

EE 5654 - Digital Communications Spring 2005



Instructor: R. Michael Buehrer

Lecture #18 - Capacity and Channel Coding
with Block Codes



Channel Capacity

- Let X represent the transmitted symbol at the input to a channel and let Y represent the symbol received at the output of a channel.



- If X and Y are related by some probabilistic distribution, then it is possible to define a quantity called channel capacity

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

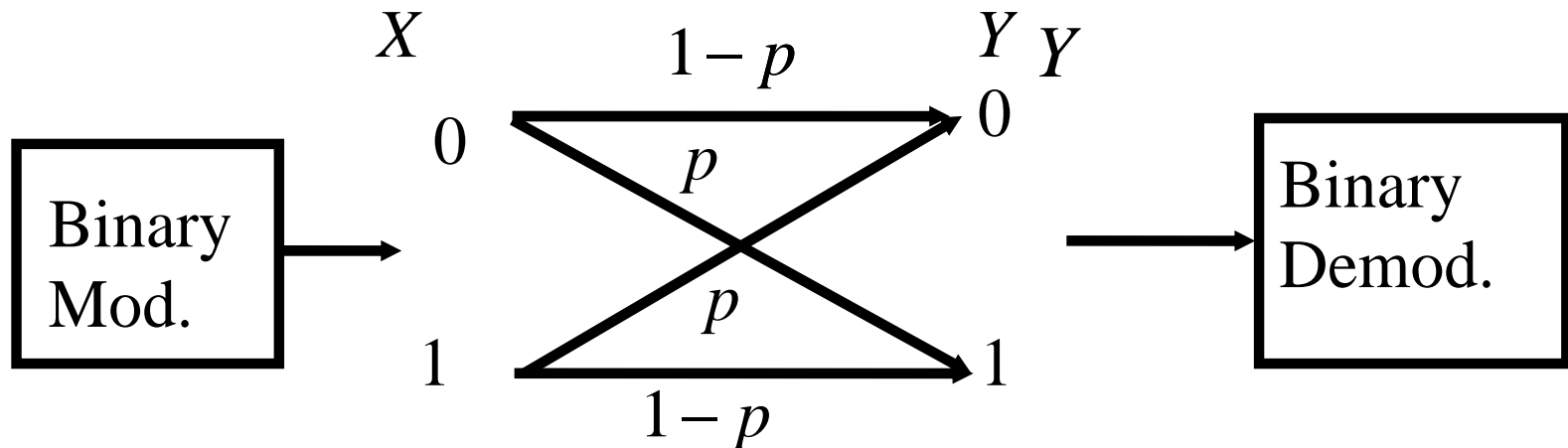
- where $I(X; Y)$ is the "mutual information" between X and Y .



Channel Capacity (continued)

- Channel capacity C has units of information rate (either “bits/sec” or “bits/sec/Hz”).
- C represents the fastest theoretical rate at which error free transmission is possible over a channel.
- This fact is called the “Channel Coding Theorem”. It’s proof is one of the principle results of Information Theory and is beyond the scope of this course.
- We are interested in designing a system to operate as close to this fundamental limit as possible.

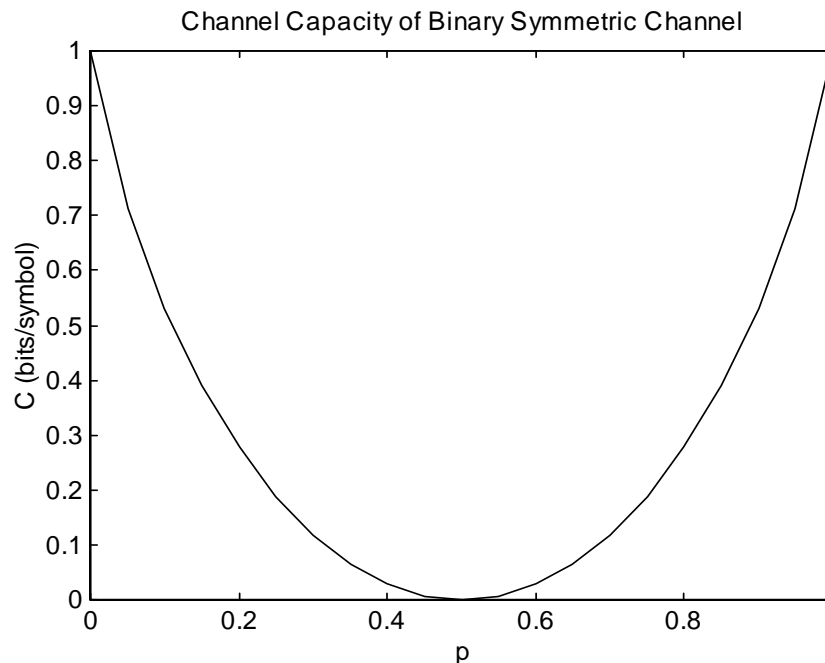
Capacity of Binary Symmetric Channel (BSC)



- p = probability of error for a binary modulation scheme.

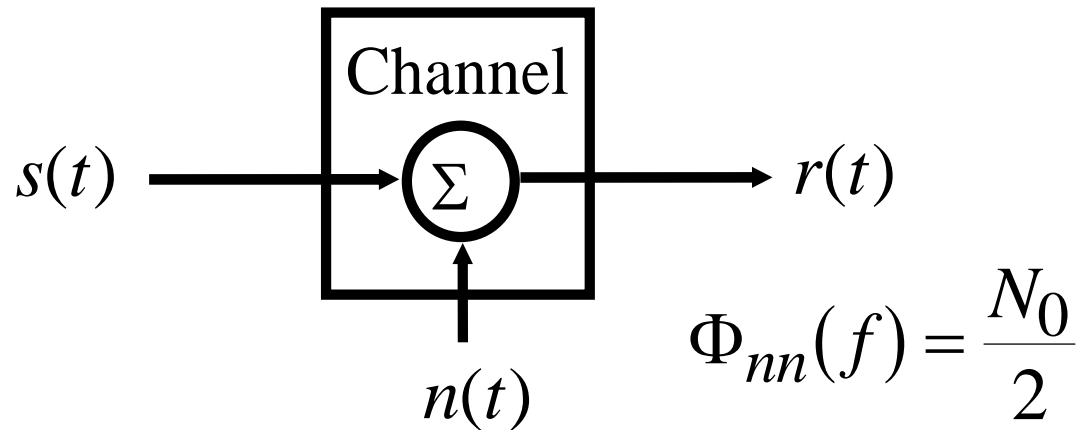
$$C = 1 + p \log_2 p + (1-p) \log_2 (1-p) \quad \frac{\text{bits}}{\text{symbol}}$$

Plot of Capacity for BSC



- Note that errorless transmission is possible even for nonzero p (just at a slower rate).
- Note that curve is symmetrical about $p=0.5$

Channel Capacity of AWGN Channel



- For AWGN channel: $C = W \log_2 \left(1 + \frac{P_{av}}{WN_0} \right) \frac{\text{bits}}{\text{second}}$
 - P_{av} = Average signal power
 - W = Available bandwidth



Capacity the AWGN Channel

- Consider a bandlimited AWGN channel where the waveform $x(t)$ is the input and the waveform $y(t)$ is the output

$$y(t) = x(t) + n(t)$$

and $n(t)$ is additive white Gaussian noise

- We know that we can express these waveforms in terms of $N=2WT$ orthonormal basis functions $f_i(t)$ where W is the bandwidth and T is the time interval
- The coefficients have conditional probabilities

$$\begin{aligned} p(y_1, y_2 \dots y_N | x_1, x_2 \dots x_N) &= \prod_{i=1}^N p(y_i | x_i) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(y_i - x_i)^2}{2\sigma_i^2}} \end{aligned}$$



Capacity (cont.)

- The capacity for this channel is defined as

$$C = \lim_{T \rightarrow \infty} \max_{p(x_j)} \frac{1}{T} I(X; Y)$$

- Further, the mutual information is defined as

$$\begin{aligned} I(\mathbf{X}_N; \mathbf{Y}_N) &= \int_{\mathbf{x}_N} \dots \int_{\mathbf{y}_N} \dots \int p(\mathbf{y}_N | \mathbf{x}_N) p(\mathbf{x}_N) \log \frac{p(\mathbf{y}_N | \mathbf{x}_N)}{p(\mathbf{x}_N)} d\mathbf{x}_N d\mathbf{y}_N \\ &= \sum_{i=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_i | x_i) p(x_i) \log \frac{p(y_i | x_i)}{p(x_i)} dy_i dx_i \end{aligned}$$



Capacity (cont.)

- This is maximized when

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma_x}} e^{-\frac{x_i^2}{2\sigma_x^2}}$$

- The maximum is

$$\begin{aligned} \max_{p(x_i)} I(\mathbf{X}_N; \mathbf{Y}_N) &= \sum_{i=1}^N \frac{1}{2} \log \left(1 + \frac{2\sigma_x^2}{N_o} \right) \\ &= \frac{N}{2} \log \left(1 + \frac{2\sigma_x^2}{N_o} \right) \\ &= WT \log \left(1 + \frac{2\sigma_x^2}{N_o} \right) \end{aligned}$$



Capacity (cont.)

- Now, defining an average power limit

$$\begin{aligned} P_{av} &= \frac{1}{T} \int_0^T E\{x^2(t)\} dt \\ &= \frac{1}{T} \sum_{i=1}^N E\{x_i^2\} \\ &= \frac{N\sigma_x^2}{T} \\ &= 2W\sigma_x^2 \end{aligned}$$

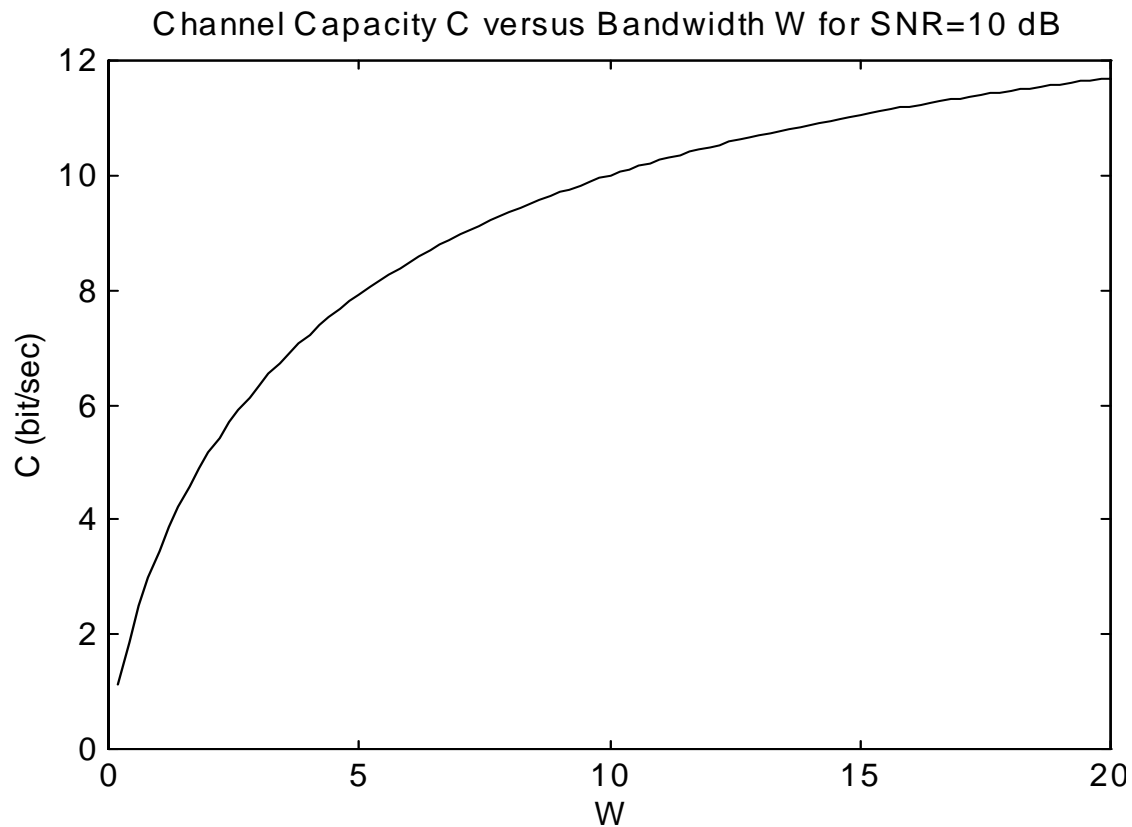
- Substituting

$$\max_{p(x_i)} I(\mathbf{X}_N; \mathbf{Y}_N) = WT \log \left(1 + \frac{P_{av}}{WN_o} \right)$$

- Thus,

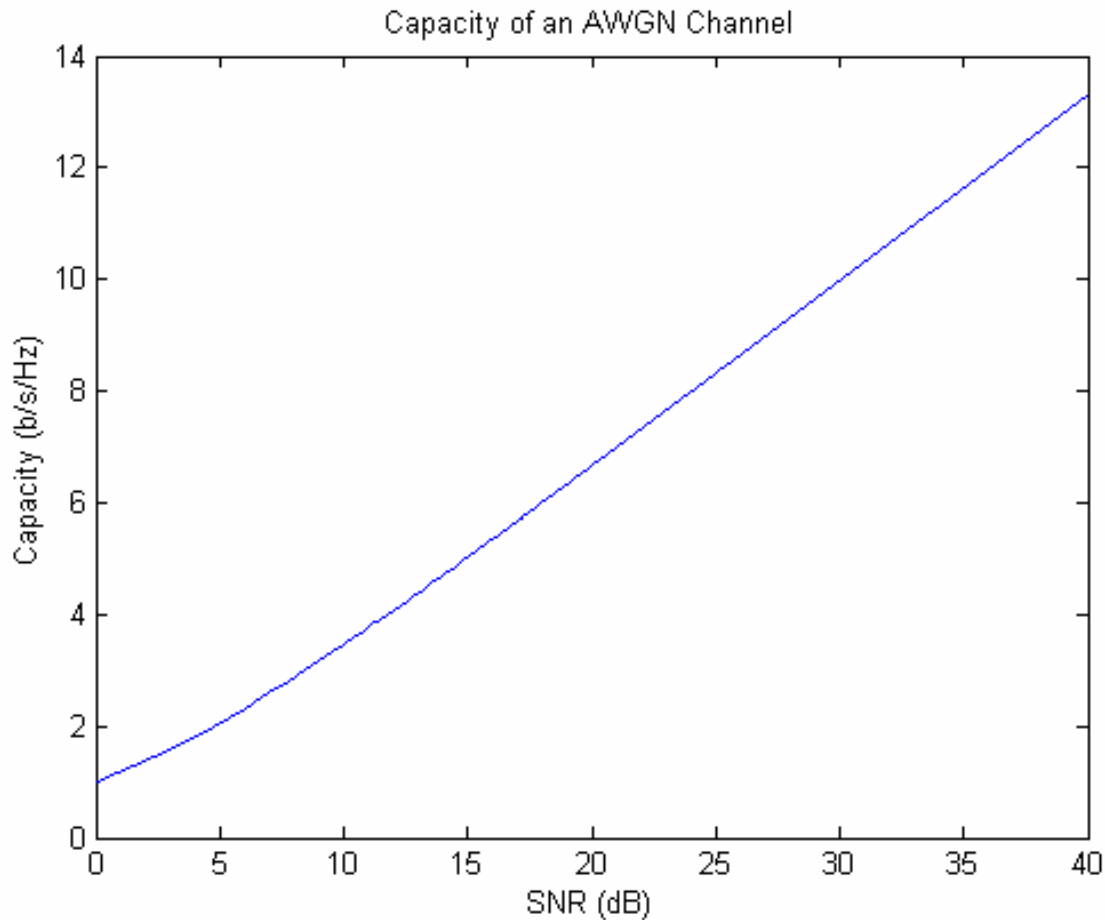
$$C = W \log \left(1 + \frac{P_{av}}{WN_o} \right)$$

Plot of Capacity for AWGN Channel



- Capacity increases with available bandwidth W

Plot of Capacity for AWGN Channel



$$\frac{C}{W} = \log_2 \left(1 + \frac{P_{av}}{WN_0} \right) \text{ bits/second/Hz}$$

Capacity increases with SNR



Capacity Limits

- The normalized capacity can be expressed as

$$\frac{C}{W} = \log_2 \left(1 + \frac{P_{av}}{WN_0} \right) \text{ bits/second/Hz}$$

- However, $P_{av} = C E_b$

$$\frac{C}{W} = \log_2 \left(1 + \frac{C E_b}{W N_0} \right) \text{ bits/second/Hz}$$

- Solving for E_b/N_0

$$\frac{E_b}{N_0} = \frac{2^{\frac{C}{W}} - 1}{C/W} \text{ bits/second/Hz}$$

- Which says that E_b/N_0 increases exponentially with C/W

- Finally,

$$\lim_{C/W \rightarrow 0} \frac{E_b}{N_0} = \lim_{C/W \rightarrow 0} \frac{2^{\frac{C}{W}} - 1}{C/W} = \ln 2$$

Minimum E_b/N_0 which will achieve capacity is $\ln(2) = 0.639$ or -



Can Modulation Alone Achieve Channel Capacity?

- Standard modulation falls well short of channel capacity
- If we use a very large dimension for our signal constellation it may be possible to achieve channel capacity.
 - M -ary FSK achieves channel capacity (but is very bandwidth inefficient).



M-ary Orthogonal Modulation

- The probability of error using the Union Bound for M -ary orthogonal modulation (e.g., M-FSK) is

$$M=2^k$$

$$P_s(e) \leq (M - 1)Q\left(\sqrt{\frac{E_b \log_2 M}{N_o}}\right) < 2^k e^{-\frac{E_b k}{2N_o}} = e^{-k\frac{E_b}{2N_o} - k \ln(2)}$$

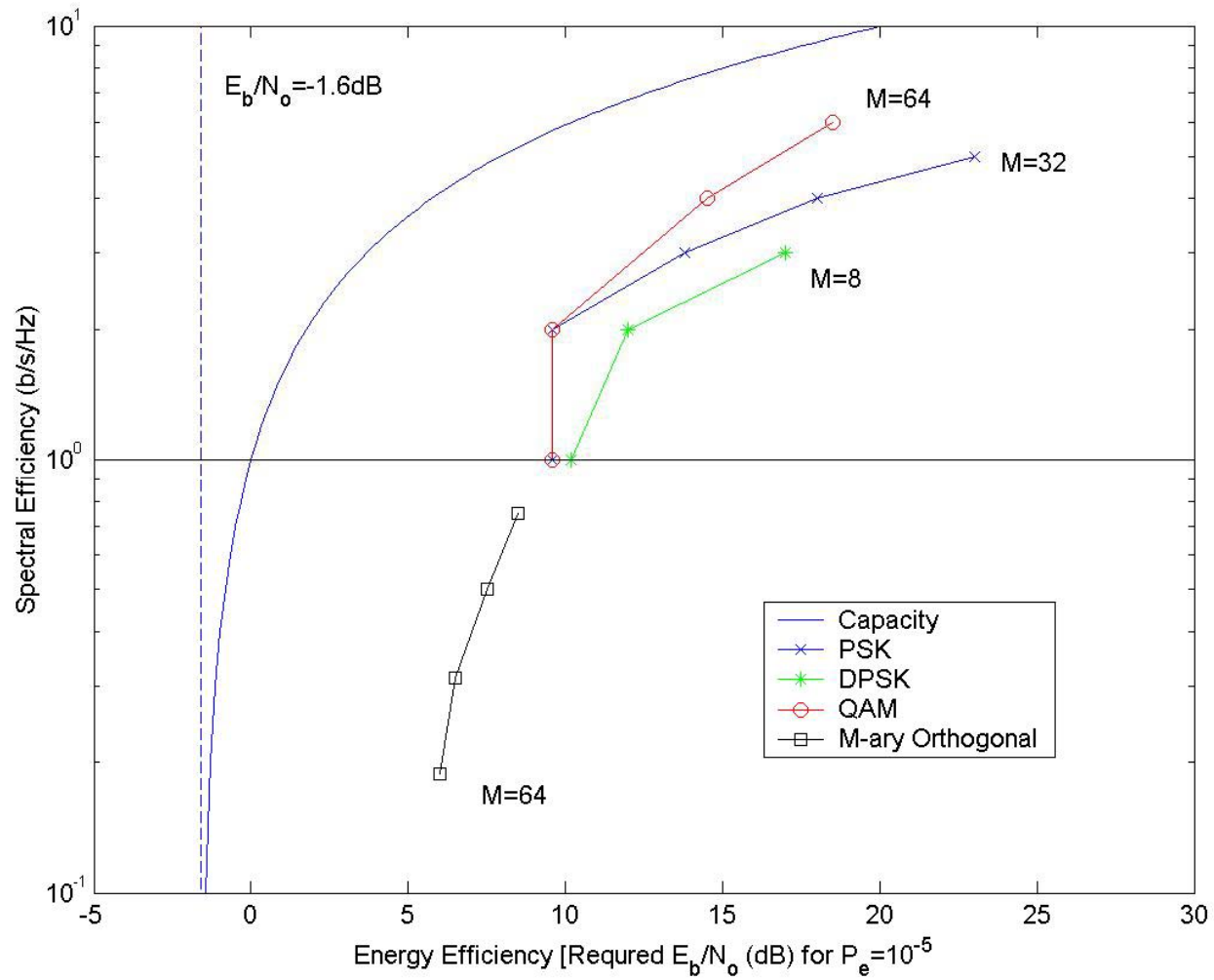
- which shows that as $M \rightarrow \textit{infinity}$, the probability of error goes to zero provided that $E_b/N_o > 2\ln(2)=1.39$
- However, more sophisticated bounds show that

$$P_s(e) < 2e^{-\frac{E_b k}{2N_o}} = e^{-k(\sqrt{E_b/N_o} - \sqrt{\ln 2})^2}$$

- which shows that as $M \rightarrow \textit{infinity}$, the probability of error goes to zero provided that $E_b/N_o > \ln(2)=0.69$ or -1.6dB



Capacity

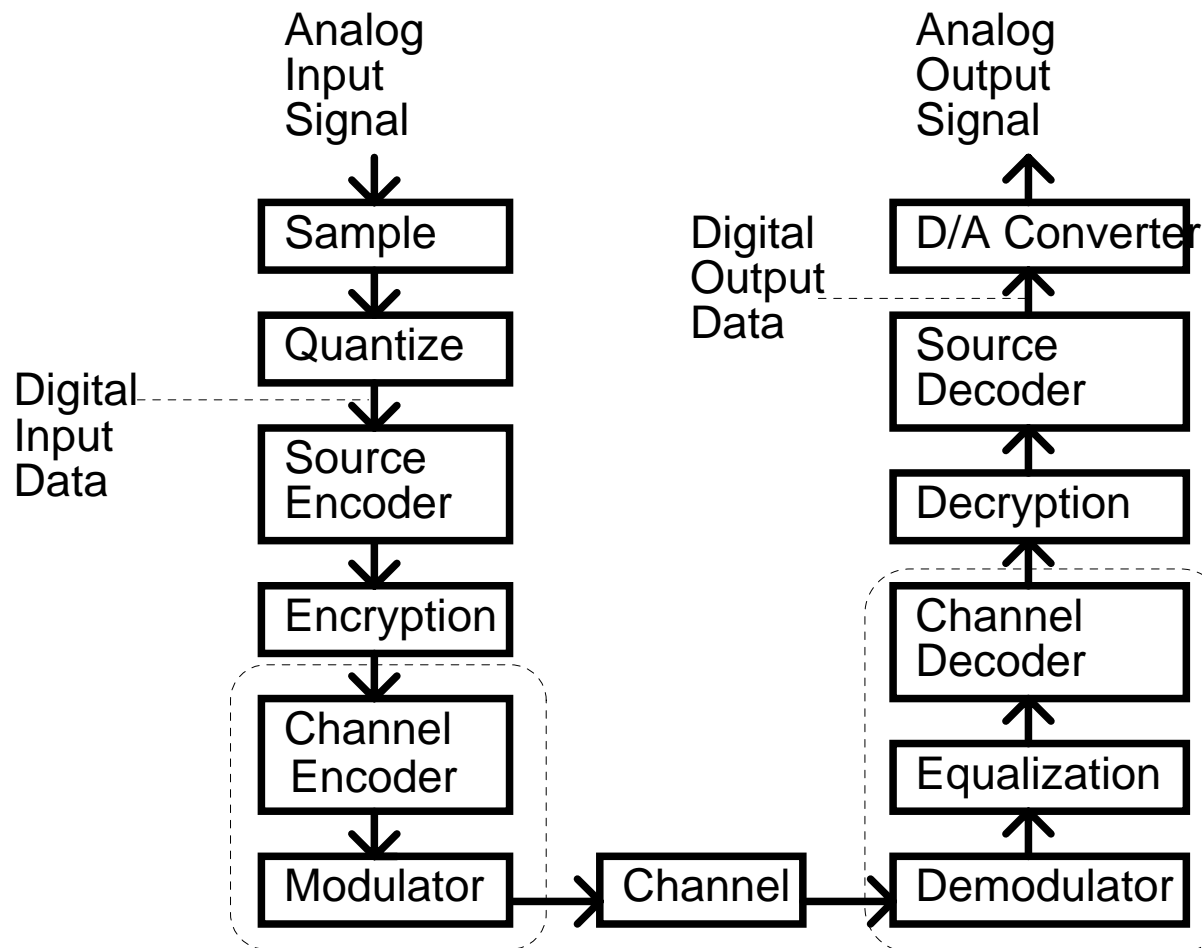




Channel Coding

- Most digital communications systems attempt to achieve channel capacity by using “error correction coding” or “channel coding” in addition to modulation.
- Channel codes selectively introduce redundant information into the transmitted data stream so that a small number of symbol errors can be corrected after the demodulator
- Channel coding improves the error performance (energy efficiency). However, we are trading for a reduction in data rate (bandwidth efficiency).
- However, the bandwidth efficiency is much better than high order M -ary modulation.

Block Diagram of Digital Communications System





Problems Associated with Channel Coding

- Although the idea of channel capacity has been known since the late 1940's, the theory did not explain how to design good codes.
- It is difficult to design a code which performs well **AND** can be easily decoded by the receiver.
- Over the last 50 years, there has been steady progress in designing good error correction codes.
- The fundamental structure of channel codes is covered in courses on Coding Theory. The intelligent design of good codes requires an understanding of finite field arithmetic.
- We focus on the selection of channel codes for the design of a communication system.



Major Types of Channel Codes

- Block Codes
 - Convolutional Codes
 - Trellis Codes (Trellis Coded Modulation)
 - Turbo Codes
-
- We consider Block Codes first.



References

- Digital Communications books with excellent introductions to error-correction coding:
 - J. G. Proakis, Digital Communications 2nd ed, McGraw-Hill, 1989.
 - R. E. Ziemer and R. L. Peterson, Introduction to Digital Communication, Macmillan, 1992.
 - A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding, McGraw-Hill, 1979.



Coding References (Continued)

- General-Texts on Coding Theory:
 - S. Lin and D. J. Costello, Error Control Coding: Fundamentals and Applications, Prentice-Hall, 1983.
 - R. Blahut, Theory and Practice of Error Control Codes, Addison-Wesley, 1983.
- Papers on Modem Design:
 - G. Ungerboeck, "Channel Coding with multilevel/phase signals," *IEEE Trans. Information Theory*, January 1982.
 - G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets, Parts I & II," *IEEE Communications Magazine*, February 1987.
 - *IEEE Communications Magazine*, December 1991.



Block Codes

- The encoder for a block code accepts blocks of k input symbols and produces blocks of n output symbols.
- Usually the input symbols and output symbols, are both bits $\{0,1\}$, but there is one notable exception (Reed-Solomon Codes) for which input and output symbols are M -ary.
- Note that there will be $n - k$ redundant symbols.
- The ratio $r = k/n \leq 1$ is called the code rate.
 - Small r -> lots of redundancy
 - Large r -> little redundancy



Typical Values

- For a practical block codes, $20 \leq n \leq 1000$ is a typical range of values.
- For practical block codes, $\frac{1}{4} \leq r \leq 1$ is typical.
- We will refer to an “ (n, k) ” block code.

Example Block Code: (7,4) Hamming Code

Input Bits				Output Bits						
x_4	x_3	x_2	x_1	c_7	c_6	c_5	c_4	c_3	c_2	c_1
0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	1	0	0	0	1
0	0	1	0	1	1	1	0	0	1	0
0	0	1	1	0	1	0	0	0	1	1
0	1	0	0	0	1	1	0	1	0	0
0	1	0	1	1	1	0	0	1	0	1
0	1	1	0	0	0	1	0	1	1	0
0	1	1	1	1	1	0	0	1	1	1
1	0	0	0	1	1	0	1	0	0	0
1	0	0	1	0	1	1	1	0	0	1
1	0	1	0	0	0	1	1	0	1	0
1	0	1	1	1	0	0	1	0	1	1
1	1	0	0	1	0	1	1	1	0	0
1	1	0	1	0	0	0	1	1	0	1
1	1	1	0	1	1	1	1	1	1	0
1	1	1	1	0	1	0	1	1	1	1



Properties of Block Codes

- Because decoding is generally the difficult problem, most block codes of interest have structure to them.
- A code $C = \{c_1, c_2, \dots, c_{2^k}\}$ is linear if:
$$c_1 \in C, c_2 \in C \Rightarrow c_1 \oplus c_2 \in C$$
 ,
where \oplus denotes modulo-2 bitwise addition.
- Example: The (7,4) Hamming Code is linear.

$$\underline{x} = (0001) \Rightarrow \underline{y} = (1010001)$$

$$\underline{x} = (0110) \Rightarrow \underline{y} = (0010110)$$

$$\underline{x} = (0111) \Rightarrow \underline{y} = (1000111)$$

Properties of Block Codes (continued)

- A code $C = \{c_1, c_2, \dots, c_{2^k}\}$ is systematic if there are k bits of the codeword which correspond directly to information bits.
- Example, the (7,4) Hamming Code is systematic:
 $c_1 = x_1, c_2 = x_2, c_3 = x_3, c_4 = x_4$
- We can think of the remaining bits as just a fancy system of parity checks:

$$c_5 = x_1 \oplus x_2 \oplus x_3$$

$$c_6 = x_2 \oplus x_3 \oplus x_4$$

$$c_7 = x_1 \oplus x_2 \oplus x_4$$

Properties of Block Codes (continued)

- A code $C = \{c_1, c_2, \dots, c_{2^k}\}$ is cyclic if:
 $(c_1, c_2, \dots, c_n) \in C \Rightarrow (c_n, c_1, \dots, c_{n-1}) \in C$
- Example: the (7,4) Hamming Code is cyclic.

$$\underline{x} = (0001) \Rightarrow \underline{c} = (1010001)$$

$$\underline{x} = (1000) \Rightarrow \underline{c} = (1101000)$$

$$\underline{x} = (0100) \Rightarrow \underline{c} = (0110100)$$

$$\underline{x} = (1010) \Rightarrow \underline{c} = (0011010)$$

$$\underline{x} = (1101) \Rightarrow \underline{c} = (0001101)$$

$$\underline{x} = (0110) \Rightarrow \underline{c} = (1000110)$$

$$\underline{x} = (0011) \Rightarrow \underline{c} = (0100011)$$

- Most practical block codes are linear and cyclic.



Distance Properties of a Block Code

- The Hamming Distance between two codewords $c_1 \in C$ and $c_2 \in C$ is the number of bits in which they differ:

$$d_H(c_1, c_2) = \sum_{i=1}^n c_{1,i} \oplus c_{2,i}$$

- Example: $d_H((1010001), (1101000)) = 4$
- The minimum distance $d_{H,\min}$ of a code is the smallest distance separating any two codewords:

$$d_{H,\min} = \min_{i \neq j} \{d_H(c_i, c_j)\}$$

- Example: $d_{H,\min} = 3$ for (7,4) Hamming Code



Decoding of Block Codes

- For a linear error correction code, $d_{H,\min}$ is the smallest weight of any nonzero codeword.
- Just as we wanted to make Euclidean distance large for modulation, we want to make Hamming distance large for our block codes.
- The decoder's job will be to choose the codeword most closely resembling the received sequence of bits.
 - Example: Suppose we transmit: $\underline{c} = (0100011)$ but receive $\underline{y} = (0100001)$ The closest match is: $\hat{\underline{c}} = (0100011)$ so we estimate that our data bits were: $\hat{\underline{x}} = (0011)$



Error Correction Capability

- Any code with minimum distance $d_{H,\min}$, can correct any combination of up to $t = \left\lfloor \frac{d_{H,\min} - 1}{2} \right\rfloor$ errors.
- We call t the error correcting capability of the code.
- There is at least one combination of $t+1$ errors which will cause an error.
- Any code with minimum distance $d_{H,\min}$ can detect any combination of up to $d_{H,\min} - 1$ errors by the channel.



Preview of Error Probability

- Assume $P_b(e)$ is the probability of bit error for the modulation on the channel, and we can assume that bit errors occur independently.
- The probability of codeword failure for an error correction code will be:

$$\begin{aligned} P_E &= 1 - \sum_{i=0}^t \binom{n}{i} P_b(e)^i (1 - P_b(e))^{n-i} \\ &= \sum_{i=t+1}^n \binom{n}{i} P_b(e)^i (1 - P_b(e))^{n-i} \end{aligned}$$