=1 \\ k\neq i}}^{M} Pr\left(A_{ik}\right), \text{ for } i = 1, 2, \ldots, M$$

  - That is, shaded area of (a) ≤ sum of the shaded areas in (b)



Illustrating the union bound, M = 4. $\mathbf{s}_1$ is transmitted message point. (a) Constellation of four message points. (b) Three constellations with a common message point and one other message point retained from the original constellation.

Union Bound on the P_e

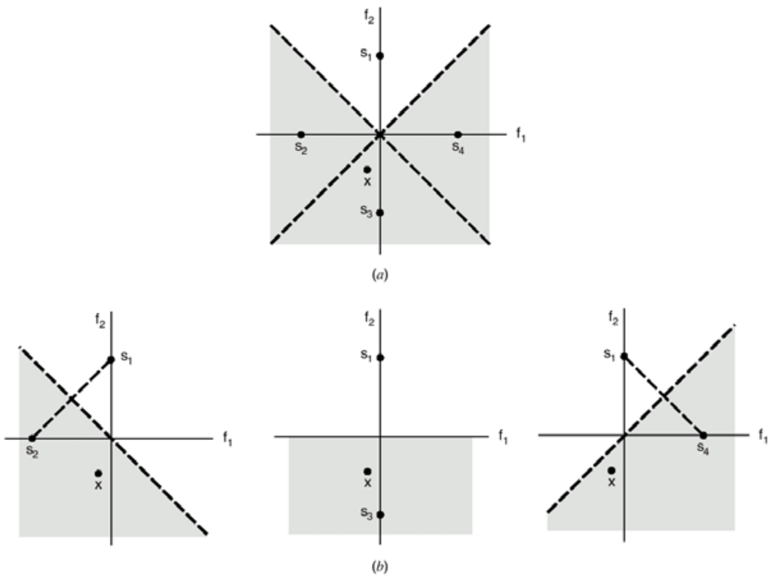
- $\Pr(A_{ik}) = \Pr_2(\mathbf{s}_i,\mathbf{s}_k)$
  - depends on only two signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$
  - It's the pairwise probability
    - If a data transmission system uses only a pair of signals, $\mathbf{s}_i$ and $\mathbf{s}_k$, then $\Pr_2(\mathbf{s}_i,\mathbf{s}_k)$ is the probability of the receiver mistaking $\mathbf{s}_k$ for $\mathbf{s}_i$ (only *two* signal vectors are compared)

$$P_e(m_i) \le \sum_{\substack{i=1 \\ k\neq i}}^{M} Pr\left(A_{ik}\right)$$
$$= \sum_{\substack{k=1 \\ k\neq i}}^{M} Pr_2\left(\mathbf{s}_i, \mathbf{s}_k\right), \text{ for } i = 1, 2, \ldots, M$$

    - Different from $Pr\left(\widehat{m} = m_k | m_i\right)$: probability of observation vector $\mathbf{x}$ is closer to signal vector $\boldsymbol{s}_k$ than *every other*, when $\mathbf{s}_i$ (or m_i) is sent

Union Bound on the P_e

- Decision boundary
  - Represented as bisector that is perpendicular to the line joining the points s_i and s_k
  - Assuming m_i (or $\mathbf{s}_i$) is sent. If $\mathbf{x}$ lies on the side of the bisector where s_k lies, error is made
    - Probability of this event



$$
\begin{aligned}
Pr_2(\mathbf{s}_i, \mathbf{s}_k) &= Pr(\mathbf{x} \text{ is closer to } \mathbf{s}_k \text{ than } \mathbf{s}_i, \\
&\qquad \text{when } \mathbf{s}_i \text{ is sent}) \\
&= \int_{d_{ik}/2}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{v^2}{N_0}\right) dv,
\end{aligned}
$$

$$
d_{ik} \triangleq \|\mathbf{s}_i - \mathbf{s}_k\|
$$

Union Bound on the P_e

- Complementary error function

$$\text{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty \exp(-z^2)\, dz$$

$$\text{Set } z = \frac{v}{\sqrt{N_0}}:$$

$$\Longrightarrow \Pr_2(\mathbf{s}_i, \mathbf{s}_k) = \frac{1}{2}\text{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right)$$

$$\Longrightarrow P_e(m_i) \le \frac{1}{2} \sum_{\substack{k=1 \\ k \ne i}}^{M} \text{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right), \text{ for } i = 1, 2, \dots, M$$

- Average over *M* symbols

$$P_e = \sum_{i=1}^{M} p_i P_e(m_i)$$

$$\le \frac{1}{2} \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \ne i}}^{M} p_i \text{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right), \text{ for } i = 1, 2, \dots, M$$

Union Bound on the P_e: Special Forms

- If signal constellation is circularly symmetric about the origin, then $P_e(m_i)$ same for all i

$$P_e \leq \frac{1}{2}\sum_{\substack{k=1\\k\neq i}}^{M} \operatorname{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right), \forall i$$

- Define $d_{\min} \triangleq \min_{k\neq i} d_{ik}, \forall i \text{ and } k$, and noting that erfc is monotonically decreasing w.r.t. u

$$\Longrightarrow \operatorname{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \leq \operatorname{erfc}\left(\frac{d_{\min}}{2\sqrt{N_0}}\right)$$

$$\therefore P_e \leq \frac{M-1}{2}\operatorname{erfc}\left(\frac{d_{\min}}{2\sqrt{N_0}}\right)$$

- Since erfc is bounded as

$$\operatorname{erfc}\left(\frac{d_{\min}}{2\sqrt{N_0}}\right) \leq \frac{1}{\sqrt{\pi}}\exp\left(-\frac{d_{\min}^2}{4N_0}\right)$$

$$\Longrightarrow P_e \leq \frac{M-1}{2\sqrt{\pi}}\exp\left(-\frac{d_{\min}^2}{4N_0}\right)$$

- That is, P_e decreases exponentially as squared minimum distance, $d_{\min}^2$

Bit Error Probability (BER)

- $\log_2 M$ bits/symbol
- Mapping binary to M-ary symbol ➔ M-tuple
  - Suppose $M = 16$: 1 0 1 1 1 0 0 1 ➔ (1011), (1001)
- Gray coding
  - Adjacent symbols only differ by one bit

$$
\begin{aligned}
P_e &= Pr\left(\bigcup_{i=1}^{\log_2 M} \{i^{th} \text{ bit is in error}\}\right) \\
&\leq \sum_{i=1}^{\log_2 M} Pr\left(i^{th} \text{ bit is in error}\right) \\
&= \log_2 M \cdot (BER)
\end{aligned}
$$

  - Note that $P_e \geq \Pr(i^{th}$ bit is in error$) = \text{BER}$

$$\therefore \frac{P_e}{\log_2 M} \leq \text{BER} \leq P_e$$

Passband Data Transmission

- Data stream is modulated onto carrier with fixed frequency limits imposed by bandpass channel
- Attention given mainly to coherent system
  - Carrier phase is sync at Rx
- M-ary signaling scheme is used in which M possible signals $s_1(t), \ldots, s_M(t)$, may be sent during signaling interval T
  - $M = 2^n$, n: # of bits
  - Symbol duration: nT_b, T_b: bit duration
  - Signals are generated by changing amplitude, phase, or frequency (or hybrid form of these) of a sinusoidal carrier in M discrete steps
  - Usually more BW efficient than binary signaling

$$\begin{aligned} \text{binary: BW} &\propto \frac{1}{T_b} \\ M\text{-ary: BW} &\propto \frac{1}{nT_b} \end{aligned}$$

- Recall that analysis of passband signal can be carried out using its baseband equivalent
- Design schemes are different
  - Maximum bandwidth efficient ρ, e.g. by trading off power (more)
  - Maximum power efficiency, e.g. by trading off P_B (higher) or bandwidth (more)
  - Minimize symbol error (P_e) or bit error (P_B), e.g. by trading bandwidth (more) or bit rate R_b

Assumptions on Passband Data Transmission

- M symbols of alphabet are equally likely with probability $p_i = \Pr(m_i) = 1/M, \forall i$
- M-ary output of the message source is injected to the encoder
  - Produces $\mathbf{s}_i$ (N complex elements)
  - Modulator constructs a distinct signal real-valued signal $s_i(t)$, from $\mathbf{s}_i$, of duration T seconds with energy

$$E_i = \int_0^T s_i^2(t)dt, \text{ for } i = 1, 2, \ldots, M$$

- Carrier is sinusoidal
  - Step change (called switching or keying) in amplitude, frequency, phase, or hybrid form in both amplitude and phase or amplitude and frequency is used by modulator to distinguish one signal from another

Assumptions on Passband Data Transmission

- Bandpass channel assumptions
  - Channel is linear, with a bandwidth that is wide enough to accommodate the transmission of the modulated signal $s_i(t)$ with little or no distortion
  - Channel noise $w(t)$ is the sample function of a white Gaussian noise process of zero mean and PSD $N_0/2$
- Receiver
  - Consists of symbol detector followed by source decoder
  - Reverses operations performed transmitter
  - Minimizes the effect of channel noise on the estimate $\hat{m}$ computed for the transmitted symbol m_i

Binary Phase-Shift Keying (BPSK)

- Pair of signals used to represent binary symbols 1 and 0

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t)$$

$$s_2(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t)$$

- E_b: transmitted signal energy per bit
- f_c chosen to be n_c/T_b, where n_c is integer
  - Ensures each transmitted bit contains an integral number of cycles of the carrier wave
- This is antipodal signals
- Envelope of signal is constrained to remain constant

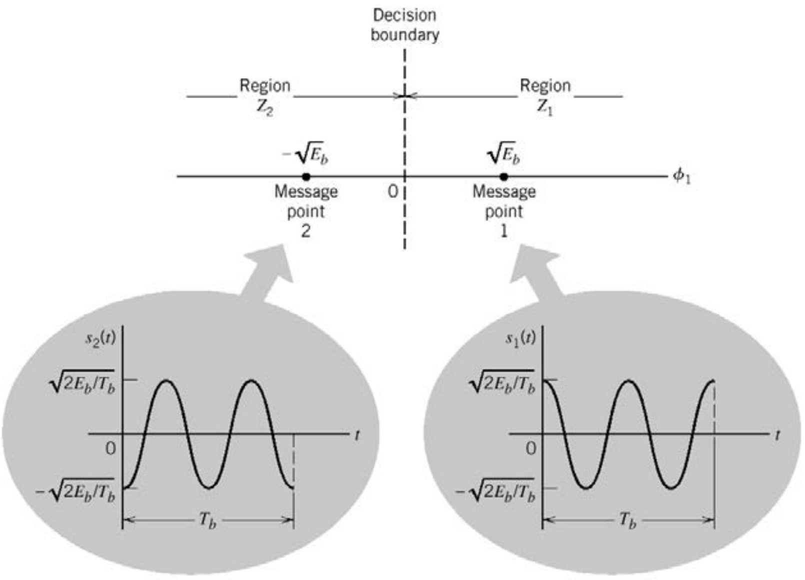
BPSK

- Note the orthonormal basis functions is

$$\phi_1(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t), \text{ for } 0 \le t < T_b$$

$$\Longrightarrow \begin{cases} s_1(t) = \sqrt{E_b}\phi_1(t), & 0 \le t < T_b \\ s_2(t) = -\sqrt{E_b}\phi_1(t), & 0 \le t < T_b \end{cases}$$

$$\Longrightarrow \begin{cases} s_{11} = \int_0^{T_b} s_1(t)\phi_1(t)dt = +\sqrt{E_b} \\ s_{21} = \int_0^{T_b} s_2(t)\phi_1(t)dt = -\sqrt{E_b} \end{cases}$$



Signal-space diagram for coherent binary PSK system. The waveforms depicting the transmitted signals $s_1(t)$ and $s_2(t)$, displayed in the inserts, assume $n_c = 2$.

P_e of BPSK

- Decision boundary
  - Midpoint of the line joining the two messages s_{11} and s_{21}
- Decision regions
  - For symbol 1 (or signal $s_1(t)$):
    - Z_1: $0 < x_1 < \infty$, where $x_1 = \int_0^{T_b} x(t)\phi_1(t)dt$
  - Conditional probability density function of rv X_1 given that symbol 0 (i.e. $s_2(t)$) was transmitted is

$$f_{X_1}(x_1|0) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_1 - s_{21})^2\right]$$
$$= \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}\left(x_1 + \sqrt{E_b}\right)^2\right]$$

P_e of BPSK

- Conditional probability of the receiver deciding in favor of symbol 1 given that symbol 0 was transmitted is

$$
\begin{aligned}
p_{10} &= \int_0^\infty f_{X_1}(x_1|0)\,dx_1 \\
&= \frac{1}{\sqrt{\pi N_0}} \int_0^\infty \exp\left[-\frac{1}{N_0}\left(x_1 + \sqrt{E_b}\right)^2\right] dx_1 \\
&= \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/N_0}}^\infty \exp\left(-z^2\right) dz \qquad \left(\text{let } z = \frac{1}{\sqrt{N_0}}\left(x_1 + \sqrt{E_b}\right)\right) \\
&= \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)
\end{aligned}
$$

- Due to symmetry, $p_{01} = p_{10}$. Hence, average symbol error probability (also equal to BER because 1 bit/symbol)

$$P_e = \text{BER} = \tfrac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$$

- As symbols 1 and 0 move apart, i.e. E_b increases, P_e decreases

M-ary PSK

- Phase of carrier takes on one of M possible values, $\theta_i = 2(i\text{-}1)\pi/M$
- That is, during signaling interval T, one of the M possible signals are

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t + \frac{2\pi}{M}(i-1)\right), \quad i = 1, 2, \ldots, M$$

E: signal energy/symbol, $f_c = n_c/T$ for fixed integer n_c

- Orthonormal basis functions are

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t), 0 \le t \le T$$

$$\phi_2(t) = \sqrt{\frac{2}{T}} \sin(2\pi f_c t), 0 \le t \le T$$

P_e of M-ary PSK

- Signal-space diagram is circulary symmetric
  - P_e is bounded by the union bound

$$P_e \le \frac{1}{2}\sum_{\substack{k=1\\k\neq i}}^{M} \operatorname{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right), \forall i$$

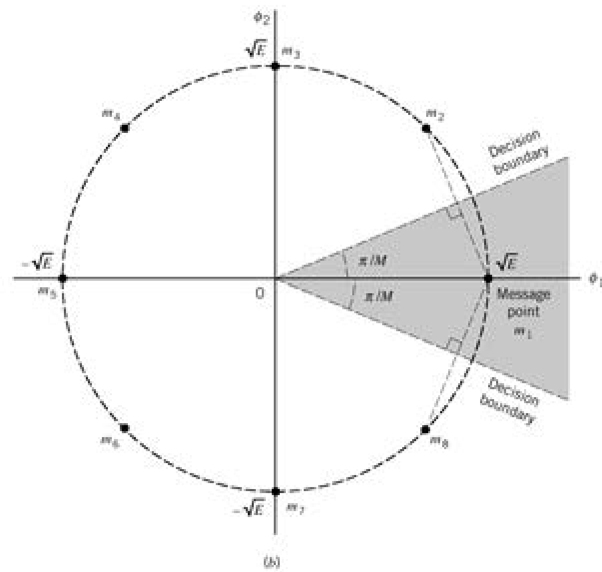
  - Assume m_1 transmitted and E/N_0 is large enough to consider the nearest two message points
  - Euclidean distance of each these two points from m_1 is

$$d_{12} = d_{18} = 2\sqrt{E}\sin\left(\frac{\pi}{M}\right)$$

  - Then average symbol error probability is

$$\Longrightarrow P_e \approx \operatorname{erfc}\left(\sqrt{\frac{E}{N_0}}\sin\left(\frac{\pi}{M}\right)\right)$$

M message points are equally spaced on a circle of radius $\sqrt{E}$ and center at the origin



Signal-space diagram for octaphase-shift keying (i.e., $M = 8$), illustrating the application of the union bound for octaphase-shift keying.

Bandwidth Efficiency of *M*-ary PSK Signals

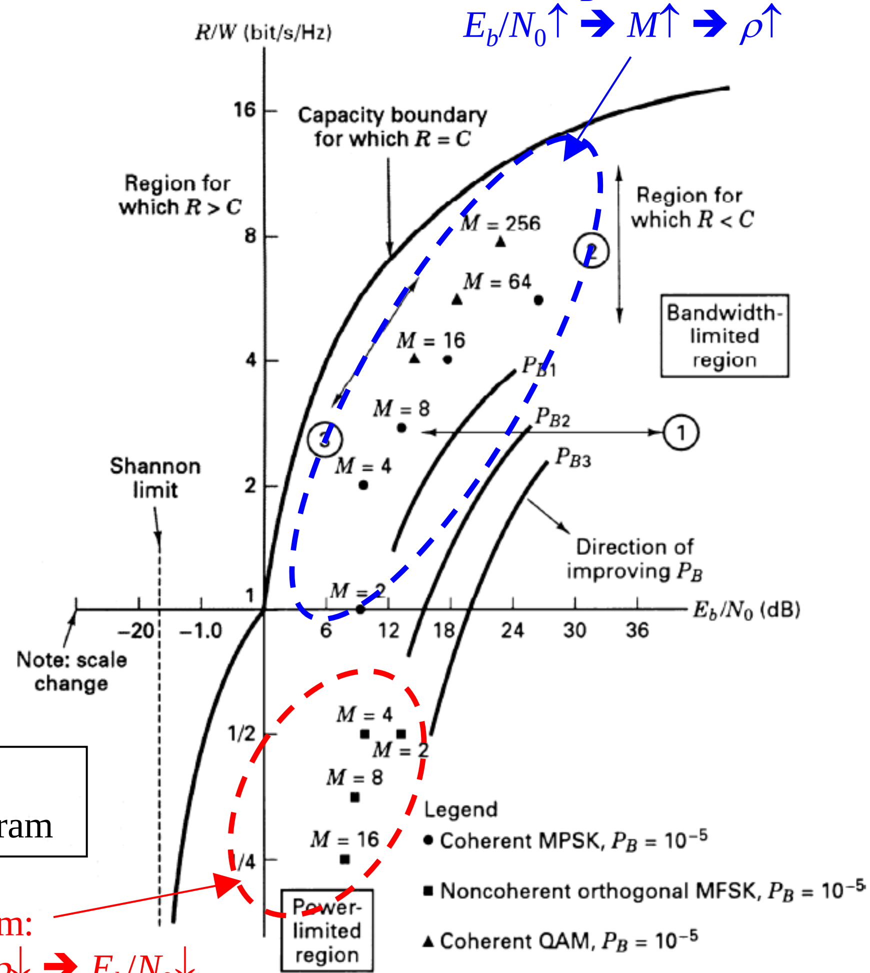
- Recall bandwidth efficiency

$$\rho \triangleq \frac{R_b}{B} = \frac{\log_2 M}{2} = \frac{\log_2 M}{BT_s} \text{bits/s/Hz}$$

Bandwidth efficiency of M-ary PSK signals

M	2	4	8	16	32	64
ρ (bits/s/Hz)	0.5	1	1.5	2	2.5	3

BW-limited system:
Fixed P_B and B,
$E_b/N_0\uparrow$ ➔ $M\uparrow$ ➔ $\rho\uparrow$



Bandwidth-efficiency diagram

Pwr-limited system:
Fixed P_B, $B\uparrow$ ➔ $\rho\downarrow$ ➔ $E_b/N_0\downarrow$

Information Theory

- Entropy
  - Formally defined as the probabilistic behavior of a source of information
  - Randomness of data
- Capacity
  - Intrinsic ability of a channel to convey information
    - Noise characteristics of the channel
- If entropy of the source is less than the capacity of the channel, error-free communication over the channel can be achieved

Information

- Source output modeled as discrete random variable S, which takes on symbols from a fixed finite alphabet
  - $\mathscr{S} = \{s_0, s_1, \ldots, s_{K\text{-}1}\}$
  - With probabilities $P(S=s_k) = p_k$, $k = 0, 1, \ldots, K\text{-}1$
  - Constraint
    $$\sum_{k=0}^{K-1} p_k = 1$$
  - Source symbols are assumed to be statistically independent
    - Discrete memoryless source
- Information defined as $I(s_k) = \log\left(\frac{1}{p_k}\right)$
  - Hence, less probability of symbol occurring, the more information it contains when it occurs

Properties of $I(s_k)$

- $I(s_k) = 0$, for $p_k = 1$
  - If we are certain of the outcome of an event, no information is gained
- $I(s_k) \geq 0$, for $0 \leq p_k \leq 1$
  - Occurrence of event $S = s_k$ provides some or no information, but never brings about a loss of information
- $I(s_k) > I(s_i)$, for $p_k < p_i$
  - Less probable an event is, the more information we gain when it occurs
- $I(s_k s_i) = I(s_k) + I(s_i)$ if s_k and s_i are statistically independent
- Base of log is arbitrary but usually 2
  - Results unit of information is called the bit

$$I(s_k) = \log\left(\frac{1}{p_k}\right) = -\log_2(p_k), \text{ for } k = 0,1,\ldots,K-1$$

  - $p_k = ½$, ➔ $I(s_k) = 1$ bit
    - One bit is the amount of information that we gain when one of two possible and equally likely events occurs

Entropy

- Entropy

$$H(\mathscr{S}) = E\left[I(s_k)\right] = \sum_{k=0}^{K-1} p_k I(s_k)$$
$$= \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

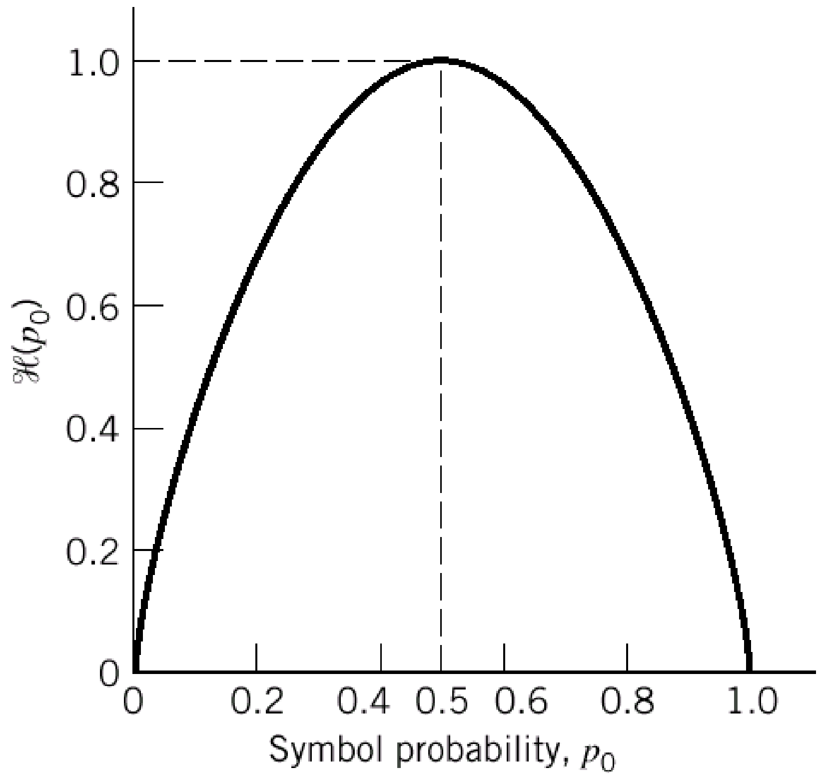
  - Measures the average information content per source symbol
  - $\mathscr{S}$ is not an argument for H, it's only a label.
  - H only depends on the probabilities of the symbols in the alphabet $\mathscr{S}$ of the source

Example: Entropy of Binary Memoryless Source

- Let symbol 0 occurs with probability p_0
- Let symbol 1 occurs with probability $p_1 = 1 - p_0$
- Source symbols are statistically independent
- Entropy is

$$H(\mathscr{S}) = -p_0 \log_2 p_0 - p_1 \log_2 p_1$$
$$= -p_0 \log_2 p_0 - (1 - p_0)\log_2 (1 - p_0) \text{ bits}$$

- Observations
  - When $p_0 = 0$, entropy = 0
    - Because $x \log x \rightarrow 0$ as $x \rightarrow 0$
  - When $p_0 = 1$, the entropy = 0
  - Entropy attains its maximum value, H_{max} = 1 bit, when $p_1 = p_0 = ½$, i.e. 1 and 0 are equiprobable
- Entropy function $H(p_0)$ is plotted on the right
  - Function of a priori probability p_0
  - Notice that $H(\mathscr{S})$ gives entropy of a discrete memoryless source with source alphabet $\mathscr{S}$ (difference is subtle)



Example – Entropy of Extended Source

- Consider a discrete memoryless source with alphabet $\mathscr{S} = \{s_0, s_1, s_2\}$ with respective probabilities $p_0=1/4$, $p_1 = 1/4$, $p_2 = ½$

➔ $H(\mathscr{S}) = p_0\log_2(1/p_0) + p_1\log_2(1/p_1) + p_2\log_2(1/p_2)$

$$= (1/4)\log_2(4) + (1/4)\log_2(4) + (1/2)\log_2(2)$$

$$= 3/2 \text{ bits}$$

Example – Entropy of Extended Source

Extended source: Each block consists of *n* successive source symbols

- Suppose now that alphabet $\mathscr{S}^2$ which consists of 9 symbols $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8\}$ with the same source alphabet as before

Alphabet particulars of second-order extension of a discrete memoryless source

Symbols of $\mathscr{S}^2$	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Corresponding sequences of symbols of $\mathscr{S}$	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Probability $p(\sigma_i)$, $i = 0, 1, \ldots, 8$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$
\begin{aligned}
H\left(\mathscr{S}^2\right) &= \sum_{i=0}^{8} p(\sigma_i)\log_2\frac{1}{p(\sigma_i)} \\
&= \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{16}\log_2(16) \\
&\quad \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{4}\log_2(4) \\
&= 3 \text{ bits}
\end{aligned}
$$

Discrete Memoryless Channel

- Discrete memoryless channel is a statistical model with an input X and an output Y that is a noisy version of X
  - X and Y are rv's
- Channel accepts an input symbol X (from alphabet $\mathscr{X}$) and emits an output symbol Y (from alphabet $\mathscr{Y}$)
- Channel is "discrete" when both alphabets $\mathscr{X}$ and $\mathscr{Y}$ have finite sizes
- Input alphabet

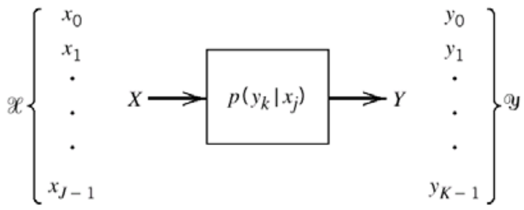
$$\mathcal{X} = \{x_0, x_1, \ldots, x_{J-1}\}$$

- Output alphabet

$$\mathcal{Y} = \{y_0, y_1, \ldots, y_{K-1}\}$$

- Transition probabilities

$$p(y_k|x_j) = Pr(Y = y_k|X = x_j), \forall j \text{ and } k,$$
$$0 \leq p(y_k|x_j) \leq 1$$



The channel or transition matrix is

$$\mathbf{P} = \begin{bmatrix} p(y_0|x_0) & p(y_1|x_0) & \cdots & p(y_{K-1}|x_0) \\ p(y_0|x_1) & p(y_1|x_1) & \cdots & p(y_{K-1}|x_1) \\ \vdots & \vdots & \ddots & \vdots \\ p(y_0|x_{J-1}) & p(y_1|x_{J-1}) & \cdots & p(y_{K-1}|x_{J-1}) \end{bmatrix} \in \mathbb{R}^{J \times K},$$

and
$\sum_{k=0}^{K-1} p(y_k|x_j) = 1, \forall j$

Probability distribution
$p(x_j) = \Pr(X=x_j)$, for $j = 0, 1, J\text{-}1$
Joint probability distribution
$p(x_j,y_k) = \Pr(Y=y_k|X=x_j)\Pr(X=x_j) = p(y_k|x_j)p(X=x_j)$
Marginal pdf Y

$$p(y_k) = \Pr(Y = y_k)$$
$$= \sum_{j=0}^{J-1} p(y_k|x_j)p(x_j), \text{ for } k = 0, 1, \ldots, K-1$$

Mutual Information

- $H(\mathscr{X})$: entropy of input
  - Represents uncertainty about channel input *before* observing channel output
- $H(\mathscr{X}|\mathscr{Y})$: conditional entropy
  - Represents uncertainty about channel input *after* observing channel output
- → $I(\mathscr{X};\mathscr{Y}) \equiv H(\mathscr{X}) - H(\mathscr{X}|\mathscr{Y})$: mutual information
  - Represent uncertainty about channel input that is *resolved* by observing the channel output
- Similarly $I(\mathscr{Y};\mathscr{X}) \equiv H(\mathscr{Y}) - H(\mathscr{Y}|\mathscr{X})$

$$H(\mathscr{X}|Y=y_k) = \sum_{j=0}^{J-1} p(x_j|y_k)\log_2\left[\frac{1}{p(x_j|y_k)}\right]$$

$$\begin{aligned}
H(\mathscr{X}|\mathscr{Y}) &= E_Y\left[H(\mathscr{X}|Y=y_k)\right] \\
&= \sum_{k=0}^{K-1} H(\mathscr{X}|Y=y_k)p(y_k) \\
&= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j|y_k)p(y_k)\log_2\left[\frac{1}{p(x_j|y_k)}\right] \\
&= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j,y_k)\log_2\left[\frac{1}{p(x_j|y_k)}\right]
\end{aligned}$$

Channel Capacity

- $I(\mathscr{X};\mathscr{Y}) \equiv H(\mathscr{X}) - H(\mathscr{X}|\mathscr{Y})$, using the equation of $H(\mathscr{X})$ and $H(\mathscr{X}|\mathscr{Y})$, and using the fact that $p(x_j|y_k)/p(x_j) = p(y_k|x_j)/p(y_k)$ (Bayes' rule)

$$I(\mathscr{X};\mathscr{Y}) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k)\log_2\left[\frac{p(x_j|y_k)}{p(x_j)}\right]$$

$$= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k)\log_2\left[\frac{p(y_k|x_j)}{p(y_k)}\right] = I(\mathscr{Y};\mathscr{X}) > 0$$

- $I(\mathscr{X};\mathscr{Y})$: measure of uncertainty about the channel input that is resolved by observing channel output
- $I(\mathscr{Y};\mathscr{X})$: measure of uncertainty about the channel output that is resolved by sending the channel input
- Cannot lose information by observing output of channel

- Note that

$$p(x_j, y_k) = p(y_k|x_j)p(x_j)$$

$$p(y_k) = \sum_{j=0}^{J-1} p(y_k|x_j)p(x_j)$$

  - MI depends not only on channel, but also on the input pdf $p(x_j)$

Channel Capacity

- Since channel is not dependent of input, define capacity of channel as

$$C \triangleq \max_{\{p(x_j)\}} I(\mathscr{X};\mathscr{Y})$$
$$s.t.\ p(x_j) \geq 0, \forall j$$
$$\sum_{j=0}^{J-1} p(x_j) = 1$$

measured in bits/channel use, or bits/transmission

  - Depends only on transition probabilities $p(y_k|x_j)$

Example: Binary Symmetric Channel

- Special case of discrete memoryless channel
- Input
  - $x_0 = 0, x_1 = 1$
- Output
  - $y_0 = 0, y_1 = 1$
- ➔ $J = K = 2$
- Note that

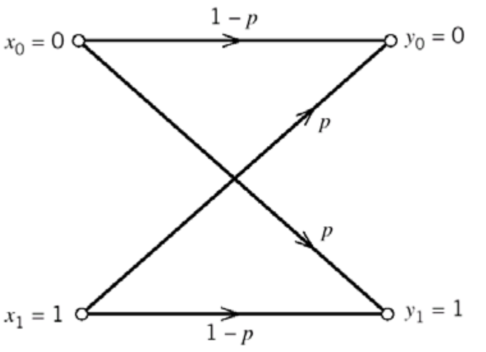
$$p(x_j, y_k) = p(y_k|x_j)p(x_j)$$

$$\Longrightarrow p(y_0|x_0)p(x_0) = p(y_1|x_1)p(x_1) = \frac{1-p}{2}$$

$$\Longrightarrow p(y_0|x_1)p(x_1) = p(y_1|x_0)p(x_0) = \frac{p}{2}$$

$$p(y_k) = \sum_{j=}^{J-1} p(y_k|x_j)p(x_j)$$

$$\Longrightarrow p(y_0) = p(y_1) = \frac{1-p}{2} + \frac{p}{2} = \frac{1}{2}$$



Transition probability diagram of binary symmetric channel.

Transition probability matrix

$$\mathbf{P} = \begin{bmatrix} p(y_0|x_0) & p(y_1|x_0) \\ p(y_0|x_1 & p(y_1|x_1) \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

Example: Binary Symmetric Channel

$$C \triangleq \max_{\{p(x_j)\}} I(\mathscr{X};\mathscr{Y}) = \max_{\{p(x_j)\}} \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k)\log_2\left[\frac{p(y_k|x_j)}{p(y_k)}\right]$$

$$s.t.\ p(x_j) \geq 0, \forall j$$

$$\sum_{j=0}^{J-1} p(x_j) = 1$$

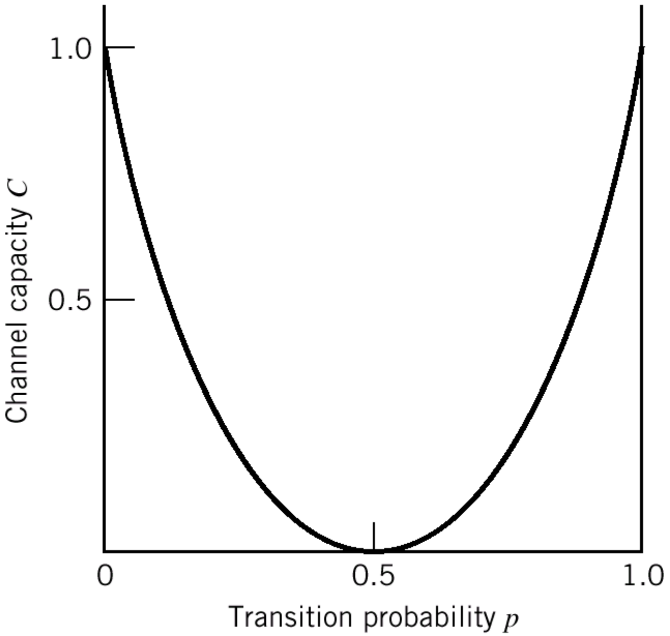
- Since $I(\mathscr{X};\mathscr{Y}) \equiv H(\mathscr{X}) - H(\mathscr{X}|\mathscr{Y})$, maximum can be achieved by maximizing $H(\mathscr{X})$
  - This happens when $p(x_0) = p(x_1) = ½$
- ➔

$$\begin{aligned} C &\triangleq \max_{\{p(x_j)\}} \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k)\log_2\left[\frac{p(y_k|x_j)}{p(y_k)}\right] \\ &= 2\cdot\frac{1-p}{2}\left[\log_2 2(1-p)\right] + 2\cdot\frac{p}{2}\left[\log_2 2p\right] \\ &= (1-p)\left[\log_2 2 + \log_2(1-p)\right] + p\left[\log_2 2 + \log_2 p\right] \\ &= 1 + (1-p)\log_2(1-p) + p\log_2 p \\ &= 1 - H(p) \end{aligned}$$

$$\begin{aligned} H(\mathscr{S}) &= -p_0\log_2 p_0 - p_1\log_2 p_1 \\ &= -p_0\log_2 p_0 \\ &\quad -(1-p_0)\log_2(1-p_0)\ \text{bits} \end{aligned}$$

Example: Binary Symmetric Channel

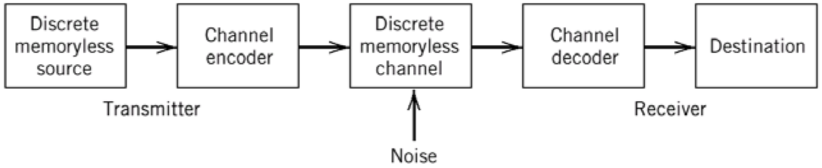
- C varies with probability of error (transition probability) p in a convex manner
  - Symmetric around $p = 1/2$
- When channel is noise free, $p = 0$, ➔ C is maximum with 1 bit/channel use (equals to the information in each channel input)
  - Coincide with minimum value of $H(p)$, which equals 0
- When $p = ½$ due to noise, the channel capacity C attains its minimum value of zero
  - Coincide with maximum value of $H(p)$, which equals 1
    - Channel is useless in this case



Variation of channel capacity of a binary symmetric channel with transition probability p.

Channel-Coding Theorem: Channel Coding

- Channel encoder
  - Introduce redundancy to increase probability that the original source sequence can be reconstructed
  - Dual of source coding
    - Reduces redundancy to improve transmission efficiency
- Assume block codes are used, i.e. message sequence is divided into block of k bits long, each k-bit block is mapped into an n-bit block, with $n > k$
  - Redundant bits added n-k
- Define code rate: $r \equiv k/n$
- Capacity-Coding Theorem includes the notion of time



Block diagram of digital communication system.

Channel-Coding Theorem for Discrete Memoryless Channel

1. Let a discrete memoryless source with an alphabet $\mathscr{S}$ have entropy $H(\mathscr{S})$ and produce symbols every T_s sec. Let a discrete memoryless channel have capacity C be used once every T_c sec. If

$$\frac{H(\mathscr{S})}{T_s} \leq \frac{C}{T_c}$$

there exists a coding scheme for which the source output can be transmitted over the channel and be reconstructed with an arbitrary small probability of error. C/T_c is called the ***critical rate*** (in bits/sec).

2. Conversely, if

$$\frac{H(\mathscr{S})}{T_s} > \frac{C}{T_c}$$

it is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error

$\frac{H(\mathscr{S})}{T_s}$ is known as the average information rate of the source

Channel-Coding Theorem

- Theorem does not show us how to construct a good code. It is simply an existence proof in that it tells us that if the average information rate is less than the critical rate, then good codes do exist
- Theorem also does not provide precise result for the probability of symbol error after decoding the channel output. It does tell us the probability of symbol error tends to zero as the length of the code increases, provided that $\frac{H(\mathscr{S})}{T_s} \leq \frac{C}{T_c}$

Differential Entropy

- Consider a continuous rv X with the probability density function $f_X(x)$.
- Differential entropy of X is defined as

$$h(X) \triangleq \int_x f_X(x) \log_2 \left[\frac{1}{f_X(x)}\right] dx$$

- It's "differential" because it's measured based on a reference $\lim_{\Delta x \to 0} \log_2 \Delta x$. To see this
  - Since discrete RV $x_k = k\Delta x$, for $k = 0, 1, 2, \ldots$, and Δx approaches zero
  - That is, X assumes a value in the interval $[x_k, x_k + \Delta x]$ with probability $f_X(x_k)\,\Delta x$ as Δx approaching to zero

$$\begin{aligned}
H(X) &= \lim_{\Delta x \to 0} \sum_k f_X(x) \Delta x \log_2 \left(\frac{1}{f_X(x_k)\Delta x}\right) \\
&= \lim_{\Delta x \to 0} \left[\sum_k f_X(x) \log_2 \left(\frac{1}{f_X(x_k)}\right) \Delta x - \log_2 \Delta x \sum_k f_X(x_k) \Delta x\right] \\
&= \int_x f_X(x) \log_2 \left(\frac{1}{f_X(x)}\right) dx - \lim_{\Delta x \to 0} \log_2 \Delta x \int_x f_X(x)\, dx \\
&= h(X) - \lim_{\Delta x \to 0} \log_2 \Delta x
\end{aligned}$$

Goes to ∞, so measure $H(X)$ based on reference

Differential Entropy

- Extension to $\mathbf{X} = [X_1, X_2, \ldots, X_n]^T$. Differential entropy is defined as the *n*-fold integral

$$h(\mathbf{X}) \triangleq \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \log_2 \left[\frac{1}{f_{\mathbf{X}}(\mathbf{x})} \right] d\mathbf{x}$$

- Consequences on dealing with cont. rv: differential entropy can be negative

Mutual Information (Cont. RV)

- Mutual information between rv's X and Y is defined as

$$I(X;Y) = \int_y \int_x f_{X,Y}(x,y) \log_2 \left[\frac{f_X(x|y)}{f_X(x)} \right] dxdy$$

- Properties of MI
  1. $I(X;Y) = I(Y;X)$
  2. $I(X;Y) \geq 0$
  3. $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$
- Conditional differential entropy

$$h(X|Y) = \int_y \int_x f_{X,Y}(x,y) \log_2 \left[\frac{1}{f_X(x|y)} \right] dxdy$$

Information Capacity

- Like to formulate information capacity theorem for *bandwidth-limited, power-limited Gaussian channels*
- Let X_k, $k = 1, 2, \ldots, K$, be a cont. rv obtained by uniform sampling of the process zero-mean stationary process $X(t)$ at Nyquist rate of $2B$ samp/sec (i.e. $X(t)$ is *bandlimited* to B)
  - Samples transmitted in T sec
  - ➔Number of samples: $K = 2BT$
- Input to channel: X_k
  - *Power limited*: $E[X_k^2] = P$
  - P: average transmitted power
- Channel output, Y_k, perturbed by *additive white Gaussian noise* (AWGN) of zero mean and PSD $N_0/2$
  - Noise is bandlimited to B Hz
    - $Y_k = X_k + N_k$, for $k = 1, 2, \ldots, K$
    - Noise sample N_k is Gaussian: zero mean and $\sigma^2 = N_0 B$ variance
  - Y_k, $\forall k$ are statistically independent

Satisfy the requirement that the channel is BW- and power-limited Gaussian

Information Capacity

$$h(X) \triangleq \int_x f_X(x) \log_2 \left[\frac{1}{f_X(x)}\right] dx$$

$$h(\mathbf{X}) \triangleq \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \log_2 \left[\frac{1}{f_{\mathbf{X}}(\mathbf{x})}\right] d\mathbf{x}$$

- Information capacity is defined as

$$C \triangleq \max_{f_{X_k}(x)} I(X_k; Y_k)$$
$$s.t.\ E[X_k^2] = P$$

- $I(X_k;Y_k) = h(Y_k) - h(Y_k|X_k) = h(Y_k) - h(N_k)$
  - 2nd equality true because can be shown that $h(Y_k|X_k) = h(N_k)$
  - Since $h(N_k)$ is indep. of the pdf of X_k, C can be obtained by maximizing $h(Y_k)$
    - Can be shown that $h(Y_k)$ is maximized iff Y_k is Gaussian distributed
      - Since N_k is Gaussian ➔ X_k is also Gaussian
- Hence, capacity can be reformulated as

$$C = I(X_k; Y_k)$$
$$\text{with } X_k \text{ Gaussian, } E[X_k^2] = P$$

Information Capacity Theorem

- Assume channel is used K times for transmission of K samples of the process $X(t)$ in T sec
  - Information capacity per unit time is $C \cdot K/T = C \cdot 2BT/T = C \cdot 2B$

$$C = B \log_2 \left(1 + \frac{P}{N_0 B}\right) \text{ bits per sec}$$

- **Information Capacity Theorem**:

  The information capacity of a continuous channel of bandwidth B Hz, perturbed by additive white Gaussian noise of power spectral density $N_0/2$ and limited in bandwidth to B, is given by

$$C = B \log_2 \left(1 + \frac{P}{N_0 B}\right) \text{ bits per sec},$$

  where P is the average transmitted power

Consequence of Information Capacity Theorem

- Dependence of C on channel bandwidth B is linear
- Dependence of C on signal-to-noise ratio $P/(N_0B)$ is logarithm
  - Easier to increase information capacity of a communication channel by expanding its bandwidth than increasing the transmitted power for a prescribed noise variance

Bandwidth-Efficiency Diagram

- Define an ideal system, i.e. $R_b = C$
  - Average transmitted power: $P = E_b C$
    - E_b: transmitted energy/bit
  - ➔ Ideal system can then be defined as

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right)$$

  - ➔ Signal energy-per-bit to noise PSD ratio, E_b/N_0, can be written as

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B}$$

- R_b/B vs. E_b/N_0 is called the bandwidth-efficiency diagram
- Taylor Series expansion:

$$2^{C/B} = e^{(C/B)\ln 2} \approx 1 + \frac{C}{B}\ln 2$$

$$\text{Also } \frac{C/B}{\log_2 e} = \ln\left(1 + \frac{P}{N_0 B}\right) \approx \frac{P}{N_0 B} \Longrightarrow C = \frac{P}{N_0}\log_2 e$$

Bandwidth-Efficiency Diagram

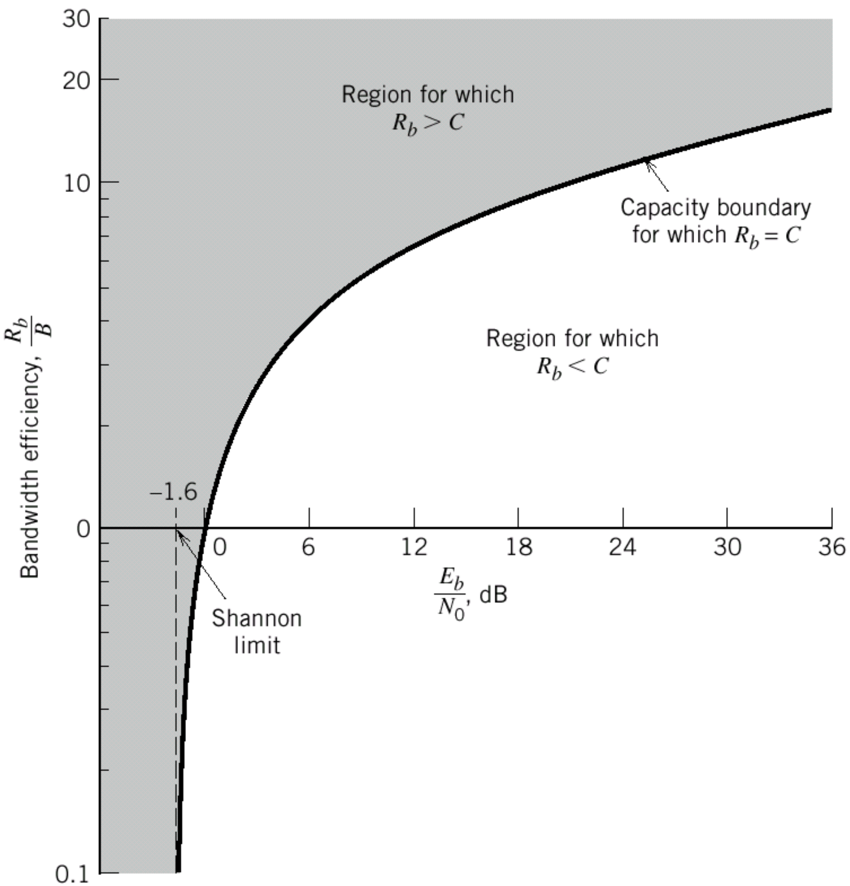
- For infinite bandwidth, E_b/N_0 approaches the limit value

$$\left(\frac{E_b}{N_0}\right)_\infty = \lim_{B\to\infty} \frac{E_b}{N_0} \approx \lim_{B\to\infty} \frac{1+\frac{C}{B}\ln 2 - 1}{C/B}$$
$$= \ln 2 = 0.693 \Leftrightarrow -1.6 \text{ dB}$$

- Corresponding limiting value of the channel capacity is

$$C_\infty = \lim_{B\to\infty} C$$
$$= \frac{P}{N_0}\log_2 e$$

- Capacity boundary: $R_b = C$
  - Separates combination of system parameters that has the potential to support error-free transmission ($R_b < C$) from those that cannot support error-free transmission ($R_b > C$)
- The diagram highlights trade-offs among E_b/N_0, R_b/B, and probability of symbol error P_e
  - Movement along a horizontal line as trading P_e vs. E_b/N_0 for fixed R_b/B (bandwidth-limited system)
  - Vertical movement: trading P_e vs. R_b/B for a fixed E_b/N_0 (power-limited system)



Concluding Remarks

- Proper modeling of (additive and convolutive) noise (incl. interference) is important
  - Probabilistic models are often used
- Design
  - Optimal design is crucial
    - Many “optimal” designs are not optimal – depends on objective
  - How do we do it? (We are engineers, this is important!)
    - Statistical signal detection and estimation theory
      - Wiener optimum filter, matched filter, adaptive filter, and many more…
    - Information theory and coding
      - Shannon says it can be done, but didn’t tell us how it can be done