Imperial College
London

EE2-4: Communication Systems

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Course Information

- Lecturer: Dr. Cong Ling (Senior Lecturer)
  – Office: Room 815, EE Building
  – Phone: 020-7594 6214
  – Email: c.ling@imperial.ac.uk
- Handouts
  – Slides: exam is based on slides
  – Notes: contain more details
  – Problem sheets
- Course homepage
  – http://www.commsp.ee.ic.ac.uk/~cling
  – You can access lecture slides/problem sheets/past papers
  – Also available in Blackboard
- Grading
  – 1.5-hour exam, no-choice, closed-book

Lectures

Introduction and background

1. Introduction
2. Probability and random processes
3. Noise

Effects of noise on analog communications

4. Noise performance of DSB
5. Noise performance of SSB and AM
6. Noise performance of FM
7. Pre/de-emphasis for FM and comparison of analog systems

Digital communications

8. Digital representation of signals
9. Baseband digital transmission
10. Digital modulation
11. Noncoherent demodulation

Information theory

12. Entropy and source coding
13. Channel capacity
14. Block codes
15. Cyclic codes

EE2-4 vs. EE1-6

- Introduction to Signals and Communications
  - How do communication systems work?
  - About modulation, demodulation, signal analysis...
  - The main mathematical tool is the Fourier transform for deterministic signal analysis.
  - More about analog communications (i.e., signals are continuous).

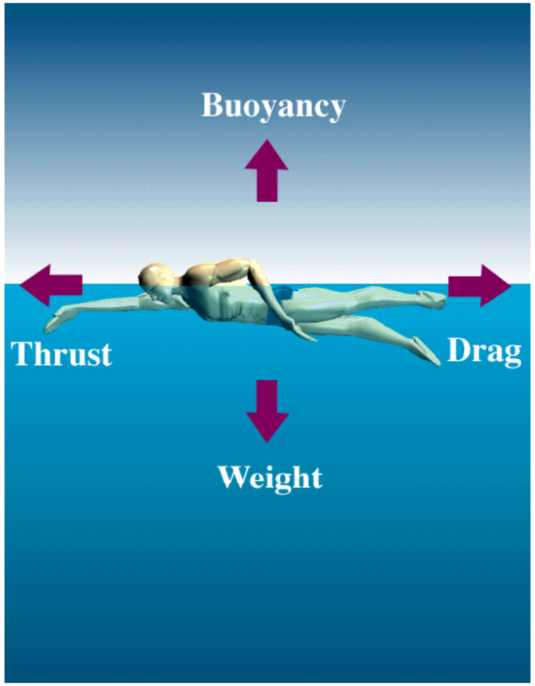
- Communications
  - How do communication systems perform in the presence of noise?
  - About statistical aspects and noise.
    - This is essential for a meaningful comparison of various communications systems.
  - The main mathematical tool is probability.
  - More about digital communications (i.e., signals are discrete).

Learning Outcomes

- Describe a suitable model for noise in communications
- Determine the signal-to-noise ratio (SNR) performance of analog communications systems
- Determine the probability of error for digital communications systems
- Understand information theory and its significance in determining system performance
- Compare the performance of various communications systems

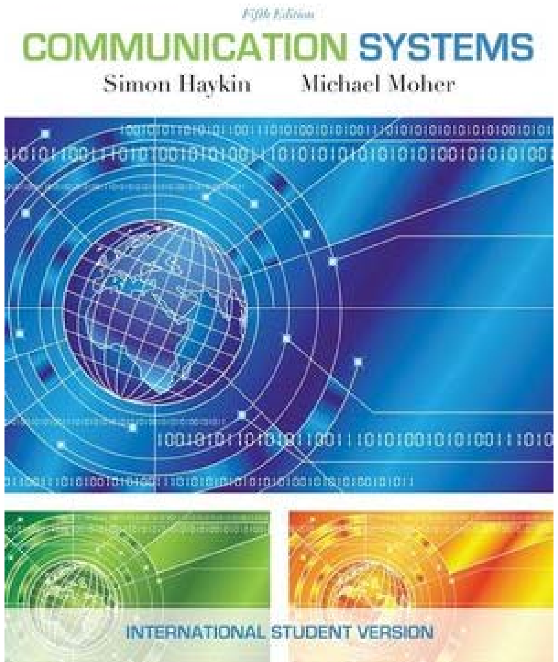
About the Classes

- You're welcome to ask questions.
  - You can interrupt me at any time.
  - Please don't disturb others in the class.
- Our responsibility is to facilitate you to learn. You have to make the effort.
- Spend time reviewing lecture notes afterwards.
- If you have a question on the lecture material after a class, then
  - Look up a book! Be resourceful.
  - Try to work it out yourself.
  - Ask me during the problem class or one of scheduled times of availability.



References

- C. Ling, Notes of Communication Systems, Imperial Collage.
- S. Haykin & M. Moher, Communication Systems, 5th ed., International Student Version, Wiley, 2009 (£43.99 from Wiley)
- S. Haykin, Communication Systems, 4th ed., Wiley, 2001
  - Owns the copyright of many figures in these slides.
  - For convenience, the note "© 2000 Wiley, Haykin/Communication System, 4th ed." is not shown for each figure.
- B.P. Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Oxford University Press, 1998
- J.G. Proakis and M. Salehi, Communication Systems Engineering, Prentice-Hall, 1994
- L.W. Couch II, Digital and Analog Communication Systems, 6th ed., Prentice-Hall, 2001



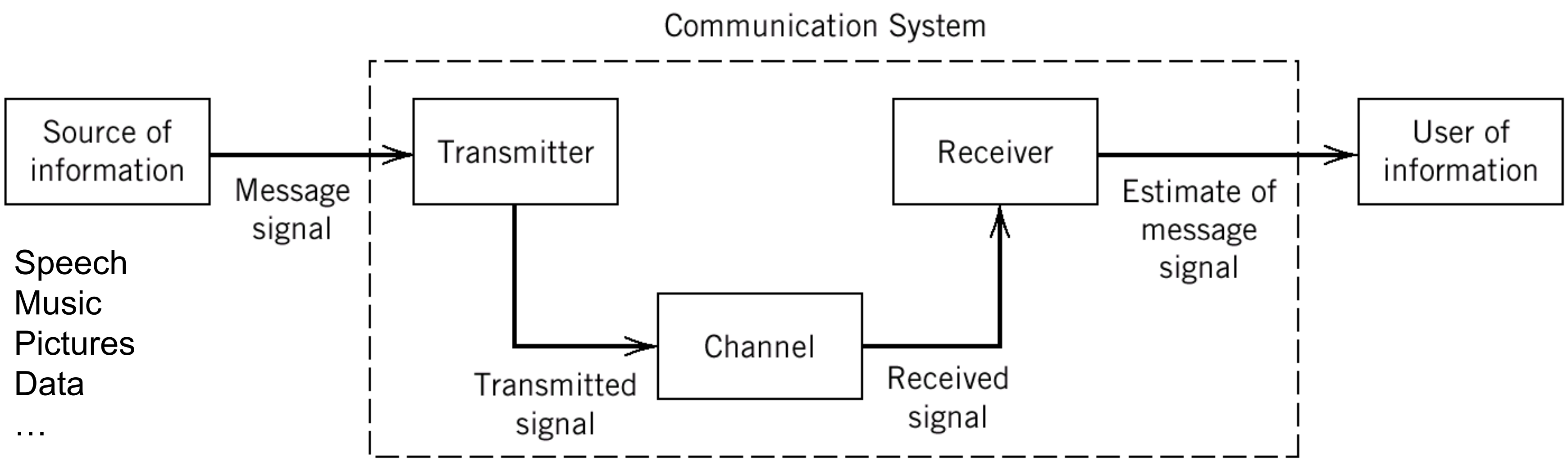
Multitude of Communications

- Telephone network
- Internet
- Radio and TV broadcast
- Mobile communications
- Wi-Fi
- Satellite and space communications
- Smart power grid, healthcare…

- Analogue communications
  - AM, FM
- Digital communications
  - Transfer of information in digits
  - Dominant technology today
  - Broadband, 3G, DAB/DVB

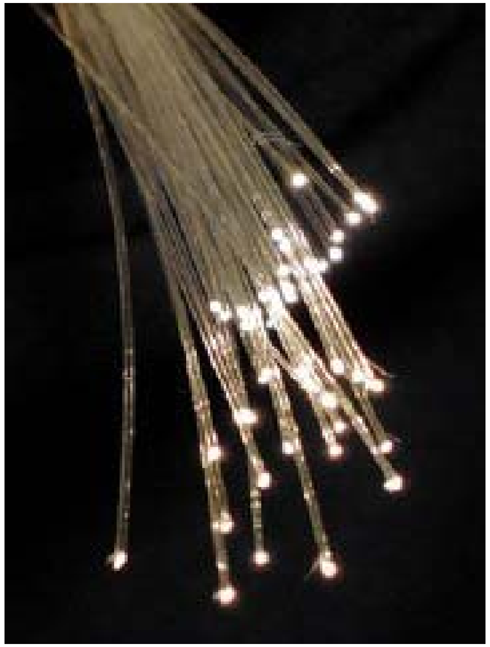
What's Communications?

- Communication involves the transfer of information from one point to another.
- Three basic elements
  - **Transmitter**: converts message into a form suitable for transmission
  - **Channel**: the physical medium, introduces distortion, noise, interference
  - **Receiver**: reconstruct a recognizable form of the message



Communication Channel

- The channel is central to operation of a communication system
  - Linear (e.g., mobile radio) or nonlinear (e.g., satellite)
  - Time invariant (e.g., fiber) or time varying (e.g., mobile radio)
- The information-carrying capacity of a communication system is proportional to the channel bandwidth
- Pursuit for wider bandwidth
  - Copper wire: 1 MHz
  - Coaxial cable: 100 MHz
  - Microwave: GHz
  - Optical fiber: THz
    - Uses light as the signal carrier
    - Highest capacity among all practical signals



Noise in Communications

- Unavoidable presence of noise in the channel
  - Noise refers to unwanted waves that disturb communications
  - Signal is contaminated by noise along the path.
- External noise: interference from nearby channels, human-made noise, natural noise...
- Internal noise: thermal noise, random emission... in electronic devices
- Noise is one of the basic factors that set limits on communications.
- A widely used metric is the signal-to-noise (power) ratio (SNR)

$$\text{SNR} = \frac{\text{signal power}}{\text{noise power}}$$

Transmitter and Receiver

- The transmitter modifies the message signal into a form suitable for transmission over the channel
- This modification often involves **modulation**
  - Moving the signal to a high-frequency carrier (up-conversion) and varying some parameter of the carrier wave
  - Analog: AM, FM, PM
  - Digital: ASK, FSK, PSK (SK: shift keying)
- The receiver recreates the original message by **demodulation**
  - Recovery is not exact due to noise/distortion
  - The resulting degradation is influenced by the type of modulation
- Design of analog communication is conceptually simple
- Digital communication is more efficient and reliable; design is more sophisticated

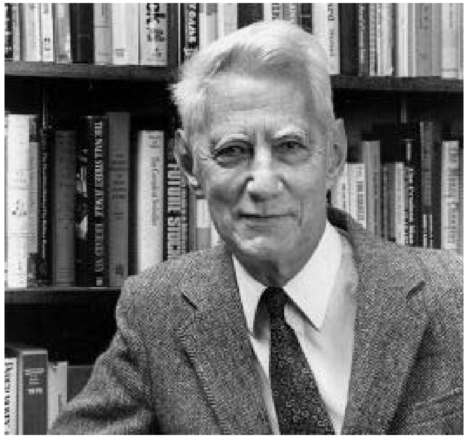
Objectives of System Design

- Two primary resources in communications
  - Transmitted **power** (should be green)
  - Channel **bandwidth** (very expensive in the commercial market)
- In certain scenarios, one resource may be more important than the other
  - Power limited (e.g. deep-space communication)
  - Bandwidth limited (e.g. telephone circuit)
- Objectives of a communication system design
  - The message is delivered both efficiently and reliably, subject to certain design constraints: power, bandwidth, and cost.
  - **Efficiency** is usually measured by the amount of messages sent in unit power, unit time and unit bandwidth.
  - **Reliability** is expressed in terms of SNR or probability of error.

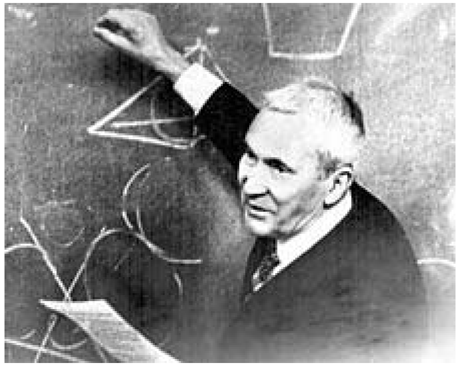
Information Theory

- In digital communications, is it possible to operate at zero error rate even though the channel is noisy?
- **Poineers: Shannon, Kolmogorov…**
  - The maximum rate of reliable transmission is calculated.
  - The famous **Shannon capacity formula** for a channel with bandwidth W (Hz)

    C = W log(1+SNR) bps (bits per second)
  - Zero error rate is possible as long as actual signaling rate is less than C.
- Many concepts were fundamental and paved the way for future developments in communication theory.
  - Provides a basis for tradeoff between SNR and bandwidth, and for comparing different communication schemes.



Shannon



Kolmogorov

Milestones in Communications

- 1837, Morse code used in telegraph
- 1864, Maxwell formulated the eletromagnetic (EM) theory
- 1887, Hertz demonstrated physical evidence of EM waves
- 1890's-1900's, **Marconi** & Popov, long-distance radio telegraph
  - Across Atlantic Ocean
  - From Cornwall to Canada

- 1875, Bell invented the telephone
- 1906, radio broadcast
- 1918, Armstrong invented superheterodyne radio receiver (and FM in 1933)
- 1921, land-mobile communication

Milestones (2)

- 1928, Nyquist proposed the sampling theorem
- 1947, microwave relay system
- 1948, information theory
- 1957, era of satellite communication began
- 1966, **Kuen Kao** pioneered fiber-optical communications (Nobel Prize Winner)
- 1970's, era of computer networks began
- 1981, analog cellular system
- 1988, digital cellular system debuted in Europe
- 2000, 3G network
- The big 3 telecom manufacturers in 2010

CISCO ERICSSON HUAWEI

Cellular Mobile Phone Network

- A large area is partitioned into cells
- Frequency reuse to maximize capacity

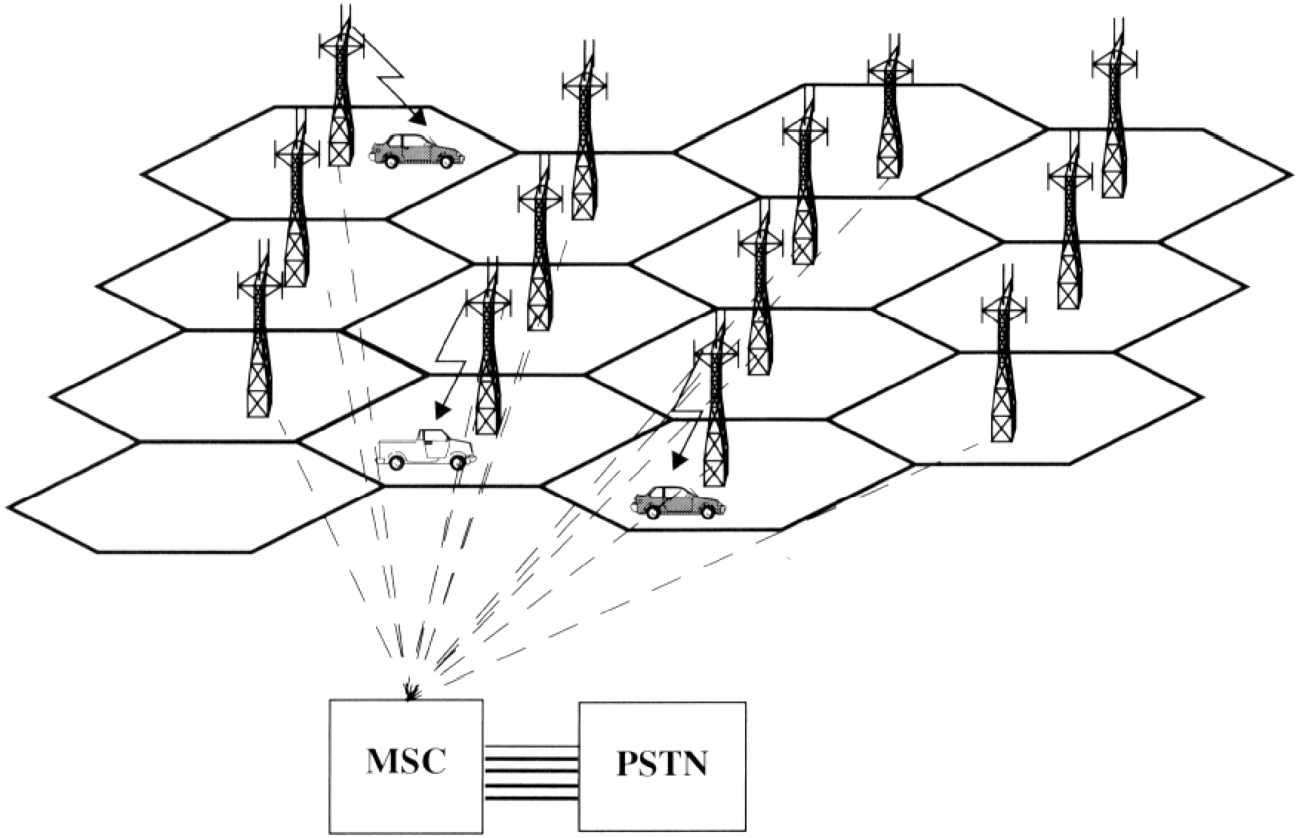
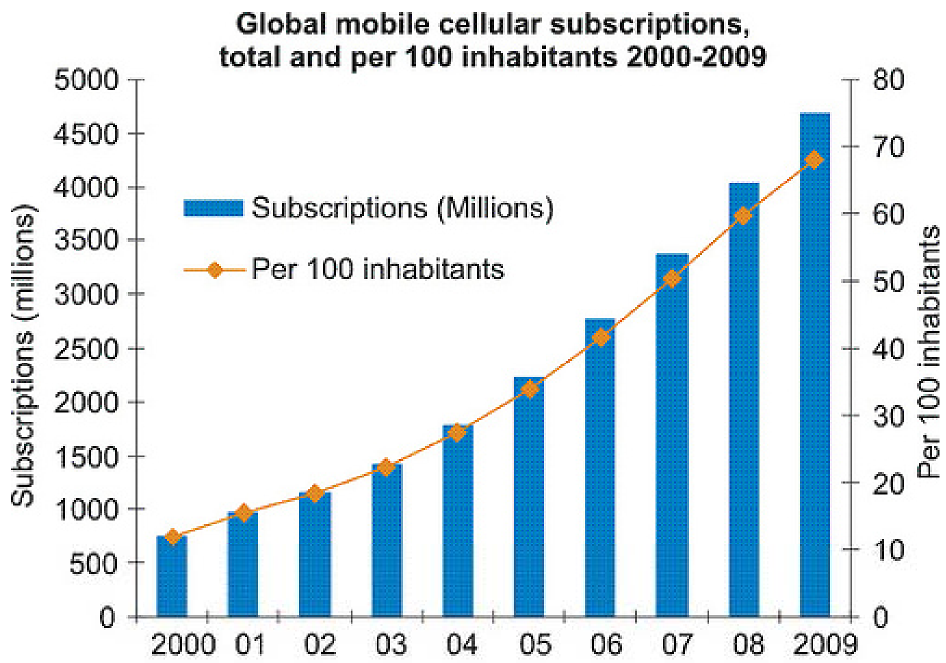


Figure 1.5 A cellular system. The towers represent base stations which provide radio access between mobile users and the mobile switching center (MSC).

Growth of Mobile Communications

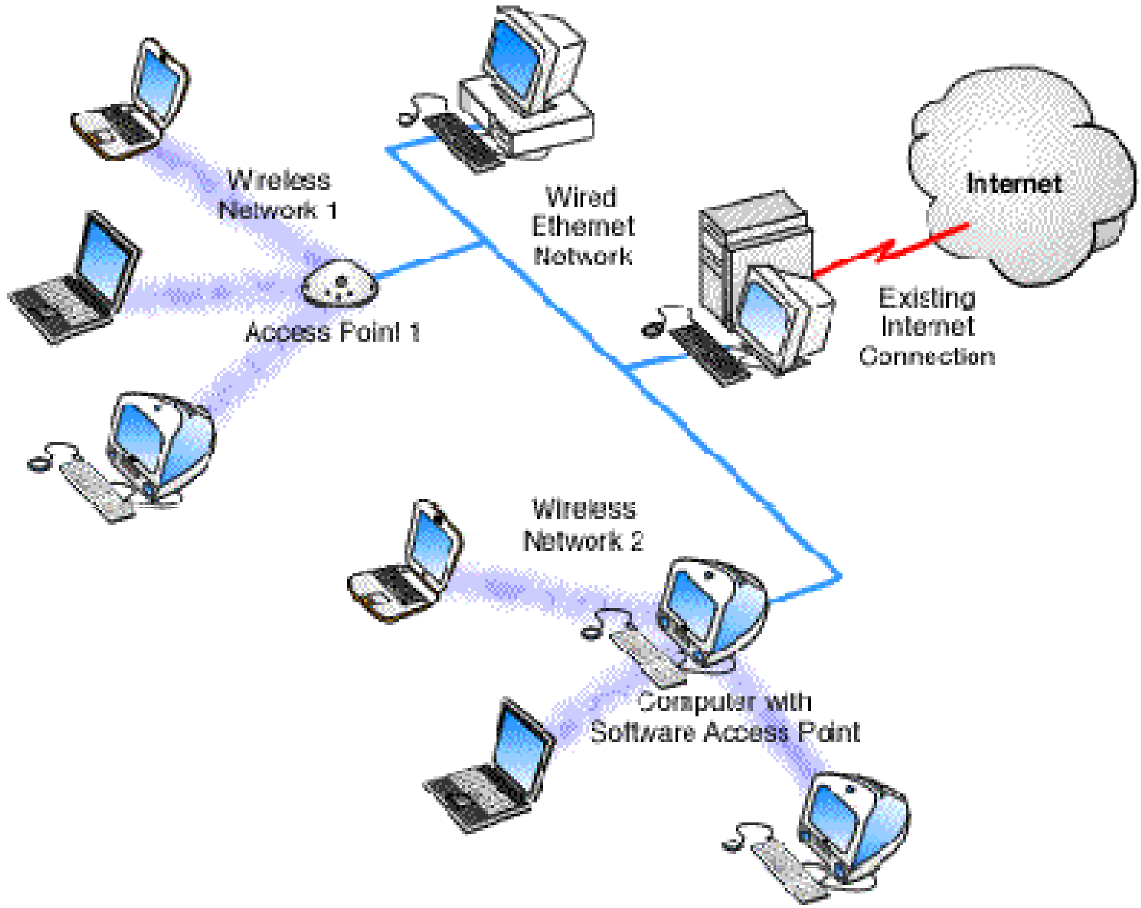
- 1G: analog communications
  - AMPS
- 2G: digital communications
  - GSM
  - IS-95
- 3G: CDMA networks
  - WCDMA
  - CDMA2000
  - TD-SCDMA
- 4G: data rate up to 1 Gbps (giga bits per second)
  - Pre-4G technologies: WiMax, 3G LTE



Source: ITU World Telecommunication/ICT Indicators database.

Wi-Fi

- Wi-Fi connects “local” computers (usually within 100m range)

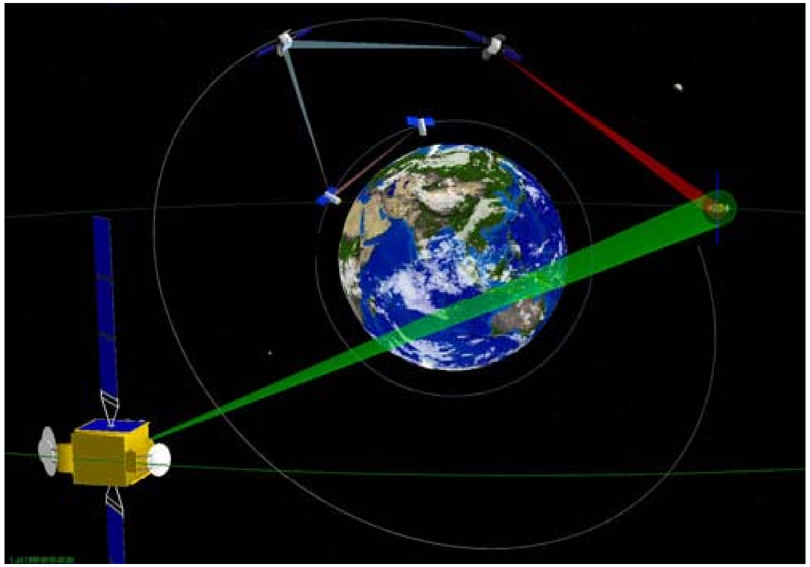


IEEE 802.11 Wi-Fi Standard

- 802.11b
  - Standard for 2.4GHz (unlicensed) ISM band
  - 1.6-10 Mbps, 500 ft range
- 802.11a
  - Standard for 5GHz band
  - 20-70 Mbps, variable range
  - Similar to HiperLAN in Europe
- 802.11g
  - Standard in 2.4 GHz and 5 GHz bands
  - Speeds up to 54 Mbps, based on orthogonal frequency division multiplexing (OFDM)
- 802.11n
  - Data rates up to 600 Mbps
  - Use multi-input multi-output (MIMO)

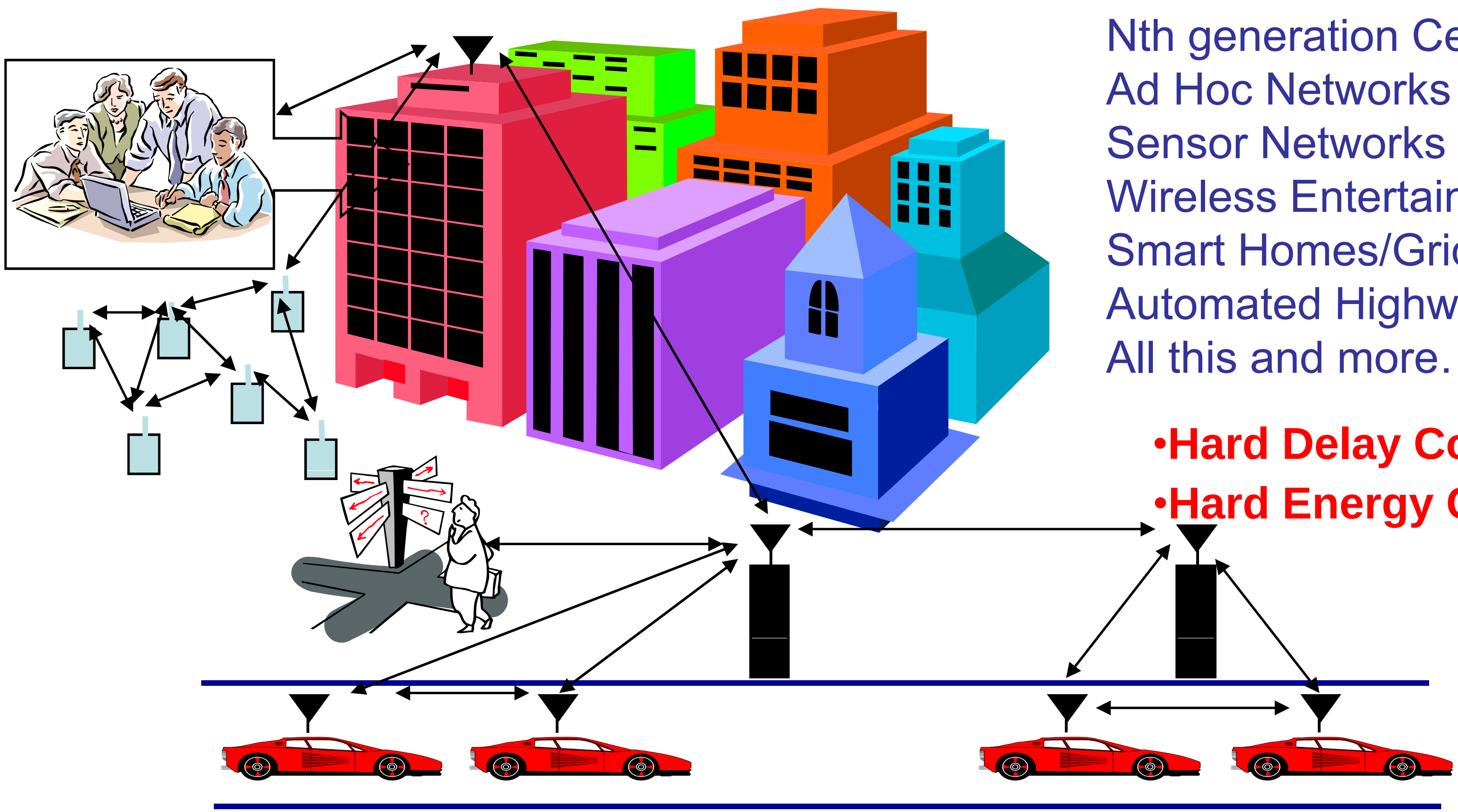
Satellite/Space Communication

- Satellite communication
  - Cover very large areas
  - Optimized for one-way transmission
    - Radio (DAB) and movie (SatTV) broadcasting
  - Two-way systems
    - The only choice for remote-area and maritime communications
    - Propagation delay (0.25 s) is uncomfortable in voice communications
- Space communication
  - Missions to Moon, Mars, …
  - Long distance, weak signals
  - High-gain antennas
  - Powerful error-control coding



Future Wireless Networks

Ubiquitous Communication Among People and Devices

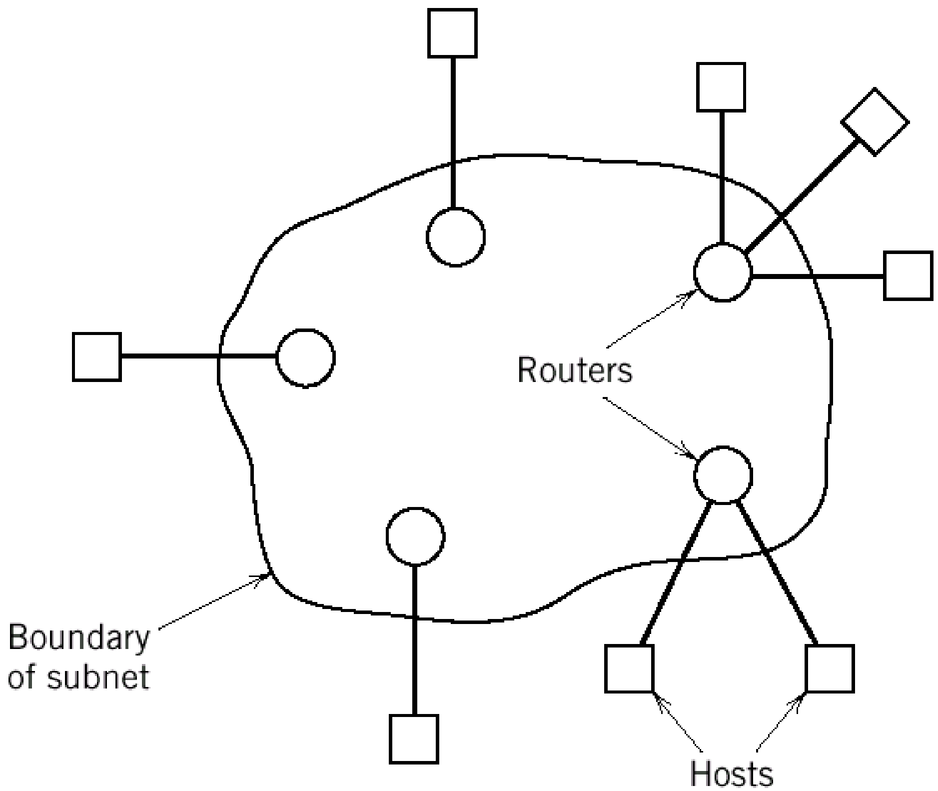


Wireless Internet access
Nth generation Cellular
Ad Hoc Networks
Sensor Networks
Wireless Entertainment
Smart Homes/Grids
Automated Highways
All this and more…

- **Hard Delay Constraints**
- **Hard Energy Constraints**

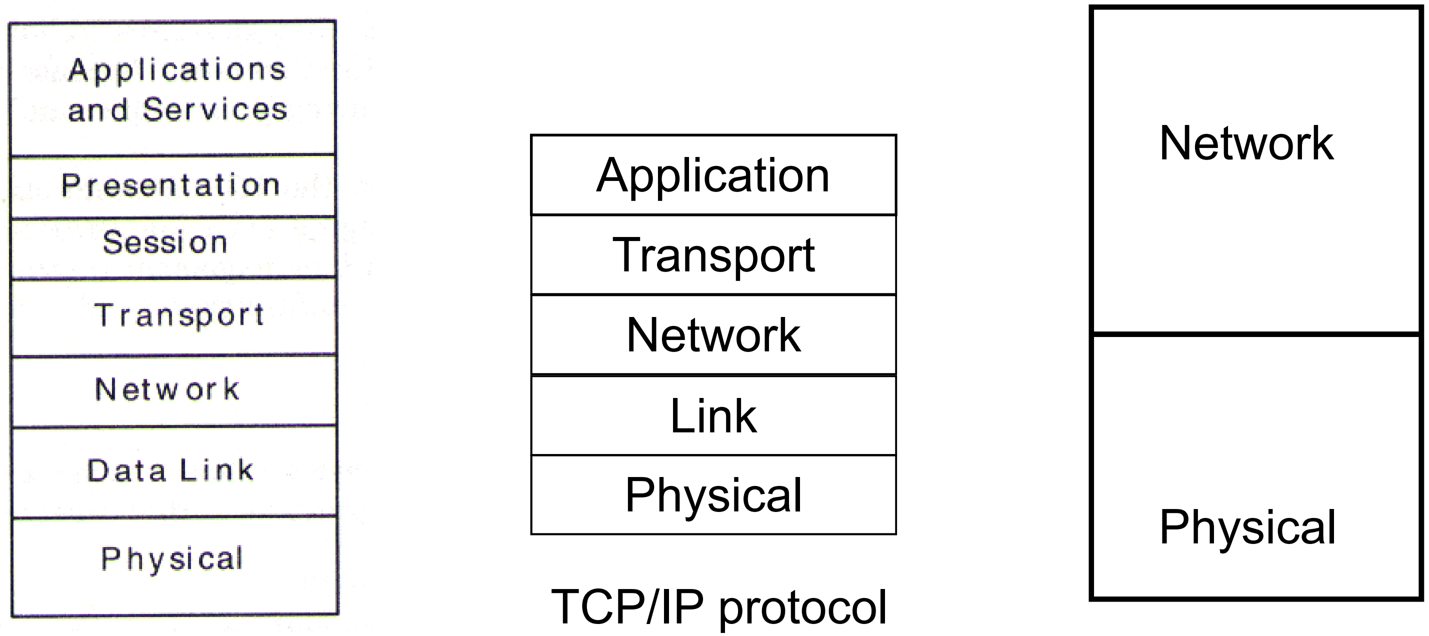
Communication Networks

- Today’s communications networks are complicated systems
  - A large number of users sharing the medium
  - Hosts: devices that communicate with each other
  - Routers: route date through the network



Concept of Layering

- Partitioned into layers, each doing a relatively simple task
- Protocol stack



OSI Model

TCP/IP protocol stack (Internet)

2-layer model

Communication Systems mostly deals with the physical layer, but some techniques (e.g., coding) can also be applied to the network layer.

Imperial College
London

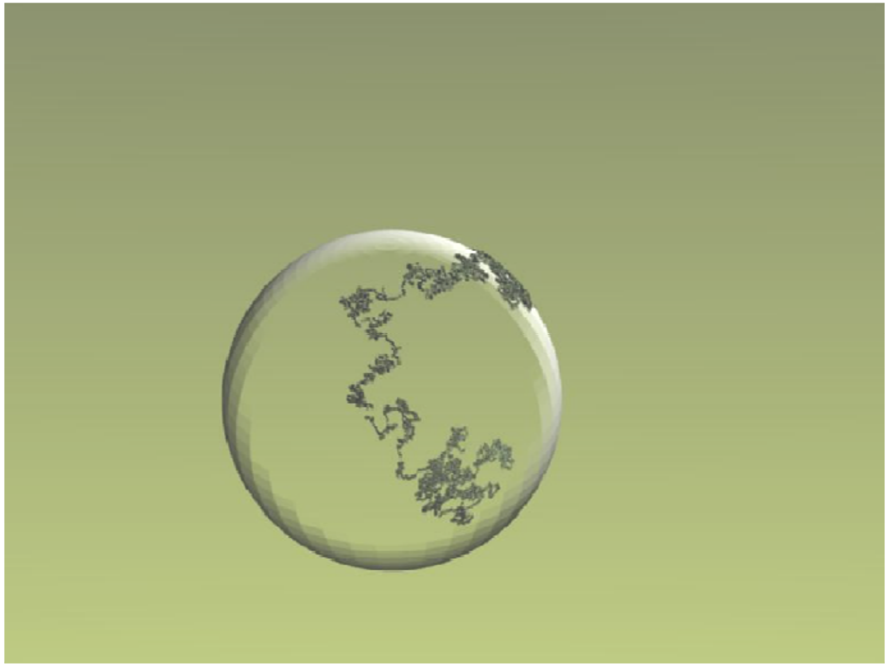
EE2-4: Communication Systems

Lecture 2: Probability and Random Processes

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline



- Probability
  - How probability is defined
  - cdf and pdf
  - Mean and variance
  - Joint distribution
  - Central limit theorem
- Random processes
  - Definition
  - Stationary random processes
  - Power spectral density
- References
  - Notes of Communication Systems, Chap. 2.3.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 5
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 11

Why Probability/Random Process?

- Probability is the core mathematical tool for communication theory.
- The stochastic model is widely used in the study of communication systems.
- Consider a radio communication system where the received signal is a random process in nature:
  - Message is random. No randomness, no information.
  - Interference is random.
  - Noise is a random process.
  - And many more (delay, phase, fading, ...)
- Other real-world applications of probability and random processes include
  - Stock market modelling, gambling (Brown motion as shown in the previous slide, random walk)…

Probabilistic Concepts

- What is a random variable (RV)?
  - It is a variable that takes its values from the outputs of a random experiment.
- What is a random experiment?
  - It is an experiment the outcome of which cannot be predicted precisely.
  - All possible identifiable outcomes of a random experiment constitute its ***sample space*** S.
  - An ***event*** is a collection of possible outcomes of the random experiment.
- Example
  - For tossing a coin, S = { H, T }
  - For rolling a die, S = { 1, 2, …, 6 }

Probability Properties

- $P_X(x_i)$: the *probability* of the random variable X taking on the value x_i
- The probability of an event to happen is a non-negative number, with the following properties:
  - The probability of the event that includes all possible outcomes of the experiment is 1.
  - The probability of two events that do not have any common outcome is the sum of the probabilities of the two events separately.
- Example
  - Roll a die: $P_X(x = k) = 1/6$ for $k = 1, 2, \ldots, 6$

CDF and PDF

- The **(cumulative) distribution function (cdf)** of a random variable X is defined as the probability of X taking a value less than the argument x:

$$F_X(x) = P(X \le x)$$

- Properties

$$F_X(-\infty) = 0, \quad F_X(\infty) = 1$$

$$F_X(x_1) \le F_X(x_2) \quad \text{if } x_1 \le x_2$$

- The **probability density function (pdf)** is defined as the derivative of the distribution function:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$

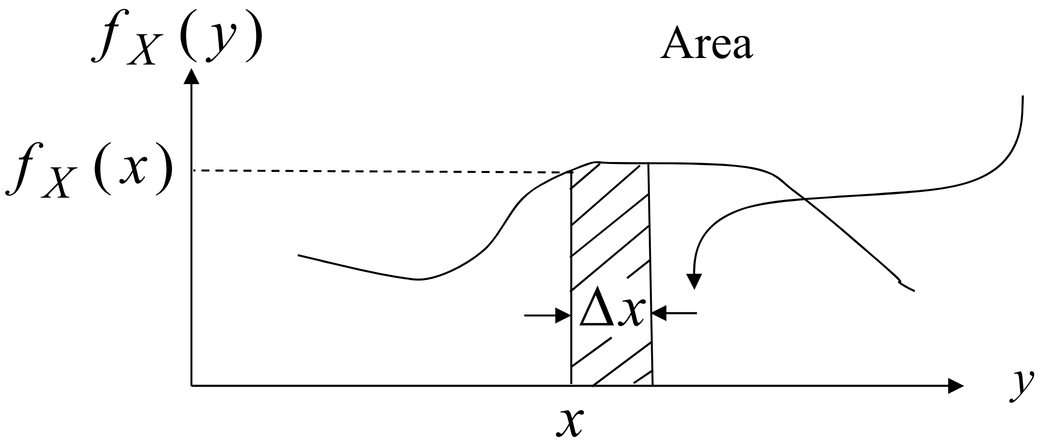
$$P(a < X \le b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(y)dy$$

$$f_X(x) = \frac{dF_X(x)}{dx} \ge 0 \ \text{ since } F_X(x) \text{ is non-decreasing}$$

Mean and Variance

- If Δx is sufficiently small,

$$P(x < X \le x + \Delta x) = \int_{x}^{x+\Delta x} f_X(y)dy \cong \underline{f_X(x)\Delta x}$$



- Mean (or expected value ⇔ DC level):

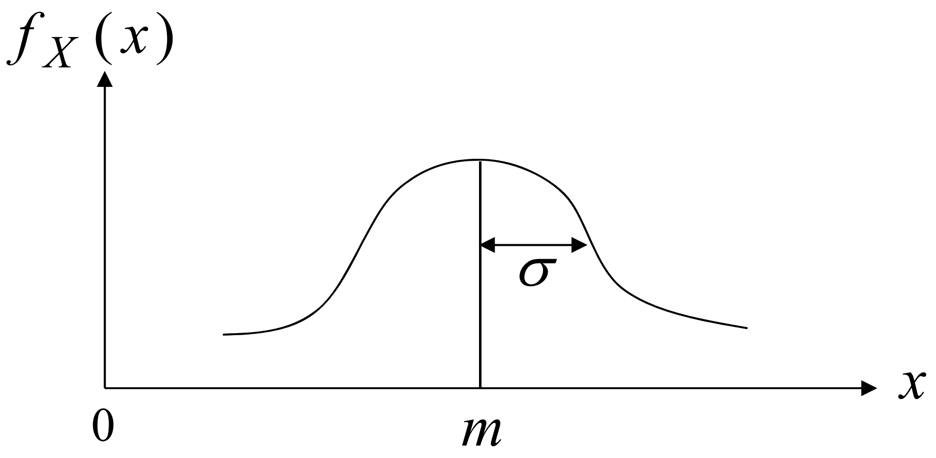
$$E[X] = \mu_X = \int_{-\infty}^{\infty} x\, f_X(x)dx$$

$E[\ ]$: expectation operator

- Variance (⇔ power for zero-mean signals):

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx = E[X^2] - \mu_X^2$$

Normal (Gaussian) Distribution



$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

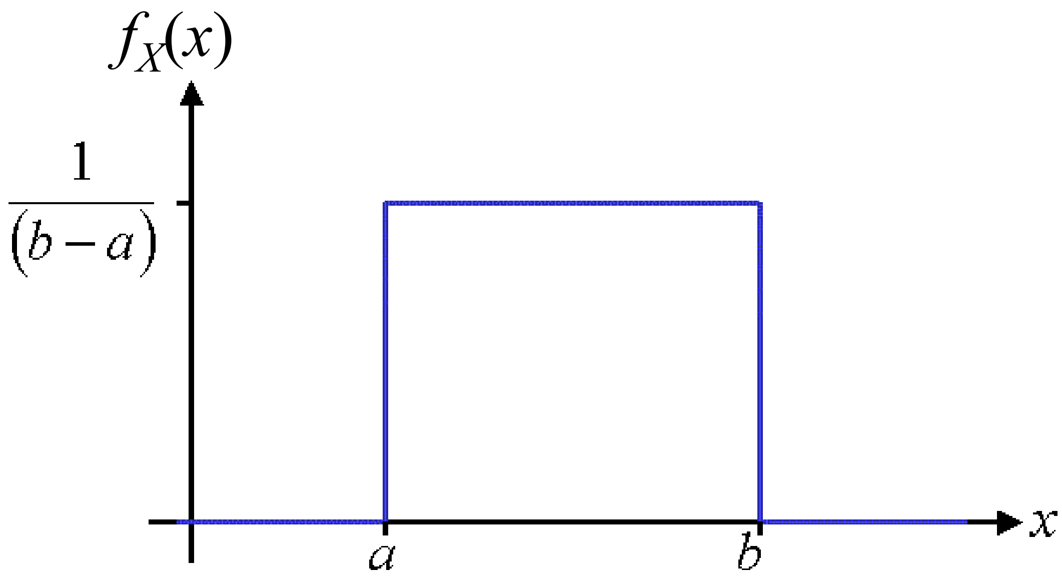
$$F_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}}\, dy$$

$E[X] = m$

$\sigma_X{}^2 = \sigma^2$

σ : rms value

Uniform Distribution



$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$

$$E[X] = \frac{a+b}{2}$$

$$\sigma_X{}^2 = \frac{(b-a)^2}{12}$$

Joint Distribution

- Joint distribution function for two random variables X and Y

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

- Joint probability density function

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

- Properties

1) $F_{XY}(\infty,\infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(u,v)\,du\,dv = 1$

2) $f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x,y)\,dy$

3) $f_Y(x) = \int_{x=-\infty}^{\infty} f_{XY}(x,y)\,dx$

4) X, Y are **independent** $\Leftrightarrow f_{XY}(x,y) = f_X(x) f_Y(y)$

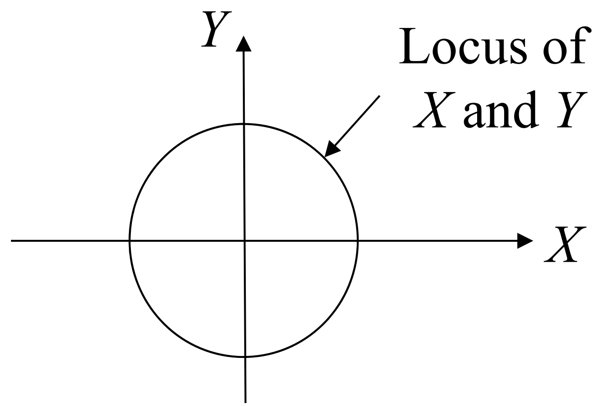
5) X, Y are **uncorrelated** $\Leftrightarrow E[XY] = E[X]E[Y]$

Independent vs. Uncorrelated

- Independent implies Uncorrelated (see problem sheet)
- Uncorrelated does not imply Independence
- For normal RVs (jointly Gaussian), Uncorrelated implies Independent (this is the only exceptional case!)
- An example of uncorrelated but dependent RV's

Let θ be uniformly distributed in $[0,2\pi]$

$$f_{\theta}(x)=\frac{1}{2\pi} \quad \text{for } 0\le x\le 2\pi$$



Define RV's *X* and *Y* as

$$X=\cos\theta \quad Y=\sin\theta$$

Clearly, *X* and *Y* are not independent.

But *X* and *Y* are uncorrelated:

$$E[XY]=\frac{1}{2\pi}\int_0^{2\pi}\cos\theta\sin\theta d\theta=0!$$

Joint Distribution of *n* RVs

- Joint cdf

$$F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n) \equiv P(X_1 \le x_1, X_2 \le x_2, \ldots X_n \le x_n)$$

- Joint pdf

$$f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n) \equiv \frac{\partial^n F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}$$

- **Independent**

$$F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$$

$$f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$$

- **i.i.d.** (independent, identically distributed)
  - The random variables are independent and have the same distribution.
  - Example: outcomes from repeatedly flipping a coin.

Central Limit Theorem

- For i.i.d. random variables,
  $$z = x_1 + x_2 + \cdots + x_n$$
  tends to Gaussian as n goes to infinity.
- Extremely useful in communications.
- That's why noise is usually Gaussian. We often say "**Gaussian noise**" or "Gaussian channel" in communications.

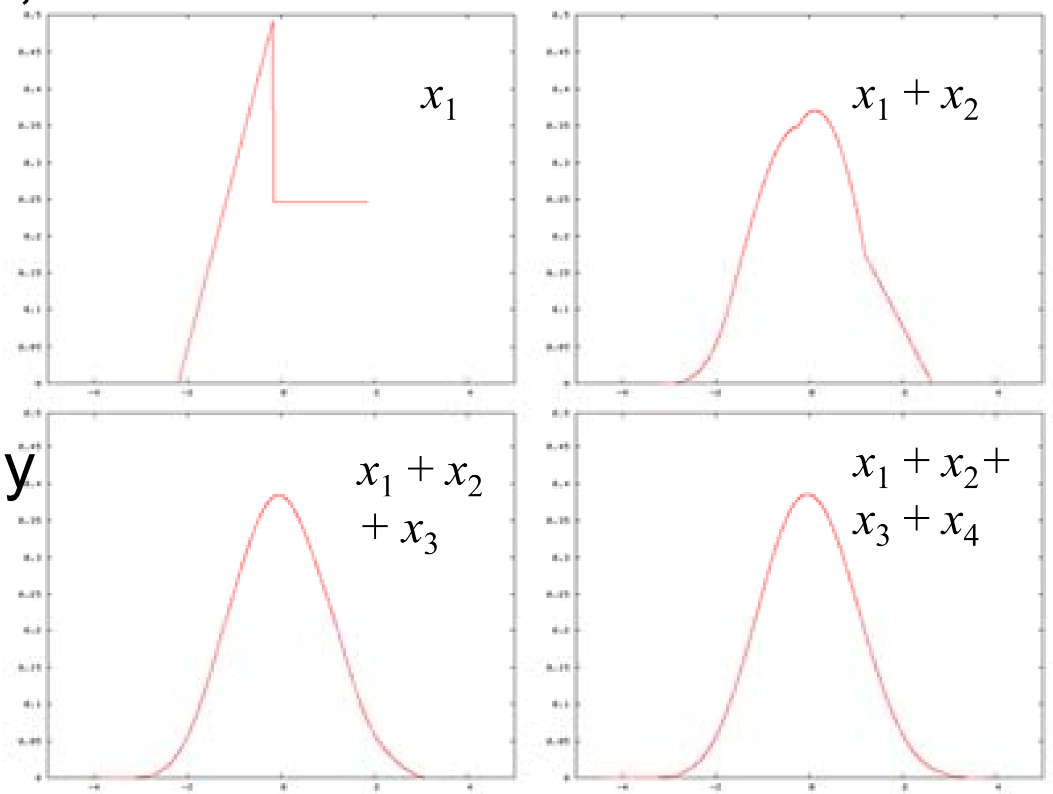


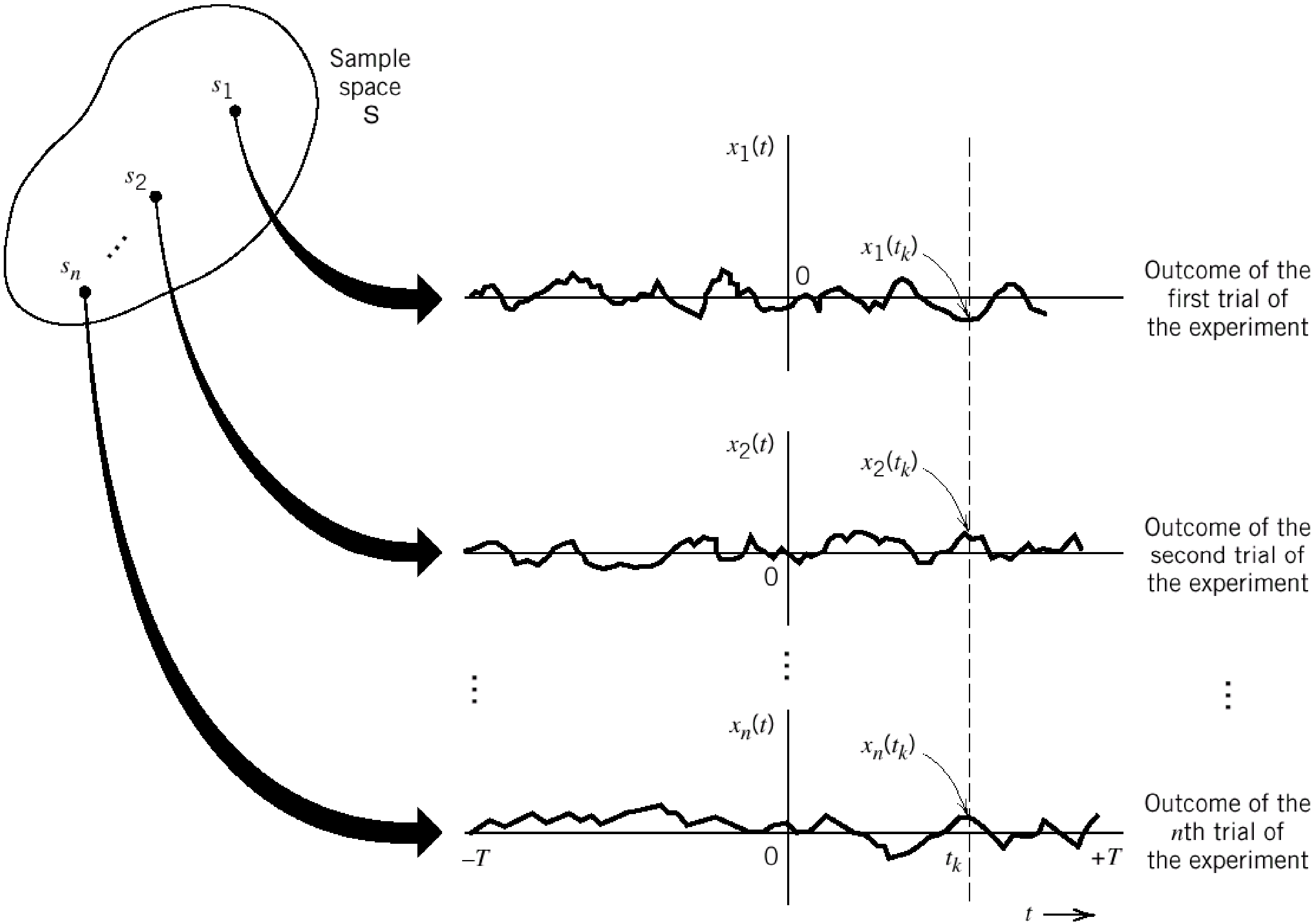
Illustration of convergence to Gaussian distribution

What is a Random Process?

- A random process is a time-varying function that assigns the outcome of a random experiment to each time instant: $X(t)$.
- For a fixed (sample path): a random process is a time varying function, e.g., a signal.
- For fixed t: a random process is a random variable.
- If one scans all possible outcomes of the underlying random experiment, we shall get an **ensemble** of signals.

- Noise can often be modelled as a **Gaussian random process**.

An Ensemble of Signals



Statistics of a Random Process

- For fixed *t*: the random process becomes a random variable, with mean

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x\, f_X(x;t)dx$$

  - In general, the mean is a function of t.

- Autocorrelation function

$$R_X(t_1,t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x,y;t_1,t_2)dxdy$$

  - In general, the autocorrelation function is a two-variable function.
  - It measures the correlation between two samples.

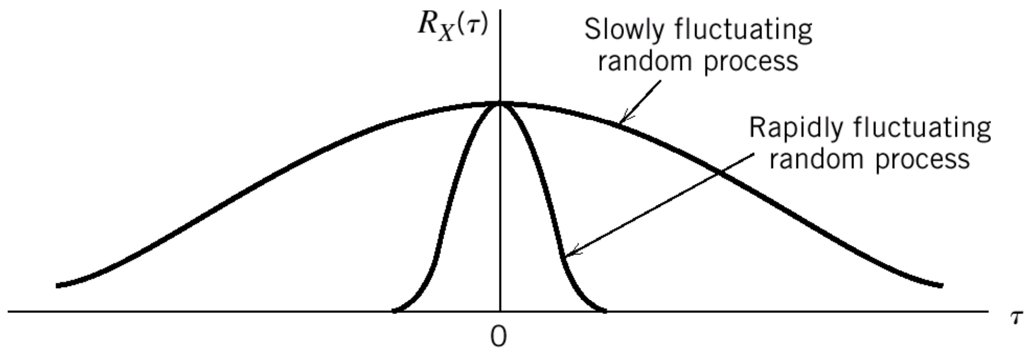
Stationary Random Processes

- A random process is (wide-sense) stationary if
  - Its mean does not depend on *t*

$$\mu_X(t) = \mu_X$$

  - Its autocorrelation function only depends on time difference

$$R_X(t, t+\tau) = R_X(\tau)$$



- In communications, noise and message signals can often be modelled as stationary random processes.

Example

- Show that sinusoidal wave with random phase

$$X(t) = A\cos(\omega_c t + \Theta)$$

with phase Θ uniformly distributed on [0,2π] is stationary.

  – Mean is a constant:

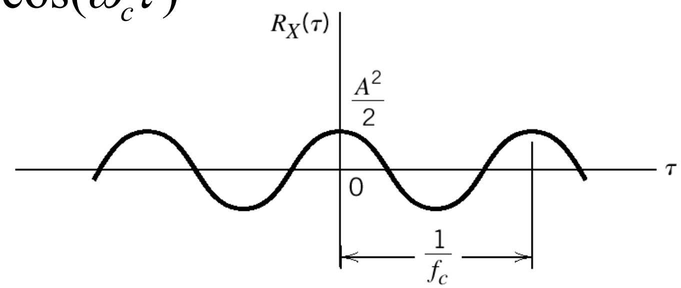
$$f_\Theta(\theta) = \frac{1}{2\pi}, \quad \theta \in [0, 2\pi]$$

$$\mu_X(t) = E[X(t)] = \int_0^{2\pi} A\cos(\omega_c t + \theta)\frac{1}{2\pi}d\theta = 0$$

  – Autocorrelation function only depends on the time difference:

$$\begin{aligned} R_X(t, t+\tau) &= E[X(t)X(t+\tau)] \\ &= E[A^2\cos(\omega_c t + \Theta)\cos(\omega_c t + \omega_c\tau + \Theta)] \\ &= \frac{A^2}{2}E[\cos(2\omega_c t + \omega_c\tau + 2\Theta)] + \frac{A^2}{2}E[\cos(\omega_c\tau)] \\ &= \frac{A^2}{2}\int_0^{2\pi}\cos(2\omega_c t + \omega_c\tau + 2\theta)\frac{1}{2\pi}d\theta + \frac{A^2}{2}\cos(\omega_c\tau) \end{aligned}$$

$$R_X(\tau) = \frac{A^2}{2}\cos(\omega_c\tau)$$



Power Spectral Density

- Power spectral density (PSD) is a function that measures the distribution of power of a random process with frequency.
- PSD is only defined for stationary processes.
- **Wiener-Khinchine relation:** The PSD is equal to the Fourier transform of its autocorrelation function:

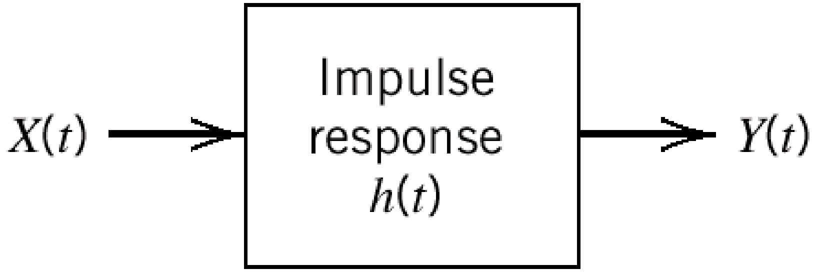
$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau$$

  - A similar relation exists for deterministic signals
- Then the average power can be found as

$$P = E[X^2(t)] = R_X(0) = \int_{-\infty}^{\infty} S_X(f) df$$

- The frequency content of a process depends on how rapidly the amplitude changes as a function of time.
  - This can be measured by the autocorrelation function.

Passing Through a Linear System



- Let Y(t) obtained by passing random process X(t) through a linear system of transfer function H(f). Then the PSD of Y(t)

$$S_Y(f) = |H(f)|^2 S_X(f) \qquad (2.1)$$

  - Proof: see Notes 3.4.2.
  - Cf. the similar relation for deterministic signals
- If X(t) is a Gaussian process, then Y(t) is also a Gaussian process.
  - Gaussian processes are very important in communications.

Imperial College
London

EE2-4: Communication Systems

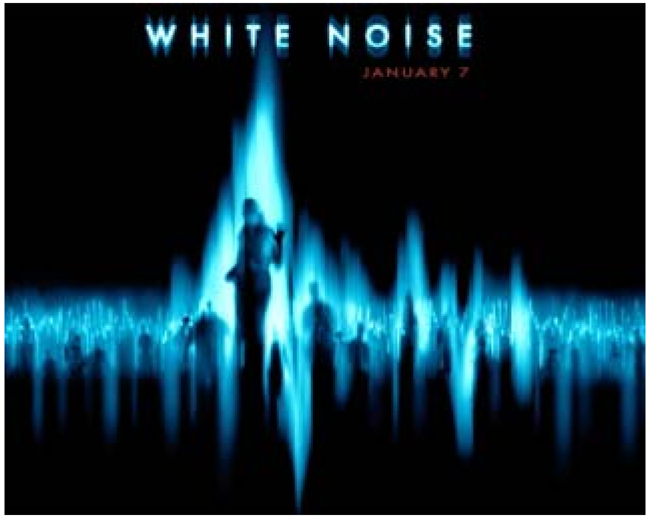
Lecture 3: Noise

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- What is noise?
- White noise and Gaussian noise
- Lowpass noise
- Bandpass noise
  - In-phase/quadrature representation
  - Phasor representation



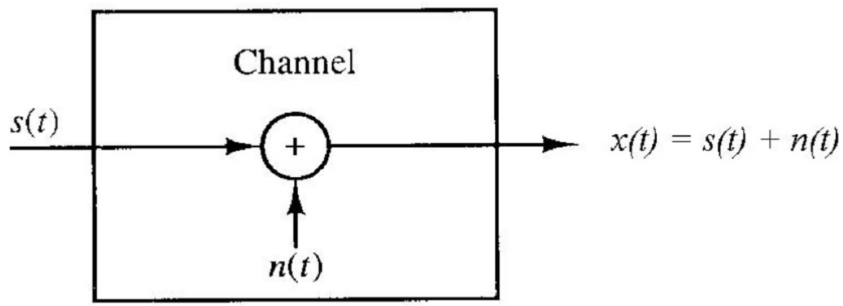
- References
  - Notes of Communication Systems, Chap. 2.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 5
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 11

Noise

- Noise is the unwanted and beyond our control waves that disturb the transmission of signals.
- Where does noise come from?
  - External sources: e.g., atmospheric, galactic noise, interference;
  - Internal sources: generated by communication devices themselves.
    - This type of noise represents a basic limitation on the performance of electronic communication systems.
    - **Shot noise:** the electrons are discrete and are not moving in a continuous steady flow, so the current is randomly fluctuating.
    - **Thermal noise:** caused by the rapid and random motion of electrons within a conductor due to thermal agitation.
- Both are often stationary and have a zero-mean **Gaussian distribution** (following from the central limit theorem).

White Noise

- The additive noise channel
  - $n(t)$ models all types of noise
  - zero mean



- White noise
  - Its power spectrum density (PSD) is constant over all frequencies, i.e.,

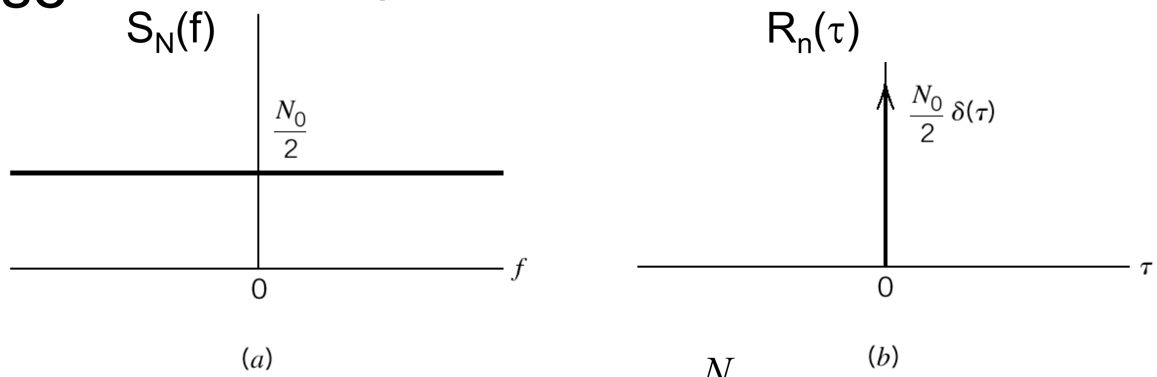
$$S_N(f) = \frac{N_0}{2}, \qquad -\infty < f < \infty$$

  - Factor 1/2 is included to indicate that half the power is associated with positive frequencies and half with negative.
  - The term ***white*** is analogous to white light which contains equal amounts of all frequencies (within the visible band of EM wave).
  - It's only defined for stationary noise.
- An infinite bandwidth is a purely theoretic assumption.

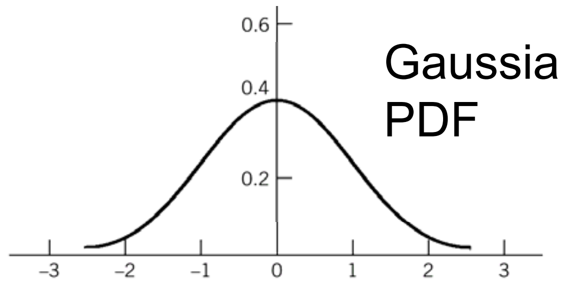
White vs. Gaussian Noise

- White noise

PSD



  - Autocorrelation function of $n(t): R_n(\tau) = \frac{N_0}{2}\delta(\tau)$
  - Samples at different time instants are uncorrelated.
- Gaussian noise: the distribution at any time instant is Gaussian
  - Gaussian noise can be colored
- White noise $\neq$ Gaussian noise
  - White noise can be non-Gaussian



Gaussian PDF

- Nonetheless, in communications, it is typically additive white Gaussian noise (AWGN).

Ideal Low-Pass White Noise

- Suppose white noise is applied to an ideal low-pass filter of bandwidth B such that

$$S_N(f) = \begin{cases} \dfrac{N_0}{2}, & |f| \leq B \\ 0, & \text{otherwise} \end{cases}$$

Power $P_N = N_0 B$

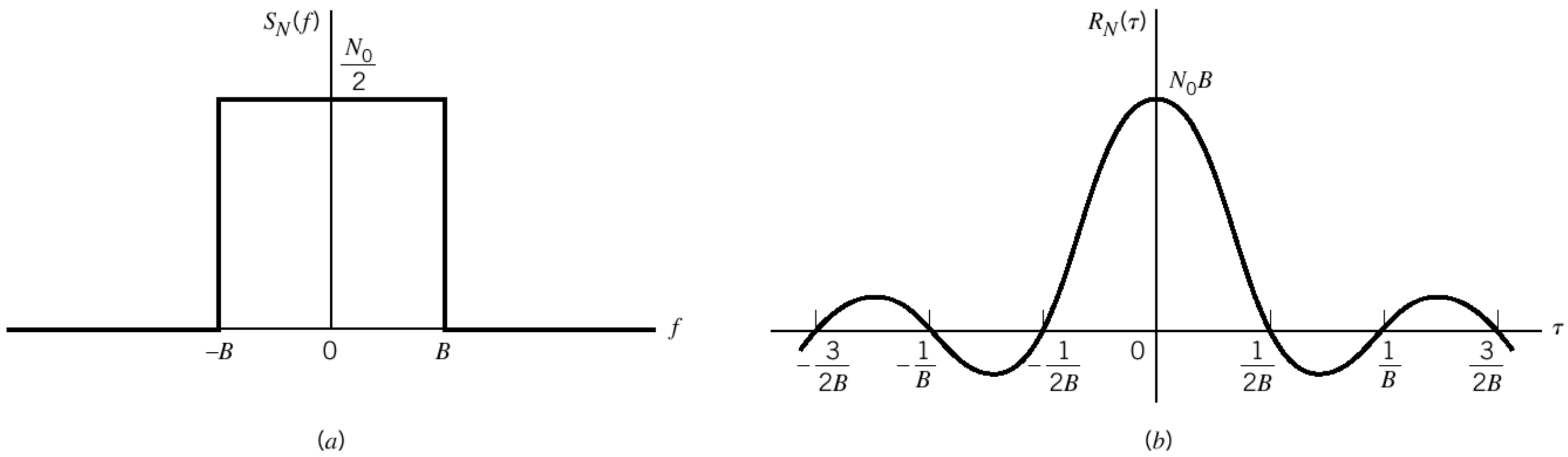
- By Wiener-Khinchine relation, autocorrelation function

$$R_n(\tau) = E[n(t)n(t+\tau)] = N_0 B \operatorname{sinc}(2B\tau) \qquad (3.1)$$

where $\operatorname{sinc}(x) = \sin(\pi x)/\pi x$.

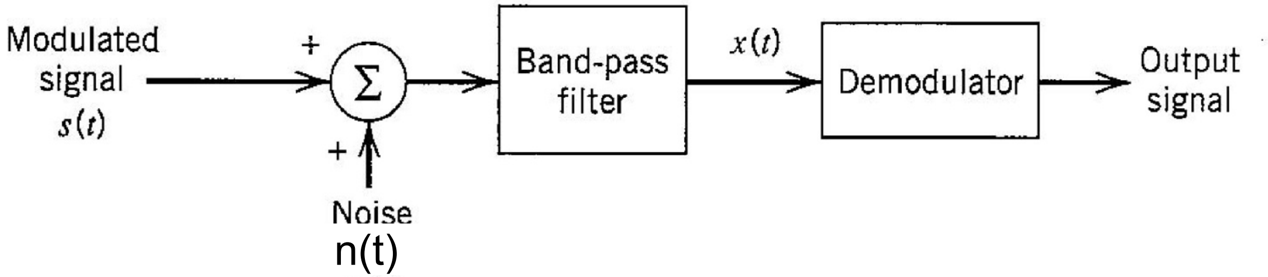
- Samples at Nyquist frequency $2B$ are uncorrelated

$$R_n(\tau) = 0, \qquad \tau = k/(2B),\ k = 1, 2, \ldots$$

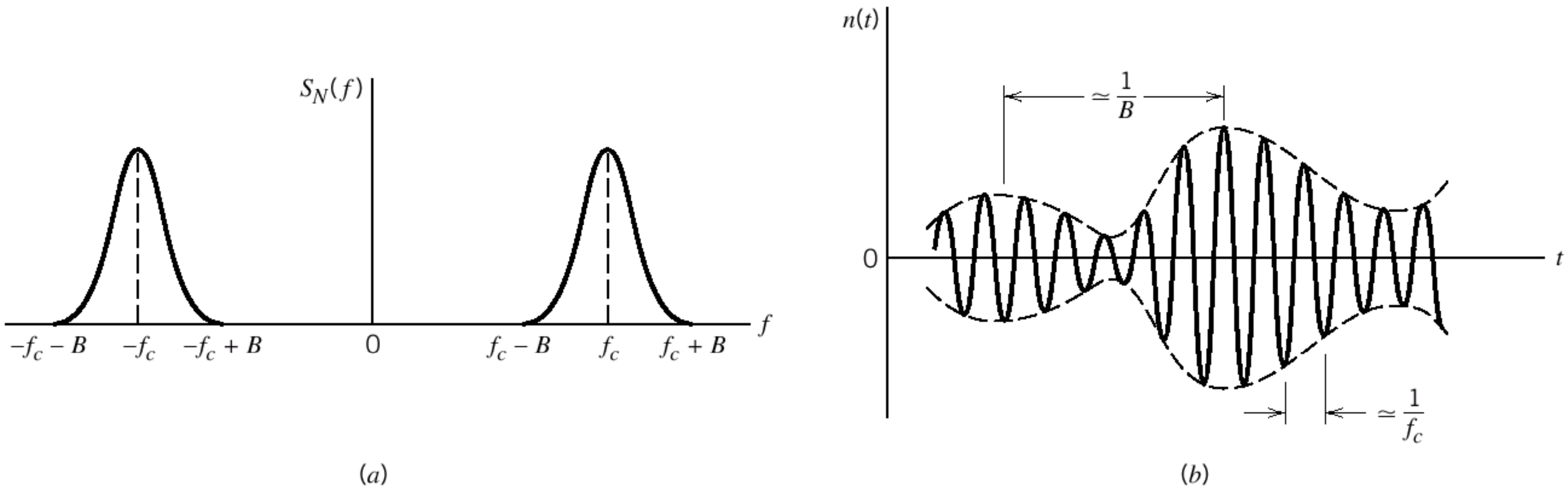


Bandpass Noise

- Any communication system that uses carrier modulation will typically have a bandpass filter of bandwidth B at the front-end of the receiver.



- Any noise that enters the receiver will therefore be bandpass in nature: its spectral magnitude is non-zero only for some band concentrated around the carrier frequency f_c (sometimes called **narrowband noise**).

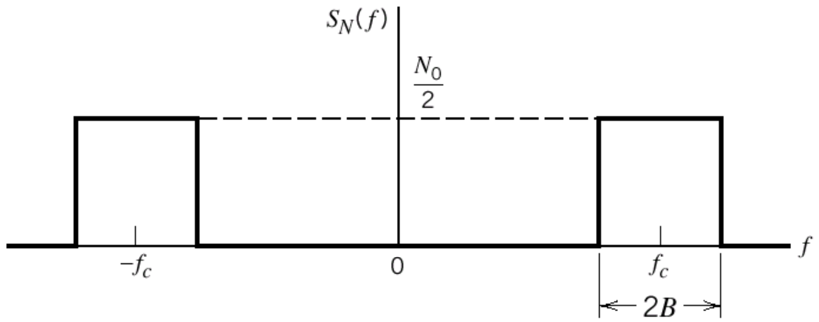


Example

- If white noise with PSD of $N_0/2$ is passed through an ideal bandpass filter, then the PSD of the noise that enters the receiver is given by

$$S_N(f) = \begin{cases} \dfrac{N_0}{2}, & |f - f_C| \le B \\ 0, & \text{otherwise} \end{cases}$$

Power $P_N = 2N_0B$



- Autocorrelation function

$$R_n(\tau) = 2N_0B\,\text{sinc}(2B\tau)\cos(2\pi f_c\tau)$$

  – which follows from (3.1) by applying the frequency-shift property of the Fourier transform

$$g(t) \Leftrightarrow G(\omega)$$
$$g(t)\cdot 2\cos\omega_0 t \Leftrightarrow [G(\omega-\omega_0)+G(\omega+\omega_0)]$$

- Samples taken at frequency $2B$ are still uncorrelated.

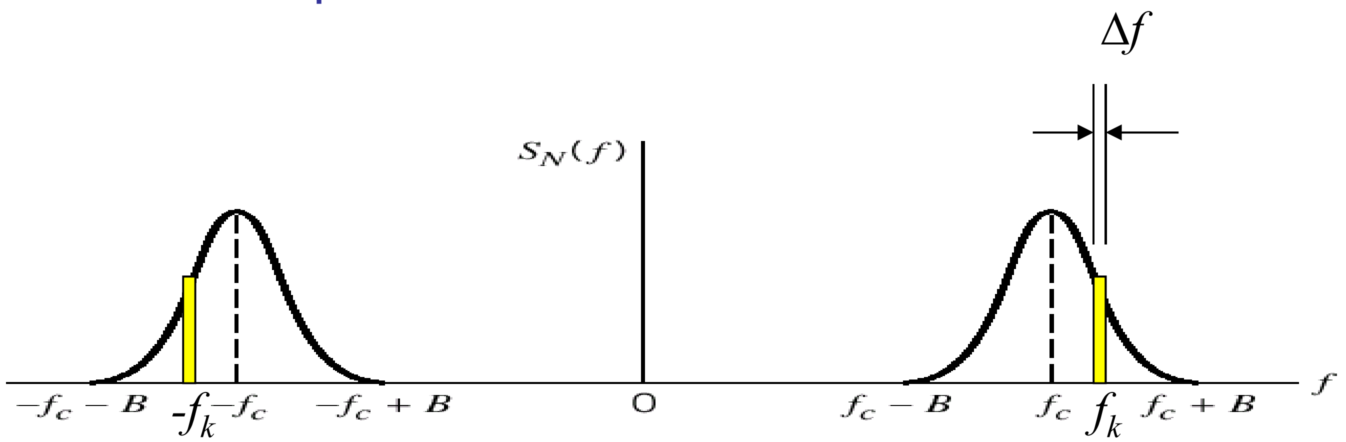
$$R_n(\tau) = 0, \quad \tau = k/(2B),\ k = 1, 2, \ldots$$

Decomposition of Bandpass Noise

- Consider bandpass noise within $|f - f_C| \le B$ with any PSD (i.e., not necessarily white as in the previous example)
- Consider a frequency slice Δf at frequencies f_k and $-f_k$.
- For Δf small:

$$n_k(t) = a_k \cos(2\pi f_k t + \theta_k)$$

  - θ_k: a random phase assumed independent and uniformly distributed in the range $[0, 2\pi)$
  - a_k: a random amplitude.



Representation of Bandpass Noise

- The complete bandpass noise waveform $n(t)$ can be constructed by summing up such sinusoids over the entire band, i.e.,

$$n(t) = \sum_k n_k(t) = \sum_k a_k \cos(2\pi f_k t + \theta_k) \qquad f_k = f_c + k\Delta f \qquad (3.2)$$

- Now, let $f_k = (f_k - f_c) + f_c$, and using $\cos(A + B) = \cos A \cos B - \sin A \sin B$ we obtain the **canonical form of bandpass noise**

$$\boxed{n(t) = n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)}$$

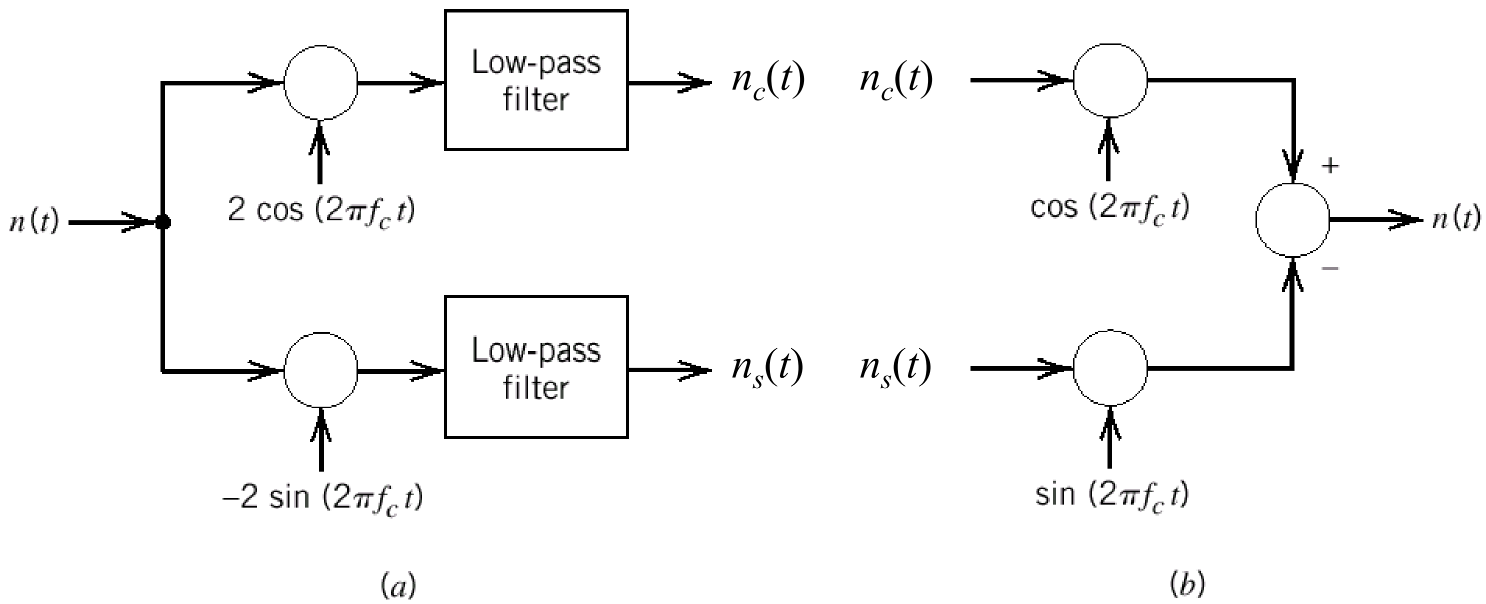
where

$$n_c(t) = \sum_k a_k \cos(2\pi(f_k - f_c)t + \theta_k)$$
$$n_s(t) = \sum_k a_k \sin(2\pi(f_k - f_c)t + \theta_k) \qquad (3.3)$$

  - $n_c(t)$ and $n_s(t)$ are ***baseband*** signals, termed the in-phase and quadrature component, respectively.

Extraction and Generation

- $n_c(t)$ and $n_s(t)$ are fully representative of bandpass noise.
  - (a) Given bandpass noise, one may extract its in-phase and quadrature components (using LPF of bandwith B). This is extremely useful in analysis of noise in communication receivers.
  - (b) Given the two components, one may generate bandpass noise. This is useful in computer simulation.



Properties of Baseband Noise

- If the noise $n(t)$ has zero mean, then $n_c(t)$ and $n_s(t)$ have zero mean.
- If the noise $n(t)$ is Gaussian, then $n_c(t)$ and $n_s(t)$ are Gaussian.
- If the noise $n(t)$ is stationary, then $n_c(t)$ and $n_s(t)$ are stationary.
- If the noise $n(t)$ is Gaussian and its power spectral density $S(f)$ is symmetric with respect to the central frequency f_c, then $n_c(t)$ and $n_s(t)$ are statistical independent.
- The components $n_c(t)$ and $n_s(t)$ have the same variance (= power) as $n(t)$.

Power Spectral Density

- Further, each baseband noise waveform will have the same PSD:

$$S_c(f) = S_s(f) = \begin{cases} S_N(f - f_c) + S_N(f + f_c), & |f| \le B \\ 0, & \text{otherwise} \end{cases} \quad (3.4)$$

- This is analogous to

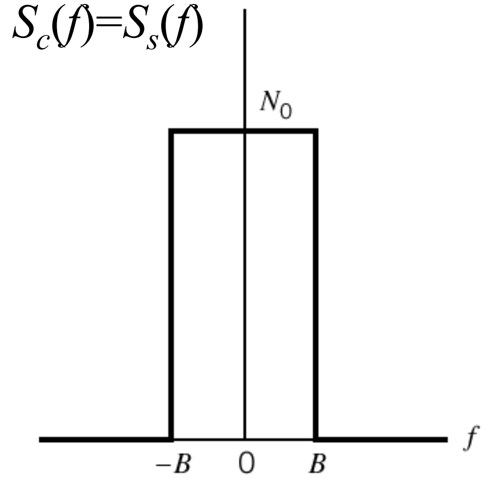
$$g(t) \Leftrightarrow G(\omega)$$
$$g(t) 2\cos\omega_0 t \Leftrightarrow [G(\omega - \omega_0) + G(\omega + \omega_0)]$$

  – A rigorous proof can be found in A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill.
  – The PSD can also be seen from the expressions (3.2) and (3.3) where each of $n_c(t)$ and $n_s(t)$ consists of a sum of closely spaced base-band sinusoids.

Noise Power

- For **ideally filtered narrowband noise**, the PSD of $n_c(t)$ and $n_s(t)$ is therefore given by

$$S_c(f) = S_s(f) = \begin{cases} N_0, & |f| \le B \\ 0, & \text{otherwise} \end{cases} \qquad (3.5)$$



- Corollary: The average power in *each* of the baseband waveforms $n_c(t)$ and $n_s(t)$ is ***identical*** to the average power in the bandpass noise waveform $n(t)$.
- For ideally filtered narrowband noise, the variance of $n_c(t)$ and $n_s(t)$ is $2N_0B$ each.

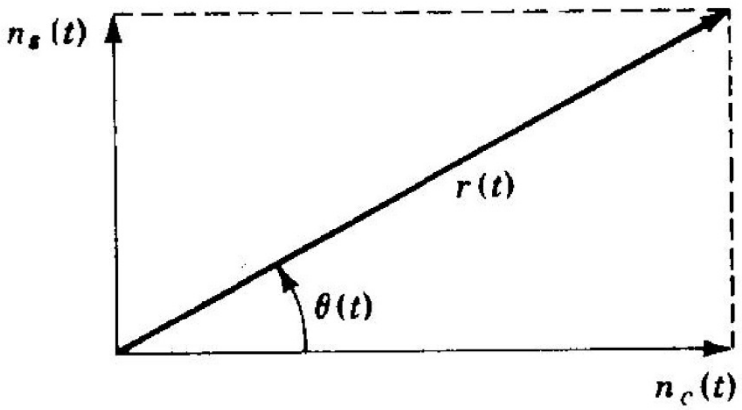
$$P_{Nc} = P_{Ns} = 2N_0B$$

Phasor Representation

- We may write bandpass noise in the alternative form:

$$n(t) = n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$
$$= r(t)\cos[2\pi f_c t + \phi(t)]$$

  - $r(t) = \sqrt{n_c(t)^2 + n_s(t)^2}$: the envelop of the noise
  - $\phi(t) = \tan^{-1}\left(\dfrac{n_s(t)}{n_c(t)}\right)$: the phase of the noise



$$\theta(t) \equiv 2\pi f_c t + \phi(t)$$

Distribution of Envelop and Phase

- It can be shown that if $n_c(t)$ and $n_s(t)$ are Gaussian-distributed, then the magnitude $r(t)$ has a **Rayleigh** distribution, and the phase $\phi(t)$ is **uniformly** distributed.

- What if a sinusoid $A\cos(2\pi f_c t)$ is mixed with noise?
- Then the magnitude will have a **Rice** distribution.

- The proof is deferred to Lecture 11, where such distributions arise in demodulation of digital signals.

Summary

- White noise: PSD is constant over an infinite bandwidth.
- Gaussian noise: PDF is Gaussian.
- Bandpass noise
  - In-phase and quadrature compoments $n_c(t)$ and $n_s(t)$ are low-pass random processes.
  - $n_c(t)$ and $n_s(t)$ have the same PSD.
  - $n_c(t)$ and $n_s(t)$ have the same variance as the band-pass noise $n(t)$.
  - Such properties will be pivotal to the performance analysis of bandpass communication systems.
- The in-phase/quadrature representation and phasor representation are not only basic to the characterization of bandpass noise itself, but also to the analysis of bandpass communication systems.

Imperial College London

EE2-4: Communication Systems

Lecture 4: Noise Performance of DSB

Dr. Cong Ling

Department of Electrical and Electronic Engineering

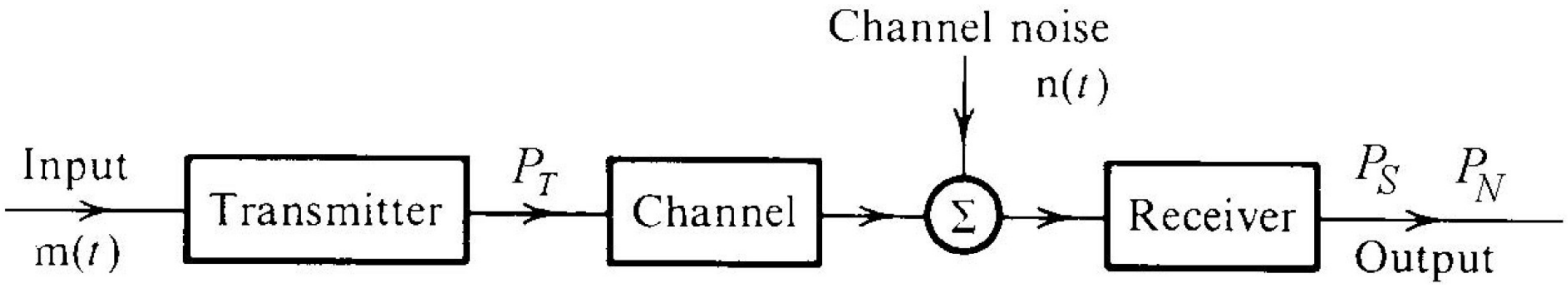
Outline

- SNR of baseband analog transmission
- Revision of AM
- SNR of DSB-SC

- References
  - Notes of Communication Systems, Chap. 3.1-3.3.2.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 6
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 12

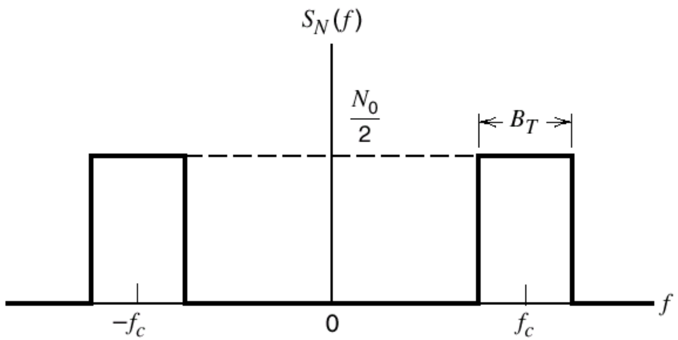
Noise in Analog Communication Systems

- How do various analog modulation schemes perform in the presence of noise?
- Which scheme performs best?
- How can we measure its performance?



Model of an analog communication system

Noise PSD: B_T is the bandwidth, $N_0/2$ is the double-sided noise PSD



SNR

- We must find a way to quantify (= to measure) the performance of a modulation scheme.
- We use the signal-to-noise ratio (SNR) at the output of the receiver:

$$SNR_o \equiv \frac{\text{average power of message signal at the receiver output}}{\text{average power of noise at the receiver output}} = \frac{P_S}{P_N}$$

  - Normally expressed in decibels (dB)
  - SNR (dB) = 10 $\log_{10}$(SNR)
  - This is to manage the wide range of power levels in communication systems
  - In honour of Alexander Bell
  - Example:
    - ratio of 2 → 3 dB; 4 → 6 dB; 10 → 10dB

dB
If x is power,
X (dB) = 10 $\log_{10}(x)$

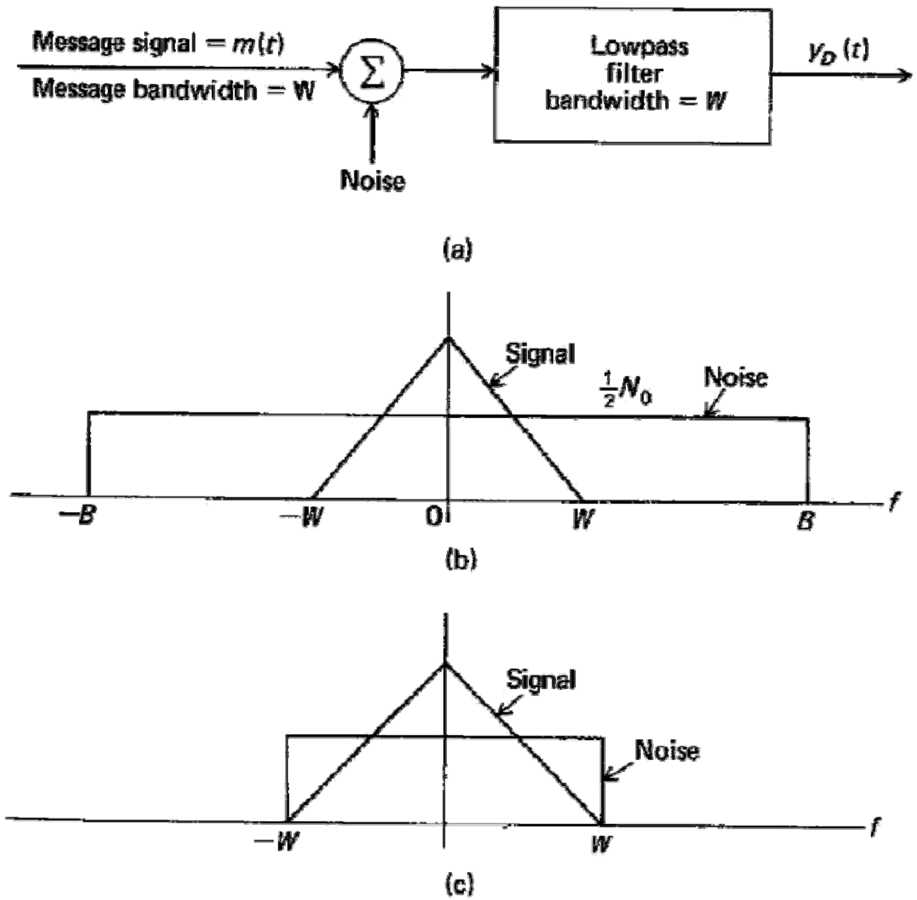
If x is amplitude,
X (dB) = 20 $\log_{10}(x)$

Transmitted Power

- P_T: The transmitted power
- Limited by: equipment capability, battery life, cost, government restrictions, interference with other channels, green communications etc
- The higher it is, the more the received power (P_S), the higher the SNR
- For a fair comparison between different modulation schemes:
  - P_T should be the same for all
- We use the **baseband** signal to noise ratio SNR_{baseband} to calibrate the *SNR* values we obtain

A Baseband Communication System

- It does not use modulation
- It is suitable for transmission over wires
- The power it transmits is identical to the message power: $P_T = P$
- No attenuation: $P_S = P_T = P$
- The results can be extended to band-pass systems



Output SNR

- Average signal (= message) power
  P = the area under the triangular curve
- Assume: Additive, white noise with power spectral density PSD = $N_0/2$
- Average noise power at the receiver
  P_N = area under the straight line = $2W \times N_0/2 = WN_0$
- SNR at the receiver output:

$$SNR_{\text{baseband}} = \frac{P_T}{N_0 W}$$

  – Note: Assume no propagation loss
- Improve the SNR by:
  – increasing the transmitted power ($P_T \uparrow$),
  – restricting the message bandwidth ($W \downarrow$),
  – making the channel/receiver less noisy ($N_0 \downarrow$).

Revision: AM

- General form of an AM signal:

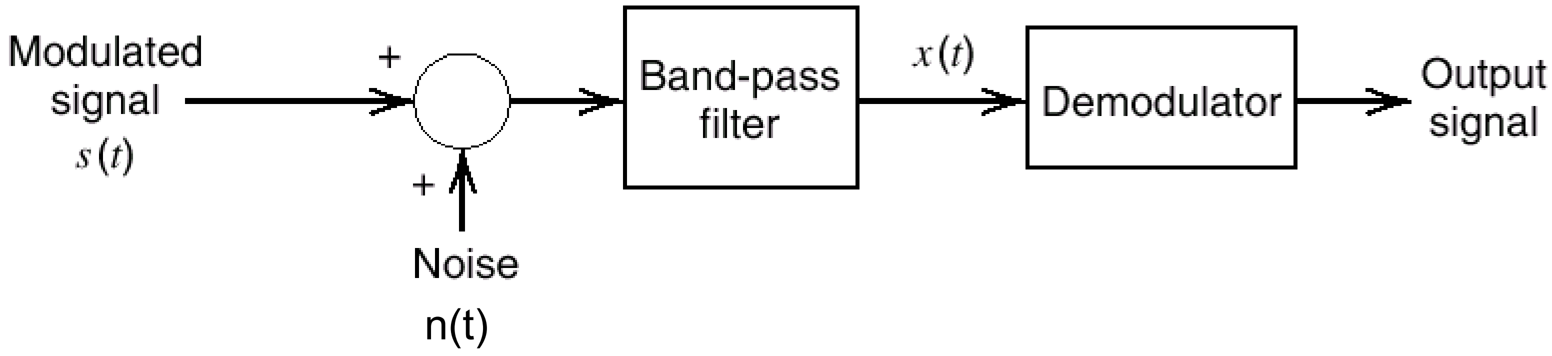
$$s(t)_{AM} = [A + m(t)]\cos(2\pi f_c t)$$

  - A: the amplitude of the carrier
  - f_c: the carrier frequency
  - $m(t)$: the message signal
- Modulation index:

$$\mu = \frac{m_p}{A}$$

  - m_p: the peak amplitude of $m(t)$, i.e., $m_p = \max|m(t)|$

Signal Recovery



Receiver model

1) $\mu \leq 1 \Rightarrow A \geq m_p$: use an envelope detector.
   This is the case in almost all commercial AM radio receivers.
   Simple circuit to make radio receivers cheap.
2) Otherwise: use synchronous detection = product detection = coherent detection

The terms detection and demodulation are used interchangeably.

Synchronous Detection for AM

- Multiply the waveform at the receiver with a local carrier of the same frequency (and phase) as the carrier used at the transmitter:

$$
\begin{aligned}
2\cos(2\pi f_c t)s(t)_{AM} &= [A+m(t)]2\cos^2(2\pi f_c t) \\
&= [A+m(t)][1+\cos(4\pi f_c t)] \\
&= A+m(t)+\cdots
\end{aligned}
$$

- Use a LPF to recover $A+m(t)$ and finally $m(t)$
- Remark: At the receiver you need a signal perfectly synchronized with the transmitted carrier

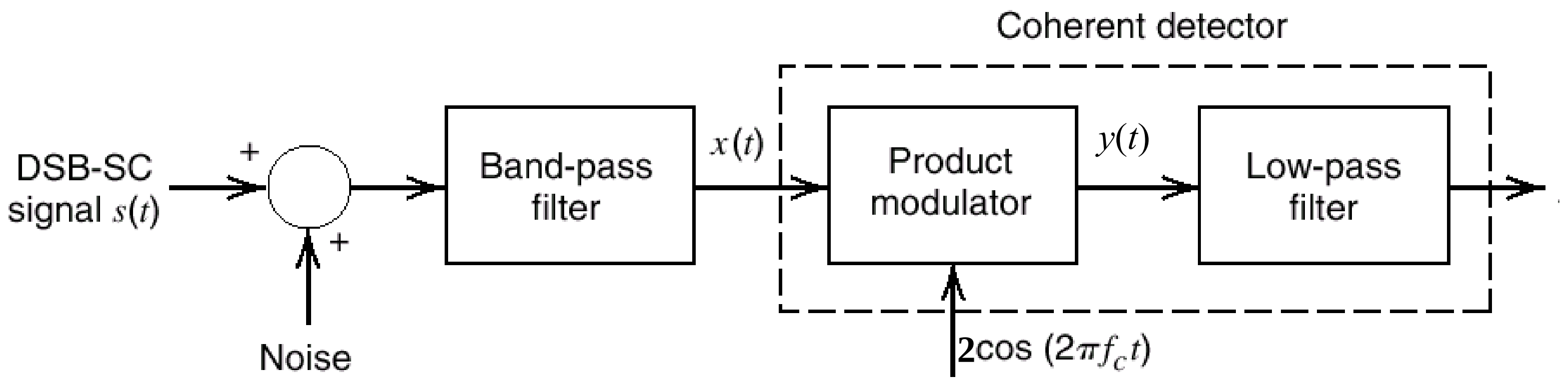
DSB-SC

- Double-sideband suppressed carrier (DSB-SC)

$$s(t)_{DSB-SC} = Am(t)\cos(2\pi f_c t)$$

- Signal recovery: with synchronous detection only
- The received noisy signal is

$$\begin{aligned} x(t) &= s(t) + n(t) \\ &= s(t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \\ &= Am(t)\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \\ &= [Am(t) + n_c(t)]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \end{aligned}$$



Synchronous Detection for DSB-SC

- Multiply with $2\cos(2f_c t)$:

$$y(t) = 2\cos(2\pi f_c t)x(t)$$
$$= Am(t)2\cos^2(2\pi f_c t) + n_c(t)2\cos^2(2\pi f_c t) - n_s(t)\sin(4\pi f_c t)$$
$$= Am(t)[1+\cos(4\pi f_c t)] + n_c(t)[1+\cos(4\pi f_c t)] - n_s(t)\sin(4\pi f_c t)$$

- Use a LPF to keep

$$\tilde{y} = Am(t) + n_c(t)$$

- Signal power at the receiver output:

$$P_S = E\{A^2 m^2(t)\} = A^2 E\{m^2(t)\} = A^2 P$$

- Power of the noise $n_c(t)$ (recall (3.5), and message bandwidth *W*):

$$P_N = \int_{-W}^{W} N_0 df = 2N_0 W$$

Comparison

- *SNR* at the receiver output:

$$SNR_o = \frac{A^2 P}{2 N_0 W}$$

- To which **transmitted** power does this correspond?

$$P_T = E\{A^2 m(t)^2 \cos^2(2\pi f_c t)\} = \frac{A^2 P}{2}$$

- So

$$SNR_o = \frac{P_T}{N_0 W} = SNR_{DSB-SC}$$

- Comparison with

$$SNR_{baseband} = \frac{P_T}{N_0 W} \quad \Rightarrow \quad SNR_{DSB-SC} = SNR_{baseband}$$

- **Conclusion:** DSB-SC system has the same SNR performance as a baseband system.

Imperial College London

EE2-4: Communication Systems

Lecture 5: Noise performance of SSB and AM

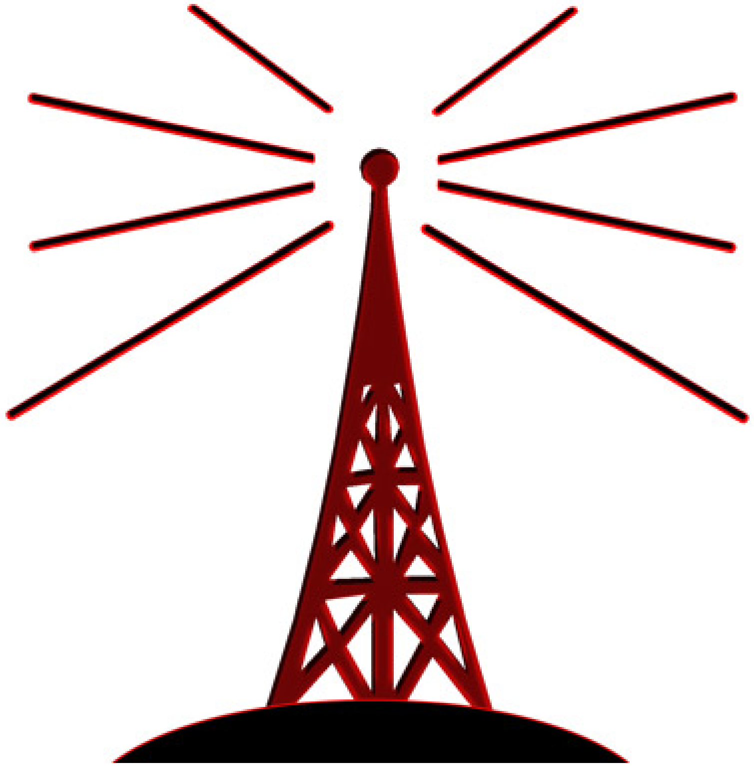
Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Noise in SSB

- Noise in standard AM
  - Coherent detection
    (of theoretic interest only)
  - Envelope detection



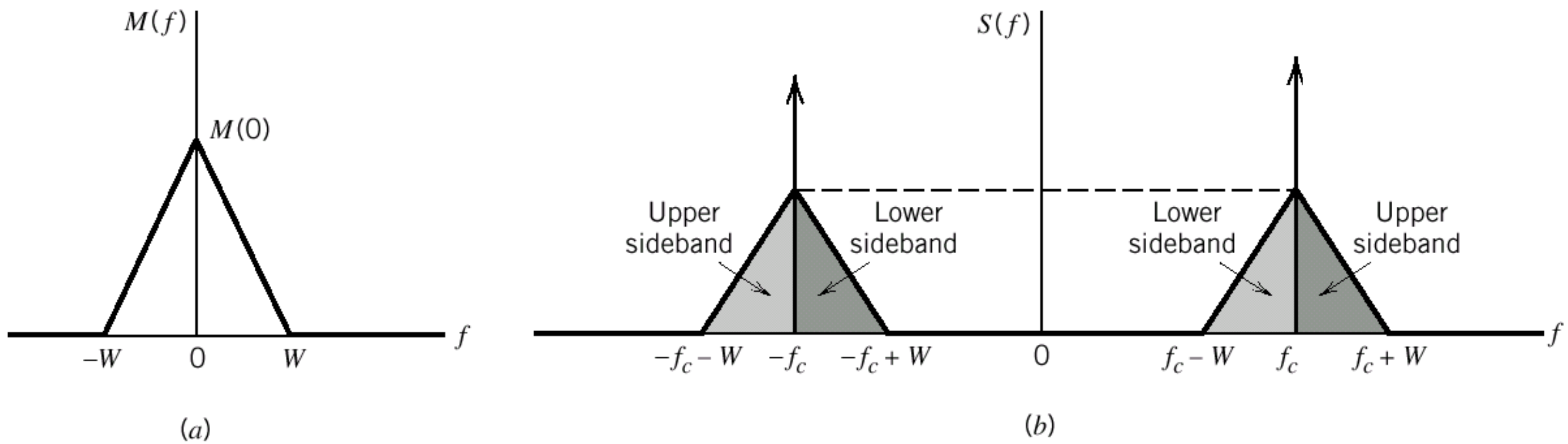
- References
  - Notes of Communication Systems, Chap. 3.3.3-3.3.4.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 6
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 12

SSB Modulation

- Consider single (lower) sideband AM:

$$s(t)_{SSB} = \frac{A}{2} m(t) \cos 2\pi f_c t + \frac{A}{2} \hat{m}(t) \sin 2\pi f_c t$$

  where $\hat{m}(t)$ is the Hilbert transform of $m(t)$.
- $\hat{m}(t)$ is obtained by passing $m(t)$ through a linear filter with transfer function $-j\text{sgn}(f)$.
- $\hat{m}(t)$ and $m(t)$ have the same power P.
- The average power is $A^2P/4$.



Noise in SSB

- Receiver signal $x(t) = s(t) + n(t)$.
- Apply a band-pass filter on the lower sideband.
- Still denote by $n_c(t)$ the lower-sideband noise (different from the double-sideband noise in DSB).
- Using coherent detection:

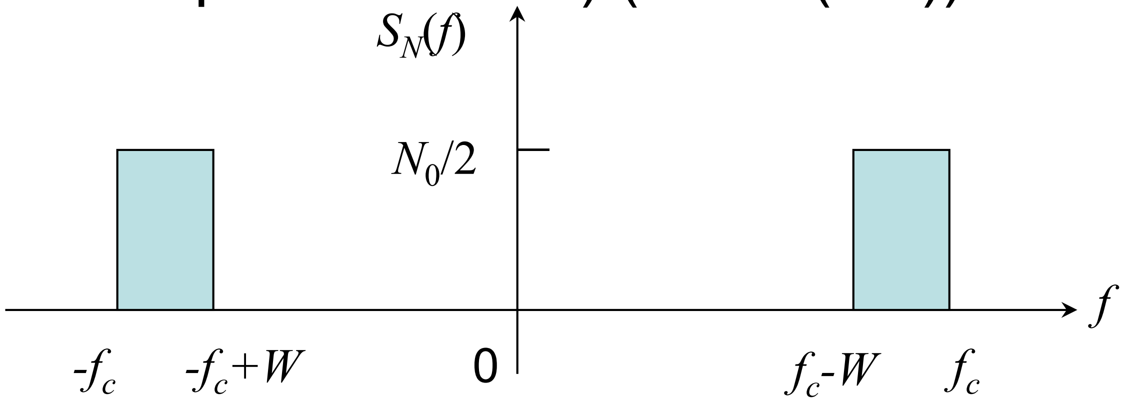
$$
\begin{aligned}
y(t) &= x(t) \times 2\cos(2\pi f_c t) \\
&= \left(\frac{A}{2} m(t) + n_c(t) \right) + \left(\frac{A}{2} m(t) + n_c(t) \right) \cos(4\pi f_c t) \\
&\quad + \left(\frac{A}{2} \hat{m}(t) - n_s(t) \right) \sin(4\pi f_c t)
\end{aligned}
$$

- After low-pass filtering,

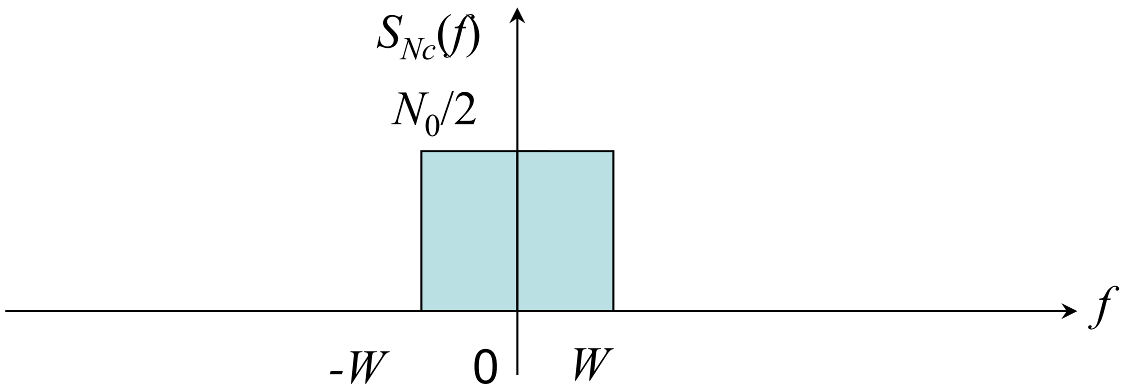
$$
y(t) = \left(\frac{A}{2} m(t) + n_c(t) \right)
$$

Noise Power

- Noise power for $n_c(t)$ = that for band-pass noise = $N_0 W$ (halved compared to DSB) (recall (3.4))



Lower-sideband noise



Baseband noise

Output SNR

- Signal power $A^2P/4$
- SNR at output

$$SNR_{SSB} = \frac{A^2 P}{4N_0 W}$$

- For a baseband system with the *same* transmitted power $A^2P/4$

$$SNR_{baseband} = \frac{A^2 P}{4N_0 W}$$

- **Conclusion:** SSB achieves the same SNR performance as DSB-SC (and the baseband model) but only requires half the band-width.

Standard AM: Synchronous Detection

- Pre-detection signal:

$$x(t) = [A + m(t)]\cos(2\pi f_c t) + n(t)$$
$$= [A + m(t)]\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$
$$= [A + m(t) + n_c(t)]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

- Multiply with $2\cos(2\pi f_c t)$:

$$y(t) = [A + m(t) + n_c(t)][1 + \cos(4\pi f_c t)]$$
$$- n_s(t)\sin(4\pi f_c t)$$

- LPF

$$\tilde{y} = A + m(t) + n_c(t)$$

Output SNR

- Signal power at the receiver output:

$$P_S = E\{m^2(t)\} = P$$

- Noise power:

$$P_N = 2N_0W$$

- SNR at the receiver output:

$$\boxed{SNR_o = \frac{P}{2N_0W} = SNR_{AM}}$$

- Transmitted power

$$P_T = \frac{A^2}{2} + \frac{P}{2} = \frac{A^2 + P}{2}$$

Comparison

- SNR of a baseband signal with the same transmitted power:

$$SNR_{baseband} = \frac{A^2 + P}{2N_0W}$$

- Thus:

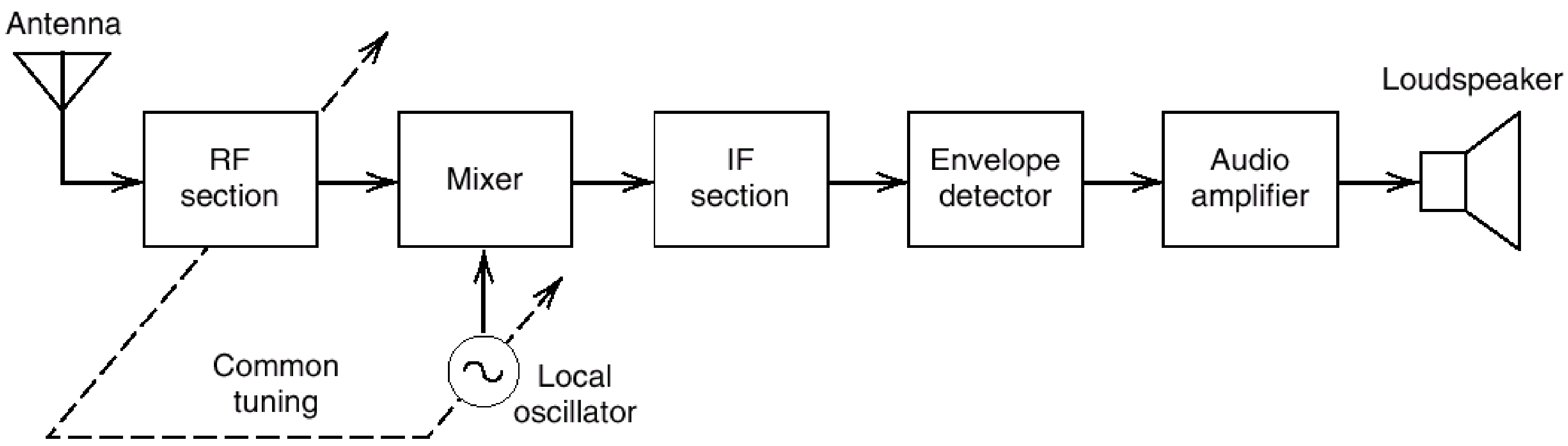
$$SNR_{AM} = \frac{P}{A^2 + P} SNR_{baseband}$$

- Note:

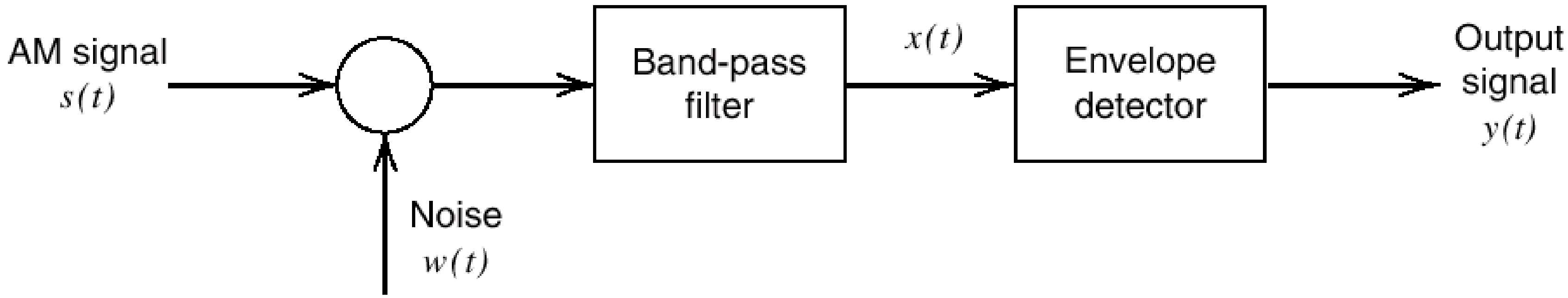
$$\frac{P}{A^2 + P} < 1$$

- **Conclusion:** the performance of standard AM with synchronous recovery is worse than that of a baseband system.

Model of AM Radio Receiver



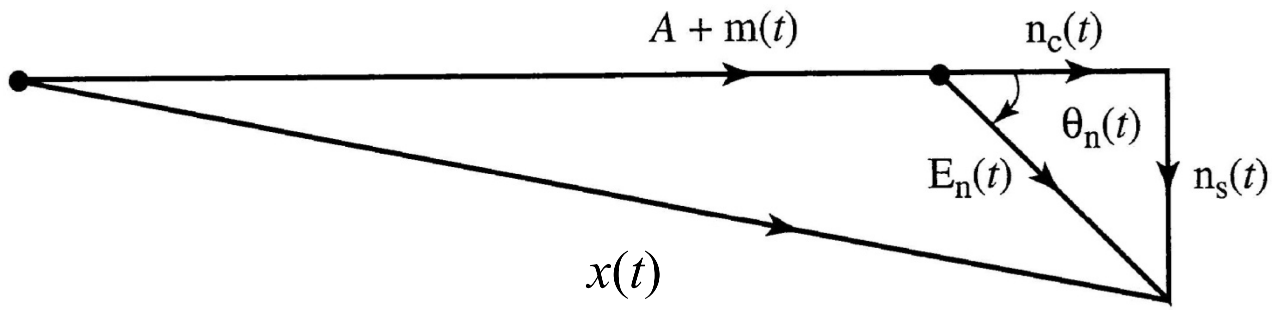
AM radio receiver of the superheterodyne type



Model of AM envelope detector

Envelope Detection for Standard AM

- Phasor diagram of the signals present at an AM receiver



- Envelope

$$y(t) = \text{envelope of } x(t)$$
$$= \sqrt{[A + m(t) + n_c(t)]^2 + n_s(t)^2}$$

- Equation is too complicated
- Must use limiting cases to put it in a form where noise and message are added

Small Noise Case

- **1st Approximation: (a)** **Small Noise Case**

$$n(t) << [A + m(t)]$$

- Then

$$n_s(t) << [A + m(t) + n_c(t)]$$

- Then

$$y(t) \approx [A + m(t) + n_c(t)]$$

Identical to the post-detection signal in the case of synchronous detection!

- Thus

$$SNR_o = \frac{P}{2N_0W} \approx SNR_{env}$$

- And in terms of baseband SNR:

$$SNR_{env} \approx \frac{P}{A^2 + P} SNR_{baseband}$$

- **Valid for small noise only!**

Large Noise Case

- **2nd Approximation: (b)** **Large Noise Case**

$$n(t) >> [A + m(t)]$$

- Isolate the small quantity:

$$y^2(t) = [A + m(t) + n_c(t)]^2 + n_s^2(t)$$

$$= (A + m(t))^2 + n_c^2(t) + 2(A + m(t))n_c(t) + n_s^2(t)$$

$$= [n_c^2(t) + n_s^2(t)]\left\{1 + \frac{(A + m(t))^2}{n_c^2(t) + n_s^2(t)^2} + \frac{2(A + m(t))n_c(t)}{n_c^2(t) + n_s^2(t)}\right\}$$

$$y^2(t) \approx [n_c^2(t) + n_s^2(t)]\left(1 + \frac{2[A + m(t)]n_c(t)}{n_c^2(t) + n_s^2(t)}\right)$$

$$= E_n^2(t)\left(1 + \frac{2[A + m(t)]n_c(t)}{E_n^2(t)}\right) \qquad E_n(t) \equiv \sqrt{n_c^2(t) + n_s^2(t)}$$

Large Noise Case: Threshold Effect

- From the phasor diagram: $n_c(t) = E_n(t)\cos\theta_n(t)$
- Then:

$$y(t) \approx E_n(t)\sqrt{1 + \frac{2[A+m(t)]\cos\theta_n(t)}{E_n(t)}}$$

- Use $\sqrt{1+x} \approx 1 + \frac{x}{2}$ for $x << 1$

$$y(t) \approx E_n(t)\left(1 + \frac{[A+m(t)]\cos\theta_n(t)}{E_n(t)}\right)$$

$$= E_n(t) + [A+m(t)]\cos\theta_n(t)$$

- Noise is multiplicative here!
- No term proportional to the message!
- Result: a **threshold effect**, as below some carrier power level (very low *A*), the performance of the detector deteriorates very rapidly.

Summary

(De-) Modulation Format	Output SNR	Transmitted Power	Baseband Reference SNR	Figure of Merit (= Output SNR / Reference SNR)
AM Coherent Detection	$\frac{P}{2N_0W}$	$\frac{A^2+P}{2}$	$\frac{A^2+P}{2N_0W}$	$\frac{P}{A^2+P}<1$
DSB-SC Coherent Detection	$\frac{A^2P}{2N_0W}$	$\frac{A^2P}{2}$	$\frac{A^2P}{2N_0W}$	1
SSB Coherent Detection	$\frac{A^2P}{4N_0W}$	$\frac{A^2P}{4}$	$\frac{A^2P}{4N_0W}$	1
AM Envelope Detection (Small Noise)	$\frac{P}{2N_0W}$	$\frac{A^2+P}{2}$	$\frac{A^2+P}{2N_0W}$	$\frac{P}{A^2+P}<1$
AM Envelope Detection (Large Noise)	Poor	$\frac{A^2+P}{2}$	$\frac{A^2+P}{2N_0W}$	Poor

A: carrier amplitude, P: power of message signal, N_0: single-sided PSD of noise, W: message bandwidth.

Imperial College London

EE2-4: Communication Systems

Lecture 6: Noise Performance of FM

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Recap of FM
- FM system model in noise
- PSD of noise
- References
  - Notes of Communication Systems, Chap. 3.4.1-3.4.2.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 6
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 12

Frequency Modulation

- Fundamental difference between AM and FM:

- AM: message information contained in the signal **amplitude** ⇒ Additive noise: corrupts directly the modulated signal.

- FM: message information contained in the signal **frequency** ⇒ the effect of noise on an FM signal is determined by the extent to which it changes the frequency of the modulated signal.

- Consequently, FM signals is less affected by noise than AM signals

Revision: FM

- A carrier waveform

$$s(t) = A\cos[\theta_i(t)]$$

  – where $\theta_i(t)$: the **instantaneous phase angle**.

- When

$$s(t) = A\cos(2\pi f\, t) \Rightarrow \theta_i(t) = 2\pi f\, t$$

  we may say that

$$\frac{d\theta}{dt} = 2\pi f \Rightarrow f = \frac{1}{2\pi}\frac{d\theta}{dt}$$

- Generalisation: **instantaneous frequency**:

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt}$$

FM

- In FM: the instantaneous frequency of the carrier varies linearly with the message:

$$f_i(t) = f_c + k_f m(t)$$

  - where k_f is the **frequency sensitivity** of the modulator.
- Hence (assuming $\theta_i(0)=0$):

$$\theta_i(t) = 2\pi \int_0^t f_i(\tau)d\tau = 2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau$$

- Modulated signal:

$$s(t) = A\cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau\right]$$

- Note:
  - (a) The envelope is constant
  - (b) Signal *s*(t) is a non-linear function of the message signal *m*(t).

Bandwidth of FM

- $m_p = \max|m(t)|$: peak message amplitude.
- $f_c - k_f m_p$ < instantaneous frequency < $f_c + k_f m_p$
- Define: **frequency deviation** = the deviation of the instantaneous frequency from the carrier frequency:

$$\Delta f = k_f m_p$$

- Define: **deviation ratio:**

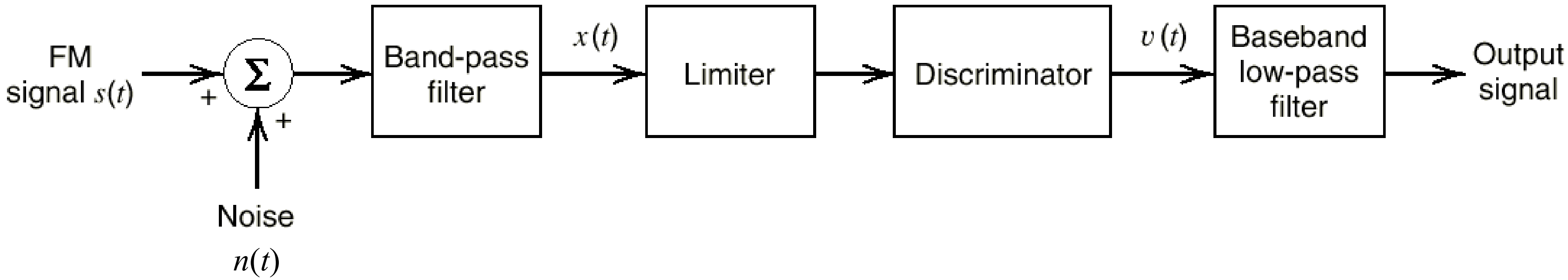
$$\beta = \Delta f / W$$

  - *W*: the message bandwidth.
  - Small β: FM bandwidth ≈ 2x message bandwidth (**narrow-band FM**)
  - Large β: FM bandwidth >> 2x message bandwidth (**wide-band FM**)
- **Carson's rule of thumb**:

$$B_T = 2W(\beta+1) = 2(\Delta f + W)$$

  - $\beta << 1 \Rightarrow B_T \approx 2W$ (as in AM)
  - $\beta >> 1 \Rightarrow B_T \approx 2\Delta f$

FM Receiver



- **Bandpass filter**: removes any signals outside the bandwidth of $f_c \pm B_T/2$ ⇒ the predetection noise at the receiver is bandpass with a bandwidth of B_T.
- FM signal has a constant envelope ⇒ use a **limiter** to remove any amplitude variations
- **Discriminator**: a device with output proportional to the deviation in the instantaneous frequency ⇒ it recovers the message signal
- **Final baseband low-pass filter**: has a bandwidth of W ⇒ it passes the message signal and removes out-of-band noise.

Linear Argument at High SNR

- FM is nonlinear (modulation & demodulation), meaning superposition doesn't hold.

- Nonetheless, it can be shown that for *high SNR*, noise output and message signal are approximately independent of each other:
  Output ≈ Message + Noise (i.e., no other nonlinear terms).

- Any (smooth) nonlinear systems are locally linear!
- This can be justified rigorously by applying Taylor series expansion.
- Noise does not affect power of the message signal at the output, and vice versa.
- ⇒ We can compute the signal power for the case without noise, and accept that the result holds for the case with noise too.
- ⇒ We can compute the noise power for the case without message, and accept that the result holds for the case with message too.

Output Signal Power Without Noise

- Instantaneous frequency of the input signal:

$$f_i = f_c + k_f m(t)$$

- Output of discriminator:

$$k_f m(t)$$

- So, output signal power:

$$P_S = k_f^2 P$$

  - P : the average power of the message signal

Output Signal with Noise

- In the presence of additive noise, the real predetection signal is

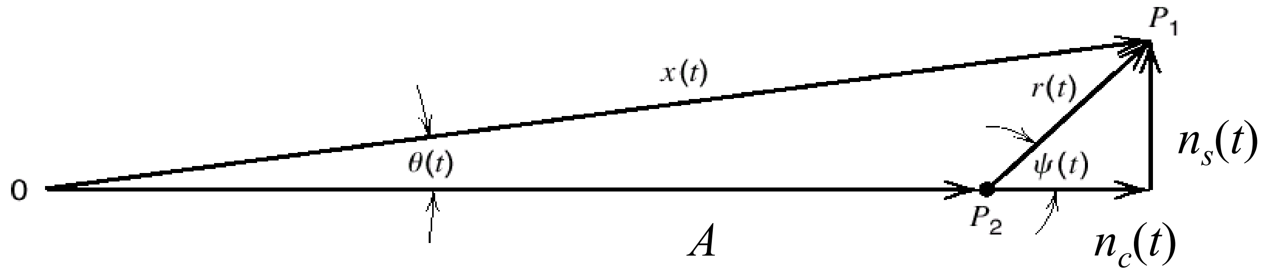
$$x(t) = A\cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau\right]$$
$$+ n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

- It can be shown (by linear argument again): For high SNR, noise output is approximately independent of the message signal
  ⇒ In order to calculate the power of output noise, we may assume there is no message
  ⇒ i.e., we only have the carrier plus noise present:

$$\tilde{x}(t) = A\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

Phase Noise

Phasor diagram of the FM carrier and noise signals



- Instantaneous phase noise:

$$\theta_i(t) = \tan^{-1}\frac{n_s(t)}{A+n_c(t)}$$

- For large carrier power (large *A*):

$$\theta_i(t) = \tan^{-1}\frac{n_s(t)}{A} \approx \frac{n_s(t)}{A}$$

- Discriminator output = instantaneous frequency:

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} = \frac{1}{2\pi A}\frac{dn_s(t)}{dt}$$

Discriminator Output

- The discriminator output in the presence of both signal and noise:

$$k_f m(t) + \frac{1}{2\pi A}\frac{dn_s(t)}{dt}$$

- What is the PSD of

$$n_d(t) = \frac{dn_s(t)}{dt}$$

- Fourier theory:

$$\text{if} \quad x(t) \leftrightarrow X(f)$$

$$\text{then} \quad \frac{dx(t)}{dt} \leftrightarrow j2\pi f X(f)$$

- Differentiation with respect to time = passing the signal through a system with transfer function of $H(f) = j2\pi f$

Noise PSD

- It follows from (2.1) that

$$S_o(f) = |H(f)|^2 \, S_i(f)$$

  - $S_i(f)$: PSD of input signal
  - $S_o(f)$: PSD of output signal
  - $H(f)$: transfer function of the system

- Then:

$$\{\text{PSD of } n_d(t)\} = |\, j2\pi f \,|^2 \times \{\text{PSD of } n_s(t)\}$$

$$\{\text{PSD of } n_s(t)\} = \left\{ N_0 \text{ within band } \pm \frac{B_T}{2} \right\}$$

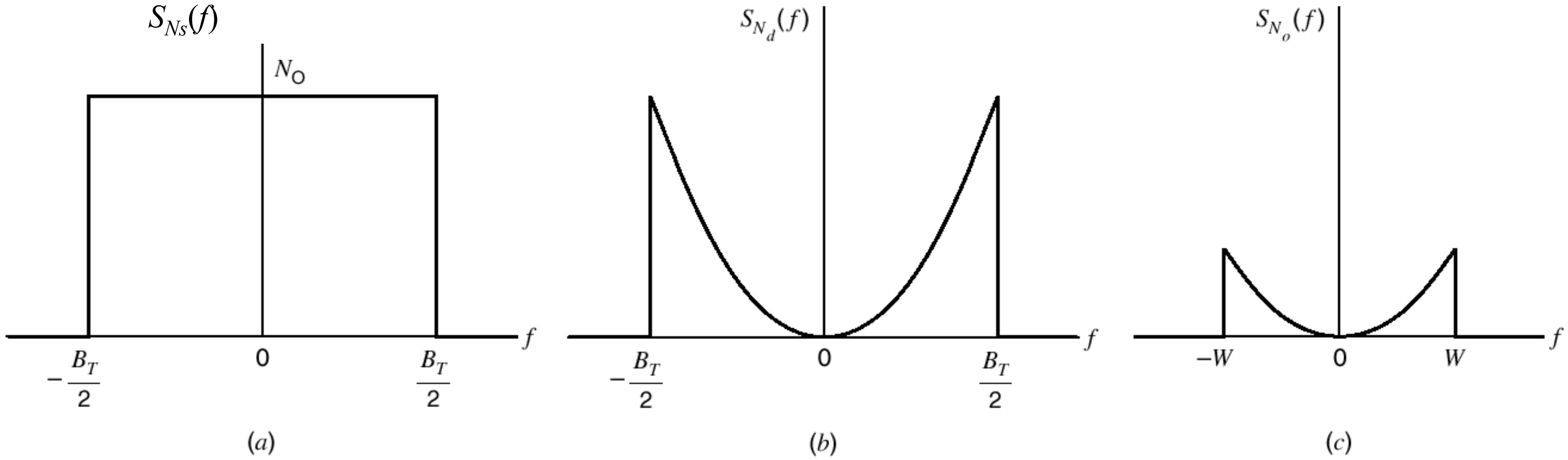
$$\{\text{PSD of } n_d(t)\} = |\, 2\pi f \,|^2 \times N_0 \qquad \left| f \right| \le B_T / 2$$

$$\left\{ \text{PSD of } f_i(t) = \frac{1}{2\pi A} \frac{dn_s(t)}{dt} \right\} = \left(\frac{1}{2\pi A} \right)^2 |\, 2\pi f \,|^2 \times N_0 = \frac{f^2}{A^2} N_0$$

- After the LPF, the PSD of noise output $n_o(t)$ is restricted in the band $\pm W$

$$S_{N_o}(f) = \frac{f^2}{A^2} N_0 \qquad \left| f \right| \le W \qquad (6.1)$$

Power Spectral Densities



(a) Power spectral density of quadrature component $n_s(t)$ of narrowband noise $n(t)$.
(b) Power spectral density of noise $n_d(t)$ at the discriminator output.
(c) Power spectral density of noise $n_o(t)$ at the receiver output.

Imperial College London

EE2-4: Communication Systems

Lecture 7: Pre/de-emphasis for FM and Comparison of Analog Systems

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Derivation of FM output SNR
- Pre/de-emphasis to improve SNR
- Comparison with AM
- References
  - Notes of Communication Systems, Chap. 3.4.2-3.5.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 6
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 12

Noise Power

- Average noise power at the receiver output:

$$P_N = \int_{-W}^{W} S_{N_o}(f)\,df$$

- Thus, from (6.1)

$$P_N = \int_{-W}^{W} \frac{f^2}{A^2} N_0\,df = \frac{2N_0W^3}{3A^2} \qquad (7.1)$$

- Average noise power at the output of a FM receiver

$$\propto \frac{1}{\text{carrier power } A^2}$$

- $A \uparrow \Rightarrow$ Noise$\downarrow$, called the *quieting effect*

Output SNR

- Since $P_S = k_f^2 P$, the output SNR

$$SNR_O = \frac{P_S}{P_N} = \frac{3A^2 k_f^2 P}{2N_0 W^3} = SNR_{FM}$$

- Transmitted power of an FM waveform:

$$P_T = \frac{A^2}{2}$$

- From $SNR_{baseband} = \frac{P_T}{N_0 W}$ and $\beta = \frac{k_f m_p}{W}$:

$$SNR_{FM} = \frac{3k_f^2 P}{W^2} SNR_{baseband} = 3\beta^2 \frac{P}{m_p^2} SNR_{baseband}$$

$$\propto \beta^2 SNR_{baseband} \quad \text{(could be much higher than AM)}$$

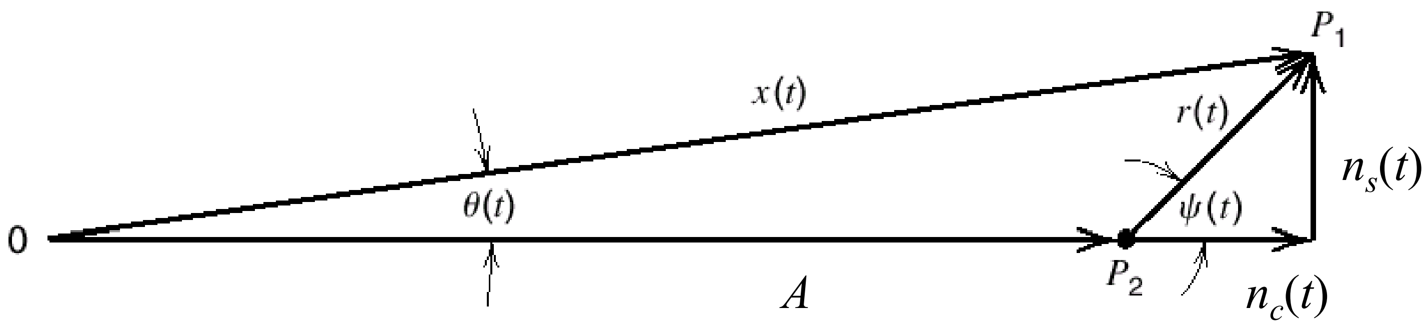
- Valid when the carrier power is large compared with the noise power

Threshold effect

- The FM detector exhibits a more pronounced threshold effect than the AM envelope detector.
- The threshold point occurs around when signal power is 10 times noise power:

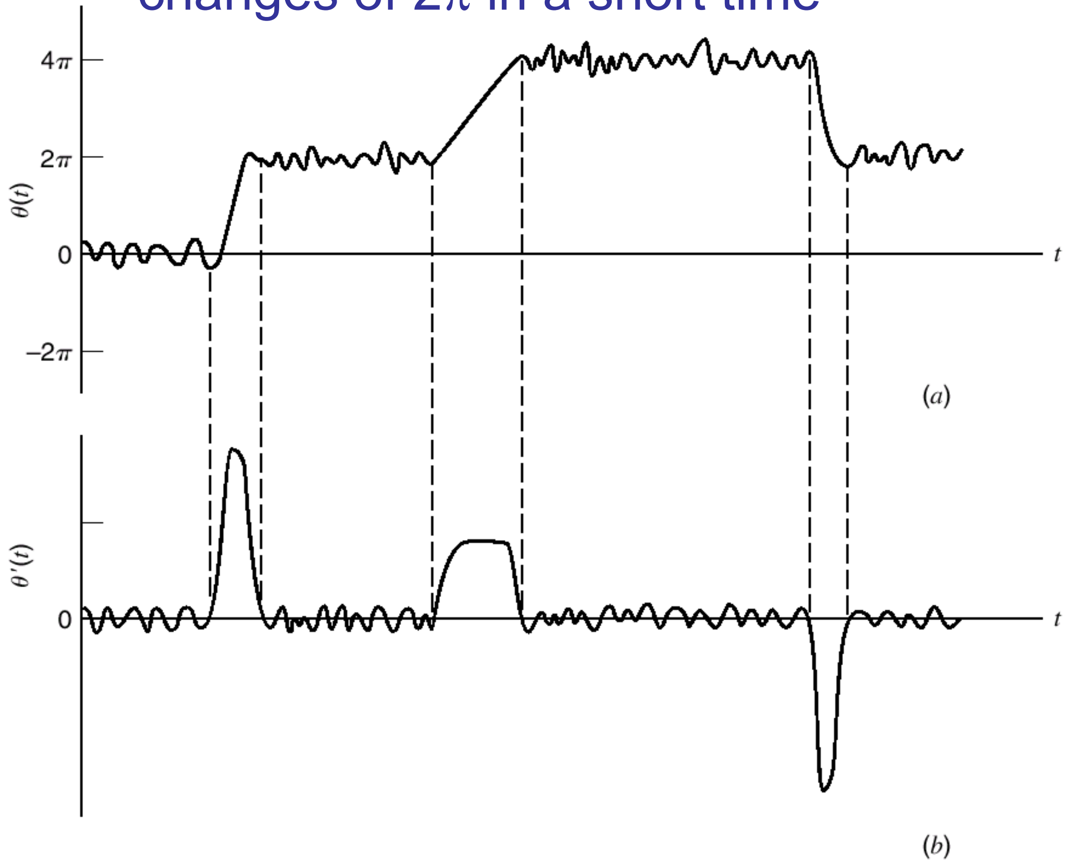
$$\frac{A^2}{2N_0B_T} = 10, \qquad B_T = 2W(\beta + 1)$$

- Below the threshold the FM receiver breaks (i.e., significantly deteriorated).
- Can be analyzed by examining the phasor diagram



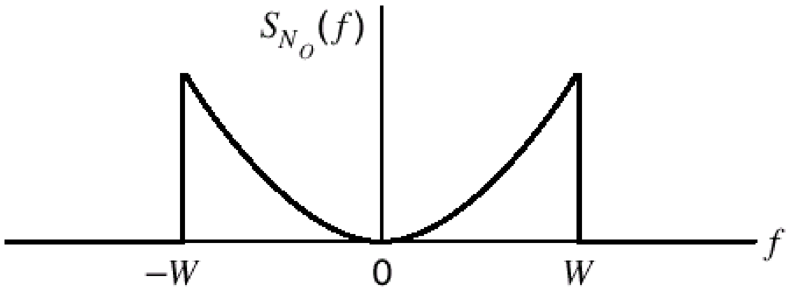
Qualitative Discussion

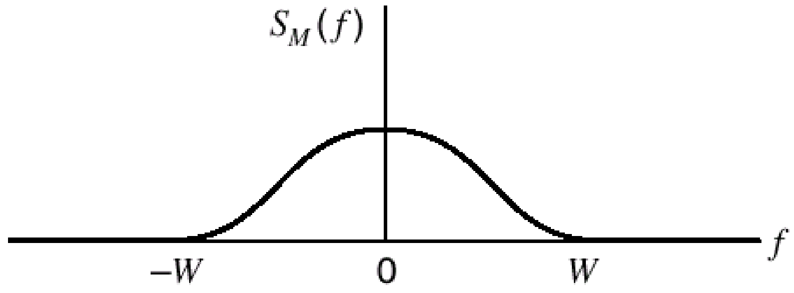
- As the noise changes randomly, the point P_1 wanders around P_2
  - High SNR: change of angle is small
  - Low SNR: P_1 occasionally sweeps around origin, resulting in changes of 2π in a short time



Illustrating impulse like components in $\theta'(t) = d\theta(t)/dt$ produced by changes of 2π in $\theta(t)$; (a) and (b) are graphs of $\theta(t)$ and $\theta'(t)$, respectively.

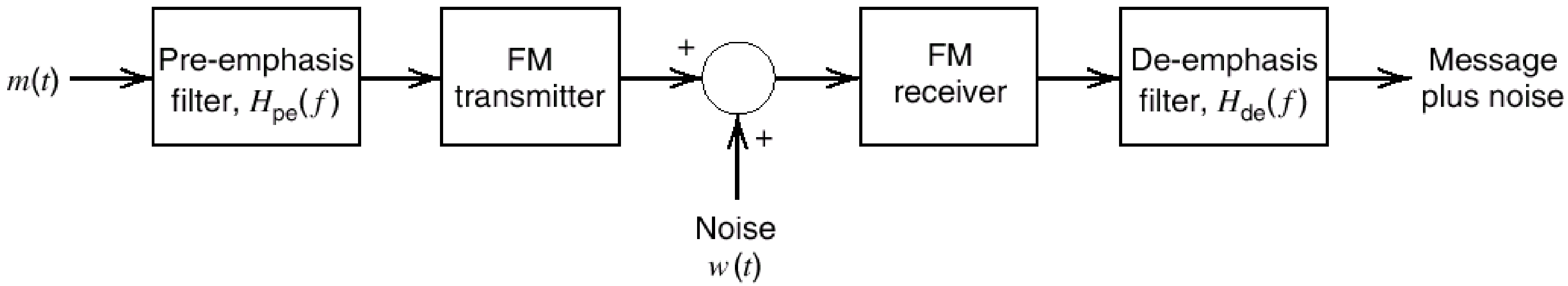
Improve Output SNR





- PSD of the noise at the detector output $\propto$ square of frequency.
- PSD of a typical message typically rolls off at around 6 dB per decade
- To increase SNR_{FM}:
  - Use a LPF to cut-off high frequencies at the output
    - Message is attenuated too, not very satisfactory
  - Use **pre-emphasis** and **de-emphasis**
    - Message is unchanged
    - High frequency components of noise are suppressed

Pre-emphasis and De-emphasis



- $H_{pe}(f)$: used to artificially emphasize the high frequency components of the message prior to modulation, and hence, before noise is introduced.
- $H_{de}(f)$: used to de-emphasize the high frequency components at the receiver, and restore the original PSD of the message signal.
- In theory, $H_{pe}(f) \propto f$, $H_{de}(f) \propto 1/f$.
- This can improve the output SNR by around 13 dB.
- **Dolby noise reduction** uses an analogous pre-emphasis technique to reduce the effects of noise (hissing noise in audiotape recording is also concentrated on high frequency).

Improvement Factor

- Assume an ideal pair of pre/de-emphasis filters

$$H_{de}(f) = 1/H_{pe}(f), \quad |f| \le W$$

- PSF of noise at the output of de-emphasis filter

$$\frac{f^2}{A^2} N_0 \; |H_{de}(f)|^2, \qquad |f| \le B_T/2, \qquad \left(\text{recall } S_{N_o}(f) = \frac{f^2}{A^2} N_0 \right)$$

- Average power of noise with de-emphasis

$$P_N = \int_{-W}^{W} \frac{f^2}{A^2} |H_{de}(f)|^2 N_0 df$$

- Improvement factor (using (7.1))

$$I = \frac{P_N \text{ without pre/de-emphasis}}{P_N \text{ with pre/de-emphasis}} = \frac{\frac{2N_0 W^3}{3A^2}}{\int_{-W}^{W} \frac{f^2}{A^2} |H_{de}(f)|^2 N_0 df} = \frac{2W^3}{3\int_{-W}^{W} f^2 |H_{de}(f)|^2 df}$$

Example Circuits

- (a) Pre-emphasis filter

$$H_{pe}(f) \cong 1 + jf / f_0$$

$$f_0 = 1/(2\pi rC), \qquad R \ll r, \quad 2\pi frC \ll 1$$

- (b) De-emphasis filter

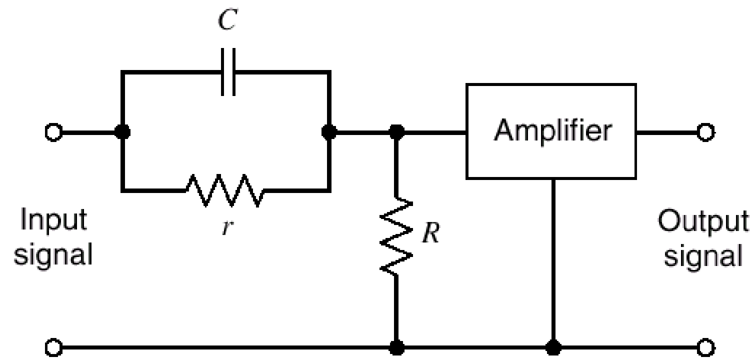
$$H_{de}(f) = \frac{1}{1 + jf / f_0}$$

- Improvement

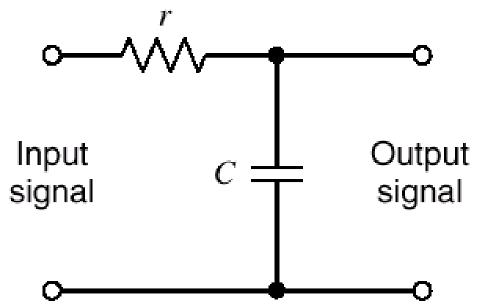
$$I = \frac{2W^3}{3\int_{-W}^{W} f^2 / (1 + f^2 / f_0^2) df}$$

$$= \frac{(W / f_0)^3}{3[(W / f_0) - \tan^{-1}(W / f_0)]}$$

- In commercial FM, $W = 15 \text{ kHz}, f_0 = 2.1 \text{ kHz}$
$\Rightarrow I = 22 \Rightarrow 13 \text{ dB}$ (a significant gain)



(a)



(b)

Comparison of Analogue Systems

- Assumptions:
  - single-tone modulation, i.e.: $m(t) = A_m \cos(2\pi f_m t)$;
  - the message bandwidth $W = f_m$;
  - for the AM system, $\mu = 1$;
  - for the FM system, $\beta = 5$ (which is what is used in commercial FM transmission, with $\Delta f = 75$ kHz, and $W = 15$ kHz).
- With these assumptions, we find that the SNR expressions for the various modulation schemes become:

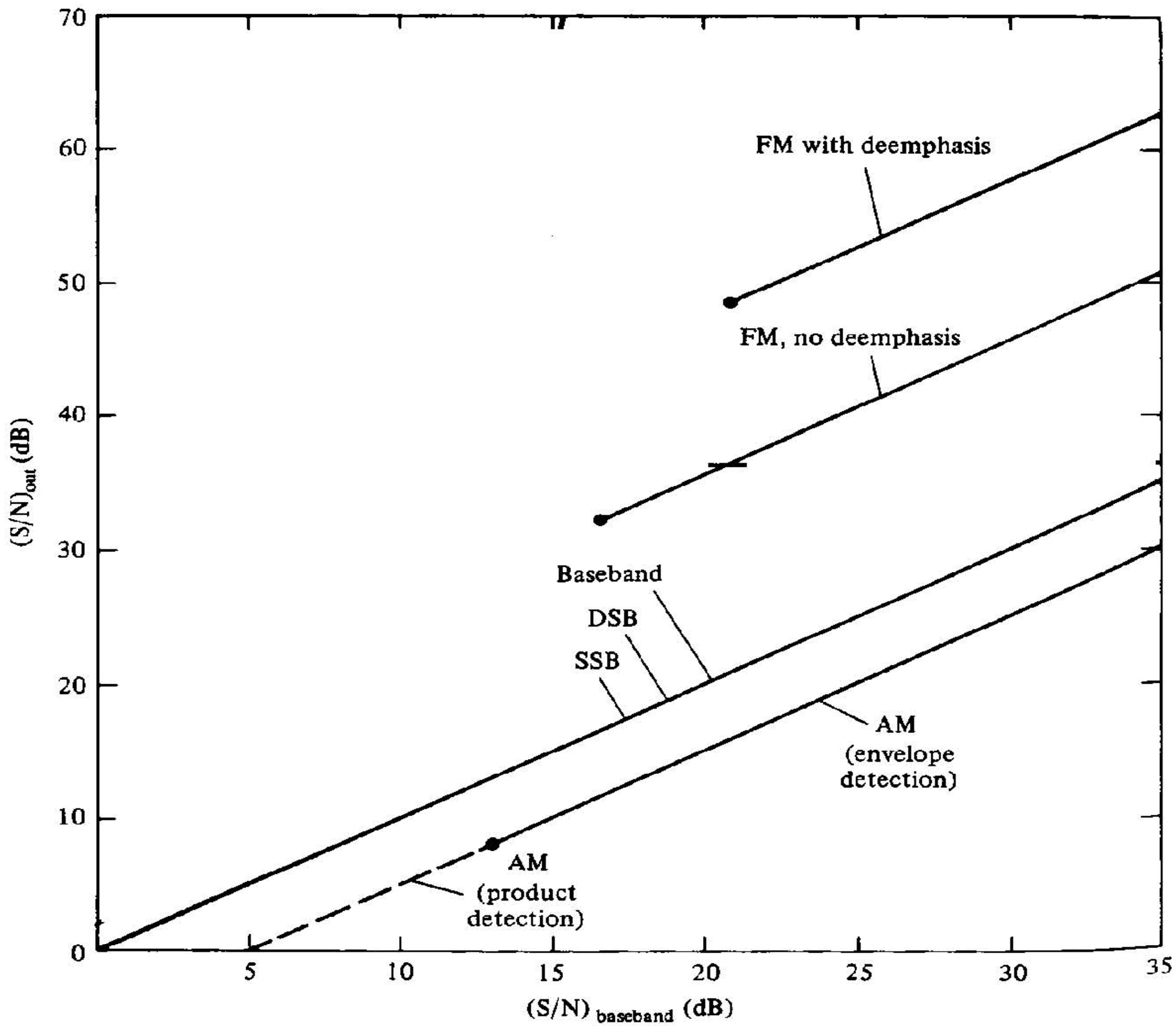
$$SNR_{DSB-SC} = SNR_{baseband} = SNR_{SSB}$$

$$SNR_{AM} = \frac{1}{3} SNR_{baseband}$$

$$SNR_{FM} = \frac{3}{2}\beta^2 SNR_{baseband} = \frac{75}{2} SNR_{baseband}$$

without pre/de-emphasis

Performance of Analog Systems



Conclusions

- (Full) AM: The SNR performance is 4.8 dB worse than a baseband system, and the transmission bandwidth is $B_T = 2W$.

- DSB: The SNR performance is identical to a baseband system, and the transmission bandwidth is $B_T = 2W$.

- SSB: The SNR performance is again identical, but the transmission bandwidth is only $B_T = W$.

- FM: The SNR performance is 15.7 dB better than a baseband system, and the transmission bandwidth is $B_T = 2(\beta + 1)W = 12W$ (with pre- and de-emphasis the SNR performance is increased by about 13 dB with the same transmission bandwidth).

Imperial College London

EE2-4: Communication Systems

Lecture 8: Digital Representation of Signals

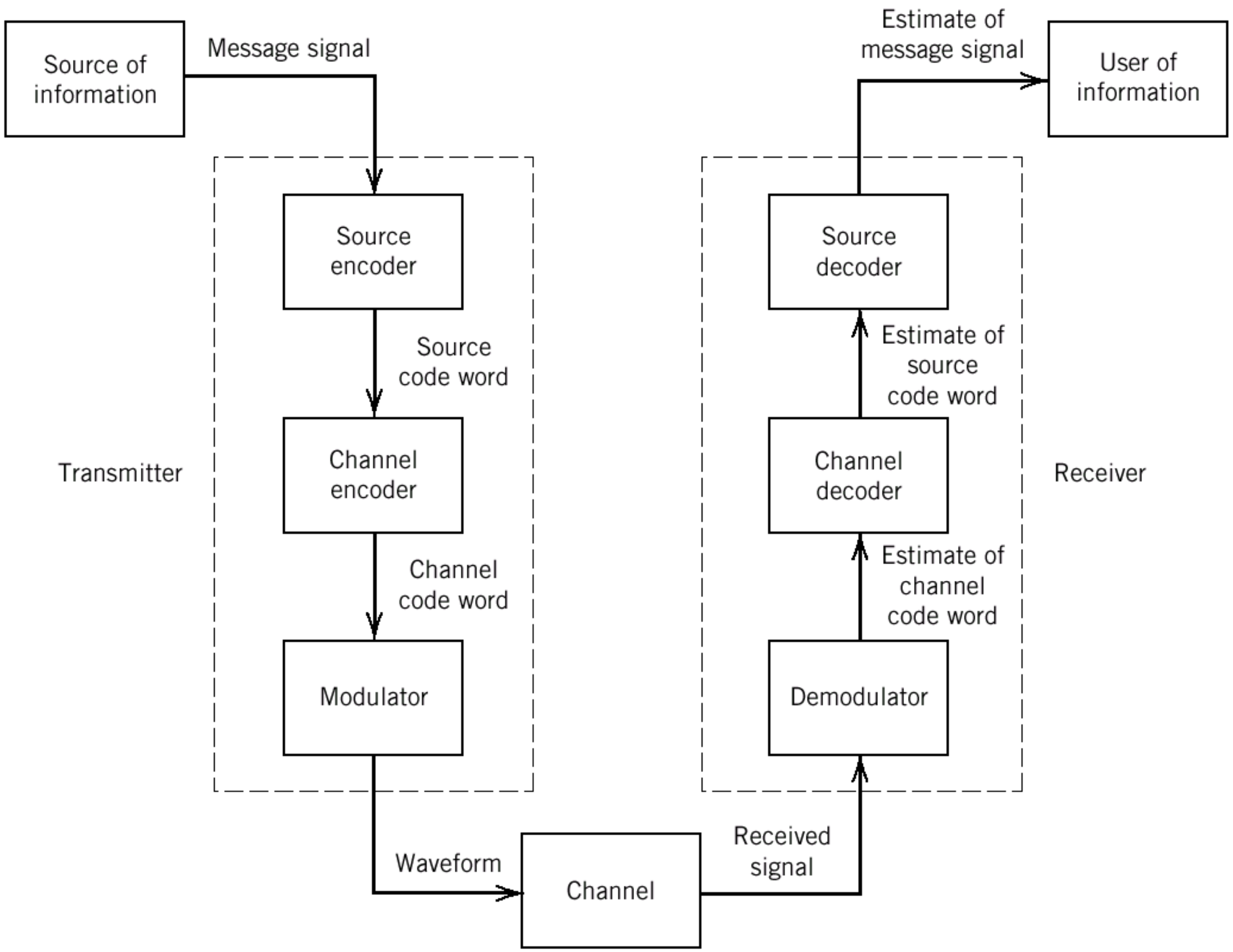
Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Introduction to digital communication
- Quantization (A/D) and noise
- PCM
- Companding
- References
  - Notes of Communication Systems, Chap. 4.1-4.3
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 7
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 6

Block Diagram of Digital Communication

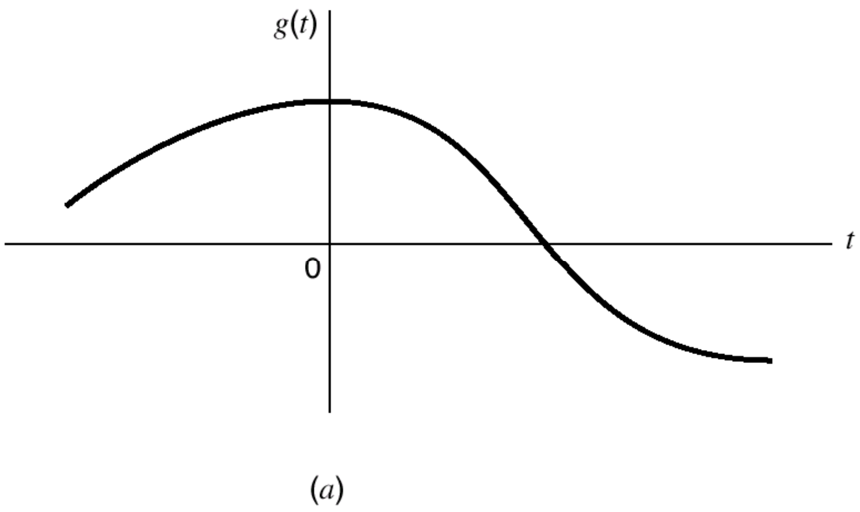


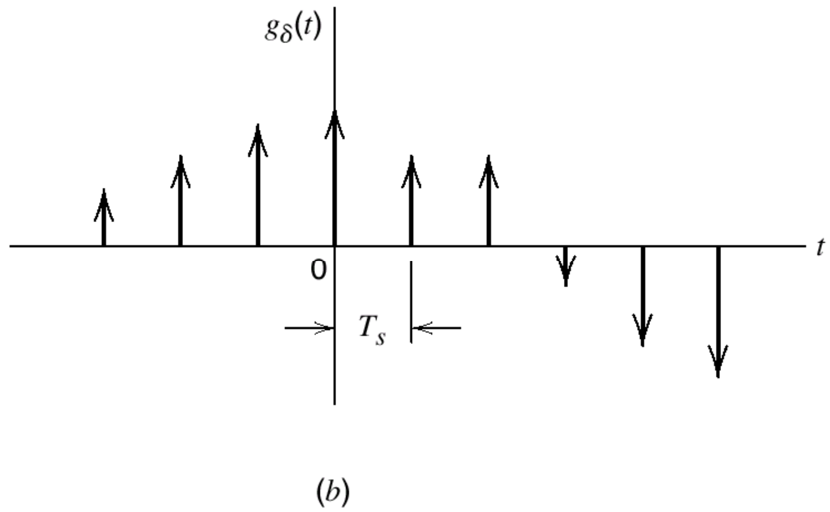
Why Digital?

- Advantages:
  - Digital signals are more immune to channel noise by using channel coding (perfect decoding is possible!)
  - Repeaters along the transmission path can detect a digital signal and retransmit a new noise-free signal
  - Digital signals derived from all types of analog sources can be represented using a uniform format
  - Digital signals are easier to process by using microprocessors and VLSI (e.g., digital signal processors, FPGA)
  - Digital systems are flexible and allow for implementation of sophisticated functions and control
  - More and more things are digital…
- For digital communication: analog signals are converted to digital.

Sampling

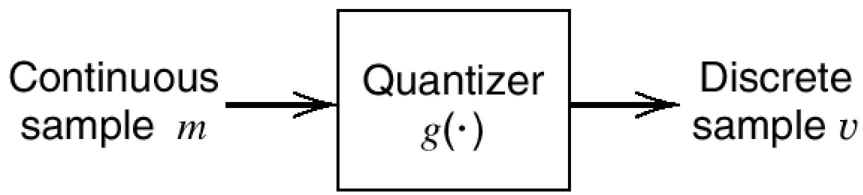
- How densely should we sample an analog signal so that we can reproduce its form accurately?
- A signal the spectrum of which is band-limited to W Hz, can be reconstructed exactly from its samples, if they are taken uniformly at a rate of $R \geq 2W$ Hz.
- Nyquist frequency: $f_s = 2W$ Hz



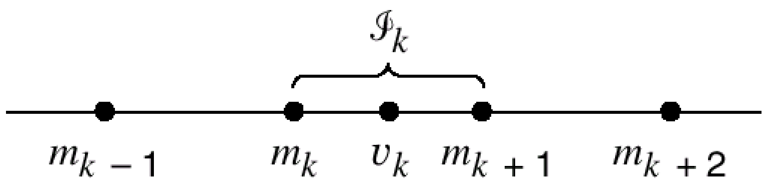


Quantization

- Quantization is the process of transforming the sample amplitude into a discrete amplitude taken from a finite set of possible amplitudes.
- The more levels, the better approximation.
- Don't need too many levels (human sense can only detect finite differences).
- Quantizers can be of a uniform or nonuniform type.



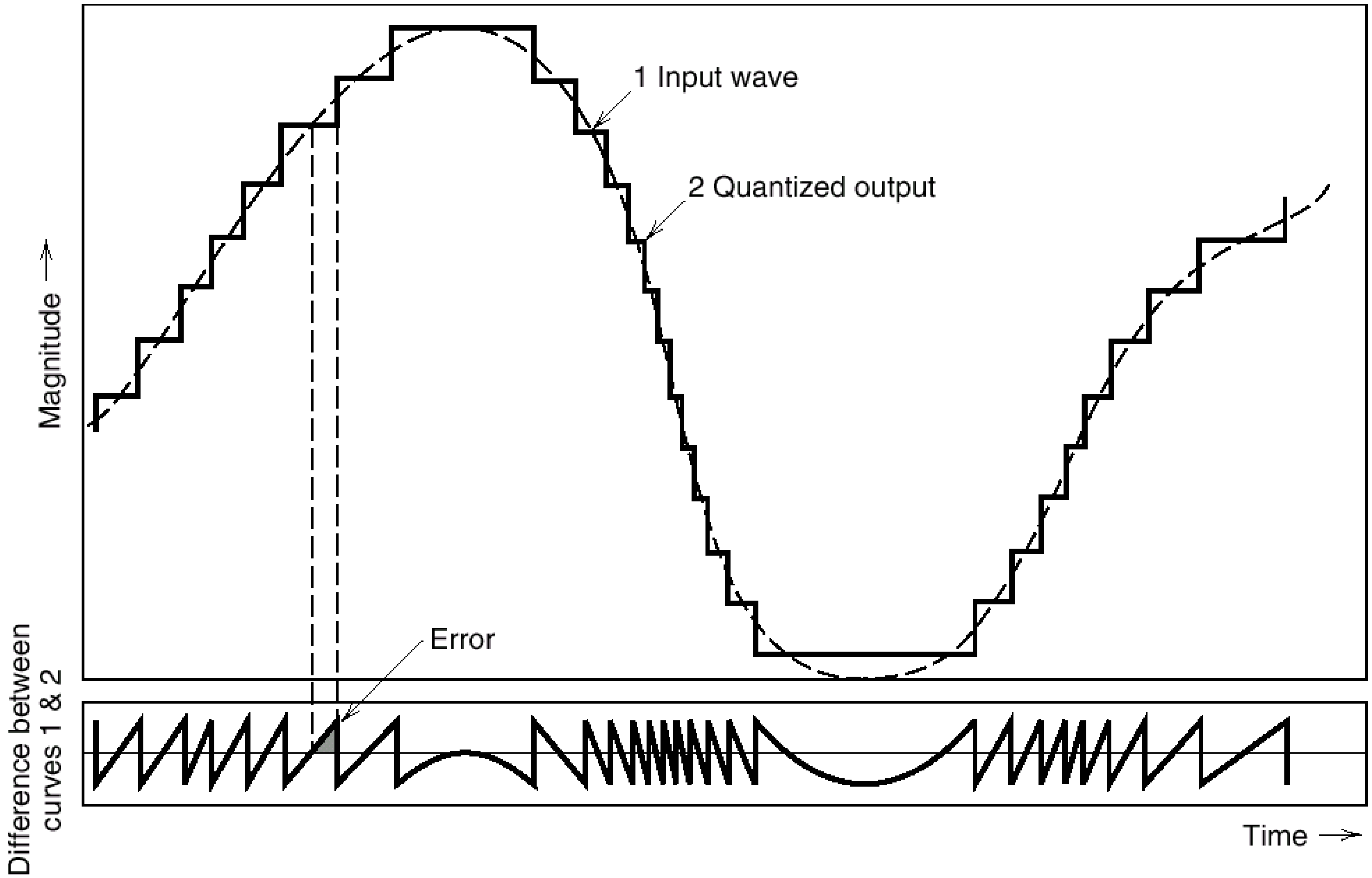
(a)



(b)

Quantization Noise

- Quantization noise: the error between the input signal and the output signal



Variance of Quantization Noise

- Δ: gap between quantizing levels (of a uniform quantizer)
- q: Quantization error = a random variable in the range

$$-\frac{\Delta}{2} \leq q \leq \frac{\Delta}{2}$$

- If Δ is sufficiently small, it is reasonable to assume that q is **uniformly distributed** over this range:

$$f_Q(q) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} \leq q \leq \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases}$$

- **Noise variance**

$$P_N = E\{e^2\} = \int_{-\infty}^{\infty} q^2 f_Q(q)dq = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} q^2 dq$$

$$= \frac{1}{\Delta}\frac{q^3}{3}\Big|_{-\Delta/2}^{\Delta/2} = \frac{1}{\Delta}\left[\frac{\Delta^3}{24} - \frac{(-\Delta)^3}{24}\right] = \frac{\Delta^2}{12}$$

SNR

- Assume: the encoded symbol has n bits
  - the maximum number of quantizing levels is $L = 2^n$
  - maximum peak-to-peak **dynamic range** of the quantizer $= 2^n \Delta$
- P: power of the message signal
- $m_p = \max |m(t)|$: maximum absolute value of the message signal
- Assume: the message signal fully loads the quantizer:

$$m_p = \frac{1}{2} 2^n \Delta = 2^{n-1} \Delta \tag{8.1}$$

- SNR at the quantizer output:

$$SNR_o = \frac{P_S}{P_N} = \frac{P}{\Delta^2 / 12} = \frac{12P}{\Delta^2}$$

SNR

- From (8.1) $$\Delta = \frac{m_p}{2^{n-1}} = \frac{2m_p}{2^n} \Rightarrow \Delta^2 = \frac{4m_p^2}{2^{2n}}$$
- $\Rightarrow$ $$SNR_o = \frac{12P}{\frac{4m_p^2}{2^{2n}}} = \frac{3P}{m_p^2} 2^{2n} \qquad (8.2)$$
-
- In dB,

$$SNR_o(\text{dB}) = 10\log_{10}(2^{2n}) + 10\log_{10}\left(\frac{3P}{m_p^2}\right)$$

$$= 20n\log_{10} 2 + 10\log_{10}\left(\frac{3P}{m_p^2}\right)$$

$$= 6n + 10\log_{10}\left(\frac{3P}{m_p^2}\right) \text{ (dB)}$$

- Hence, **each extra bit in the encoder adds 6 dB to the output SNR of the quantizer.**
- Recognize the tradeoff between SNR and *n* (i.e., rate, or bandwidth).

Example

- Sinusoidal message signal: $m(t) = A_m \cos(2\pi f_m t)$.
- Average signal power:

$$P = \frac{A_m^2}{2}$$

- Maximum signal value: $m_p = A_m$.
- Substitute into (8.2):

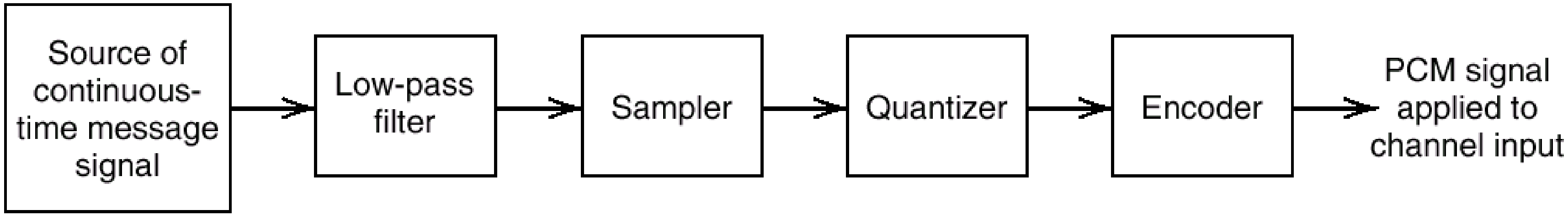
$$SNR_o = \frac{3A_m^2}{2A_m^2} 2^{2n} = \frac{3}{2} 2^{2n}$$

- In dB

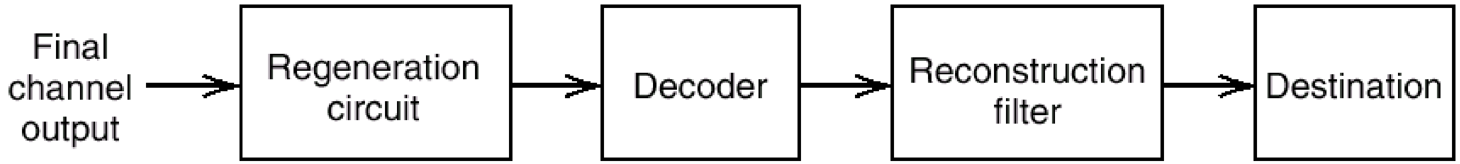
$$SNR_o(\text{dB}) = 6n + 1.8 \text{ dB}$$

- Audio CDs: $n = 16 \Rightarrow SNR > 90$ dB

Pulse-Coded Modulation (PCM)



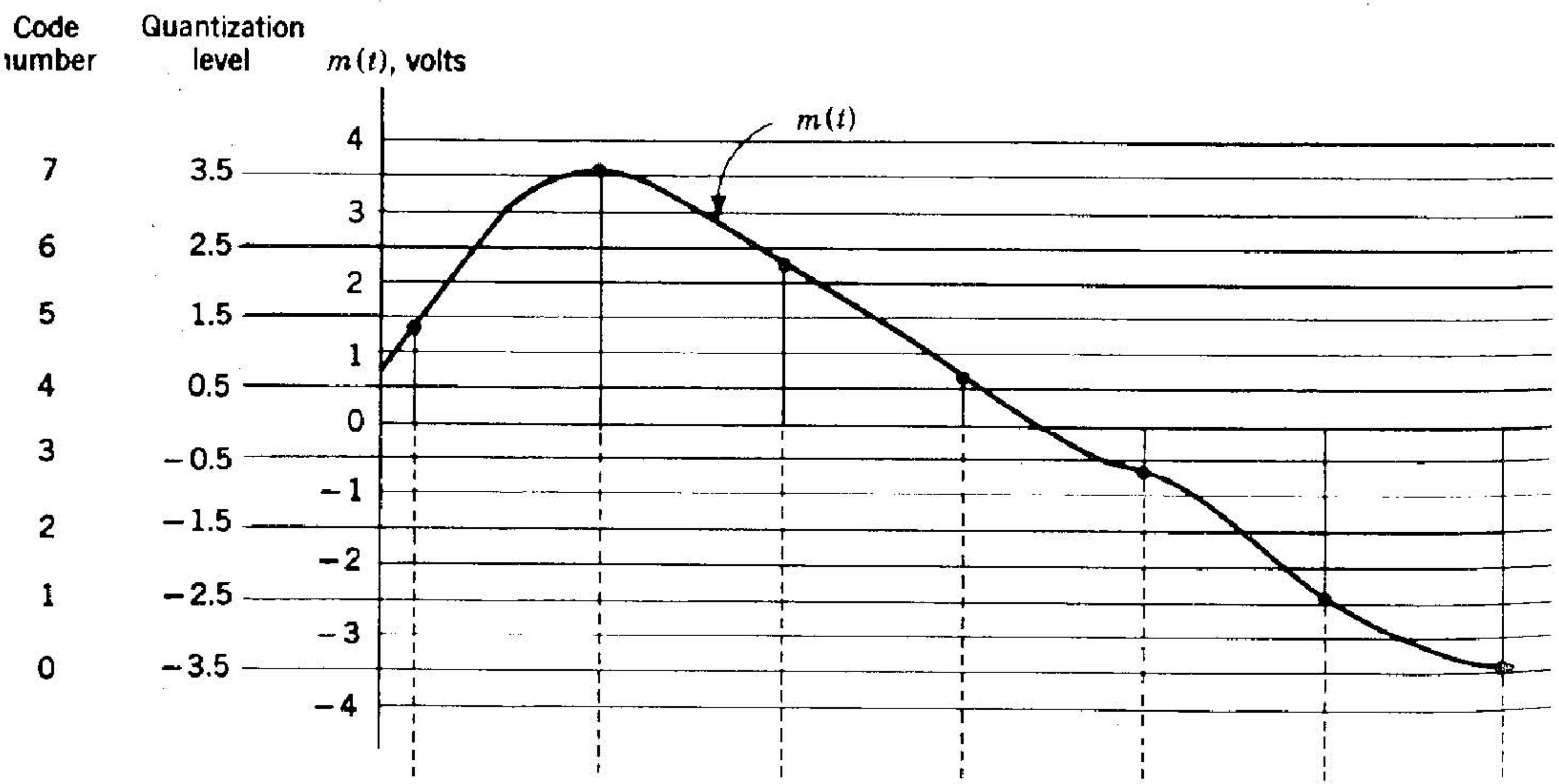
(a) Transmitter



(c) Receiver

- Sample the message signal above the Nyquist frequency
- Quantize the amplitude of each sample
- Encode the discrete amplitudes into a binary codeword
- *Caution:* PCM isn’t modulation in the usual sense; it’s a type of Analog-to-Digital Conversion.

The PCM Process



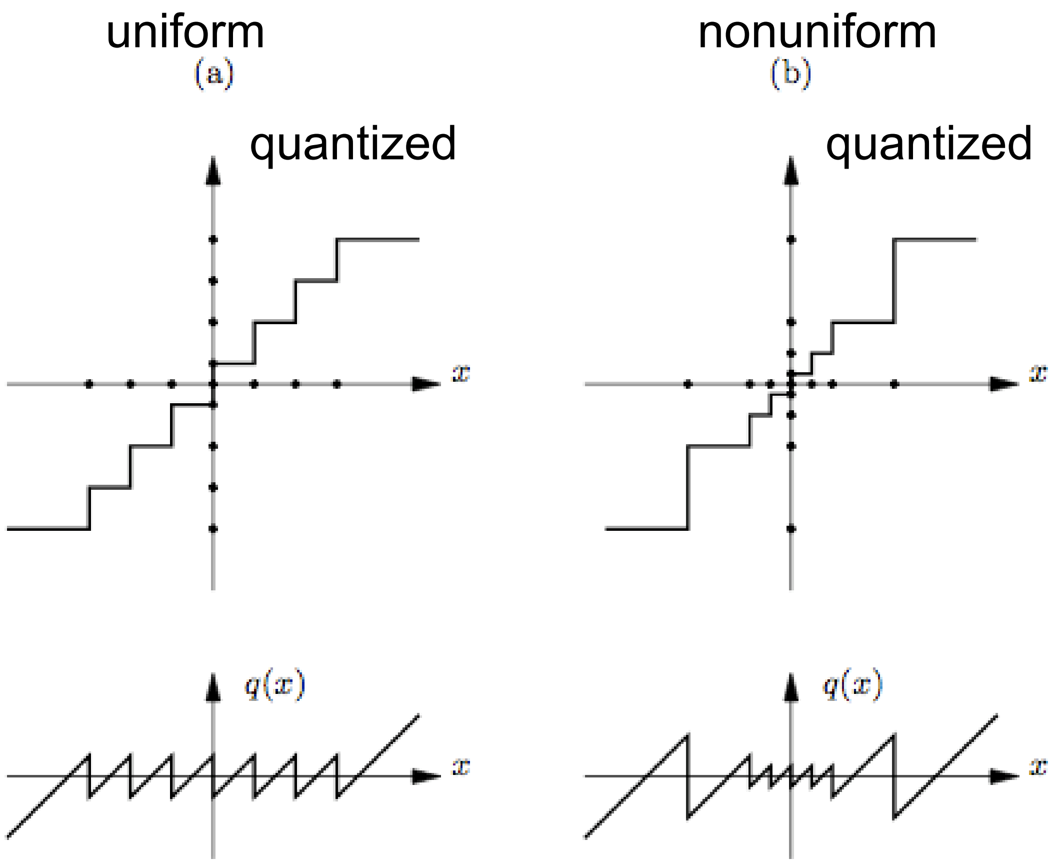
Sample value	1.3	3.6	2.3	0.7	−0.7	−2.4	−3.4
Nearest quantization level	1.5	3.5	2.5	0.5	−0.5	−2.5	−3.5
Code number	5	7	6	4	3	1	0
Binary representation	101	111	110	100	011	001	000

Problem With Uniform Quantization

- Problem: the output SNR is adversely affected by peak to average power ratio.
- Typically small signal amplitudes occur more often than large signal amplitudes.
  - The signal does not use the entire range of quantization levels available with equal probabilities.
  - Small amplitudes are not represented as well as large amplitudes, as they are more susceptible to quantization noise.

Solution: Nonuniform quantization

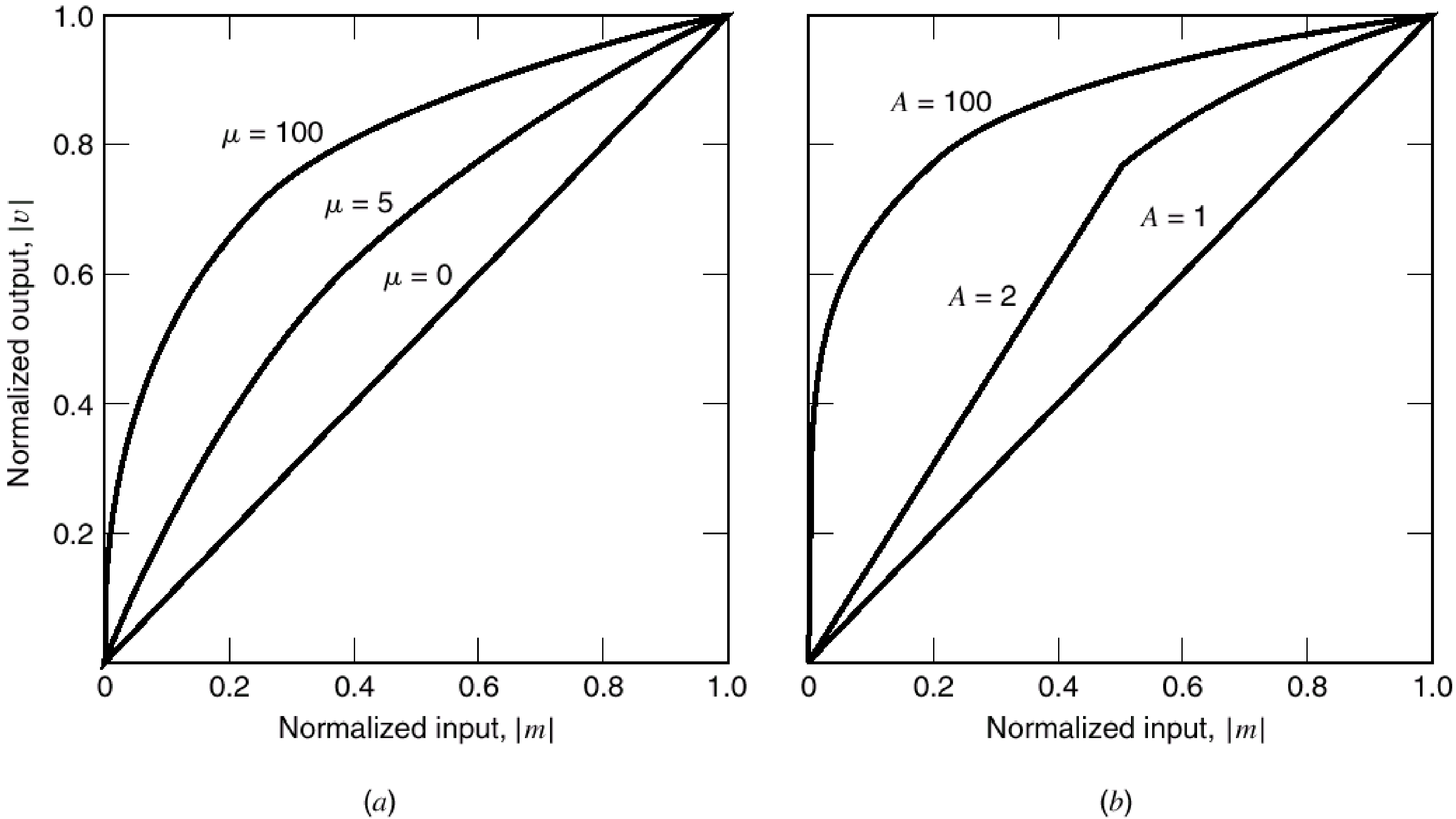
- Nonuniform quantization uses quantization levels of variable spacing, denser at small signal amplitudes, broader at large amplitudes.



Companding = Compressing + Expanding

- A practical (and equivalent) solution to nonuniform quantization:
  - Compress the signal first
  - Quantize it (using a uniform quantizer)
  - Transmit it
  - Expand it
- Companding is the corresponding to pre-emphasis and de-emphasis scheme used for FM.
- Predistort a message signal in order to achieve better performance in the presence of noise, and then remove the distortion at the receiver.
- The exact SNR gain obtained with companding depends on the exact form of the compression used.
- With proper companding, the output SNR can be made insensitive to peak to average power ratio.

μ-Law vs. A-Law



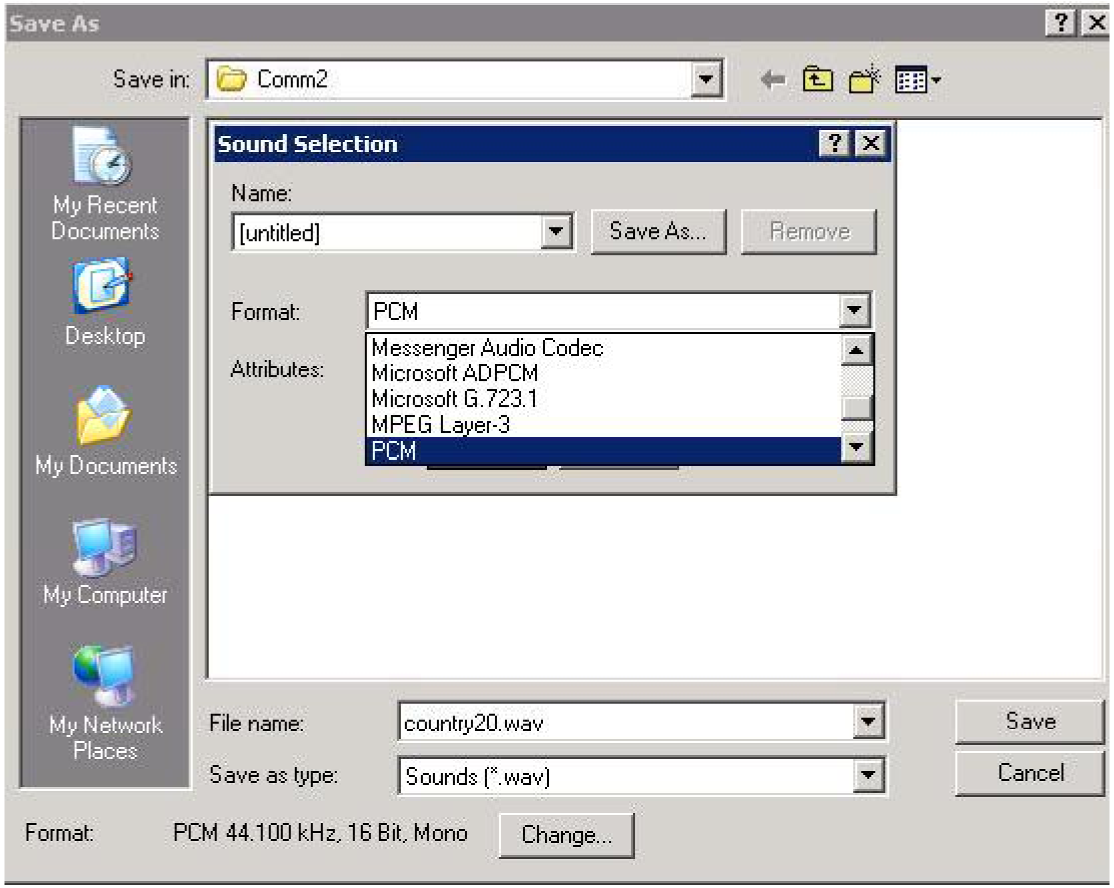
(a) μ-law used in North America and Japan, (b) A-law used in most countries of the world. Typical values in practice (T1/E1): μ = 255, A = 87.6.

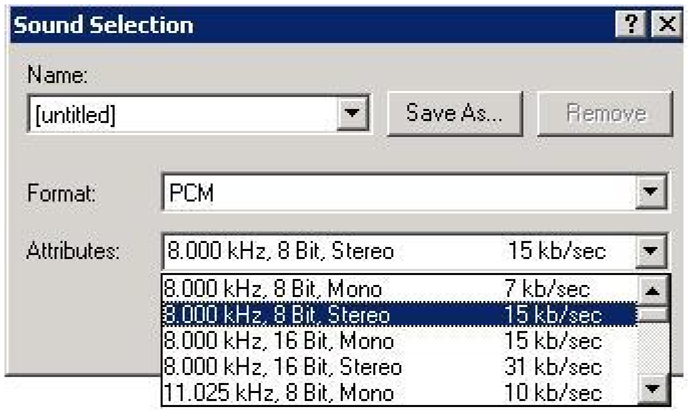
Applications of PCM & Variants

- Speech:
  - PCM: The voice signal is sampled at 8 kHz, quantized into 256 levels (8 bits). Thus, a telephone PCM signal requires 64 kbps.
    - need to reduce bandwidth requirements
  - DPCM (differential PCM): quantize the difference between consecutive samples; can save 8 to 16 kbps. ADPCM (Adaptive DPCM) can go further down to 32 kbps.
  - Delta modulation: 1-bit DPCM with oversampling; has even lower symbol rate (e.g., 24 kbps).
- Audio CD: 16-bit PCM at 44.1 kHz sampling rate.
- MPEG audio coding: 16-bit PCM at 48 kHz sampling rate compressed to a rate as low as 16 kbps.

Demo

- Music playing in Windows Sound Recorder





Summary

- Digitization of signals requires
  - Sampling: a signal of bandwidth W is sampled at the Nyquist frequency 2W.
  - Quantization: the link between analog waveforms and digital representation.
    - SNR (under high-resolution assumption)

$$SNR_o(\text{dB}) = 6n + 10\log_{10}\left(\frac{3P}{m_p^2}\right) \text{ (dB)}$$

    - Companding can improve SNR.
- PCM is a common method of representing audio signals.
  - In a strict sense, “pulse coded modulation” is in fact a (crude) source coding technique (i.e, method of digitally representing analog information).
  - There are more advanced source coding (compression) techniques in information theory.

Imperial College London

EE2-4: Communication Systems

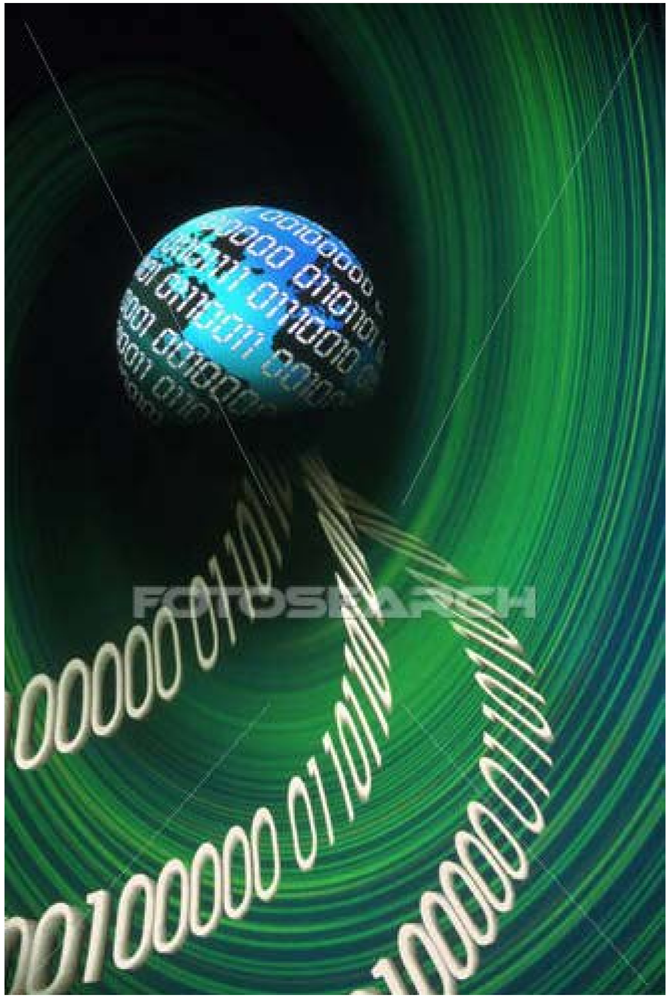
Lecture 9: Baseband Digital Transmission

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Line coding
- Performance of baseband digital transmission
  - Model
  - Bit error rate
- References
  - Notes of Communication Systems, Chap. 4.4
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 8
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 7



Line Coding

- The bits of PCM, DPCM etc need to be converted into some electrical signals.
- Line coding encodes the bit stream for transmission through a line, or a cable.
- Line coding was used former to the wide spread application of channel coding and modulation techniques.
- Nowadays, it is used for communications between the CPU and peripherals, and for short-distance baseband communications, such as the Ethernet.

Line Codes

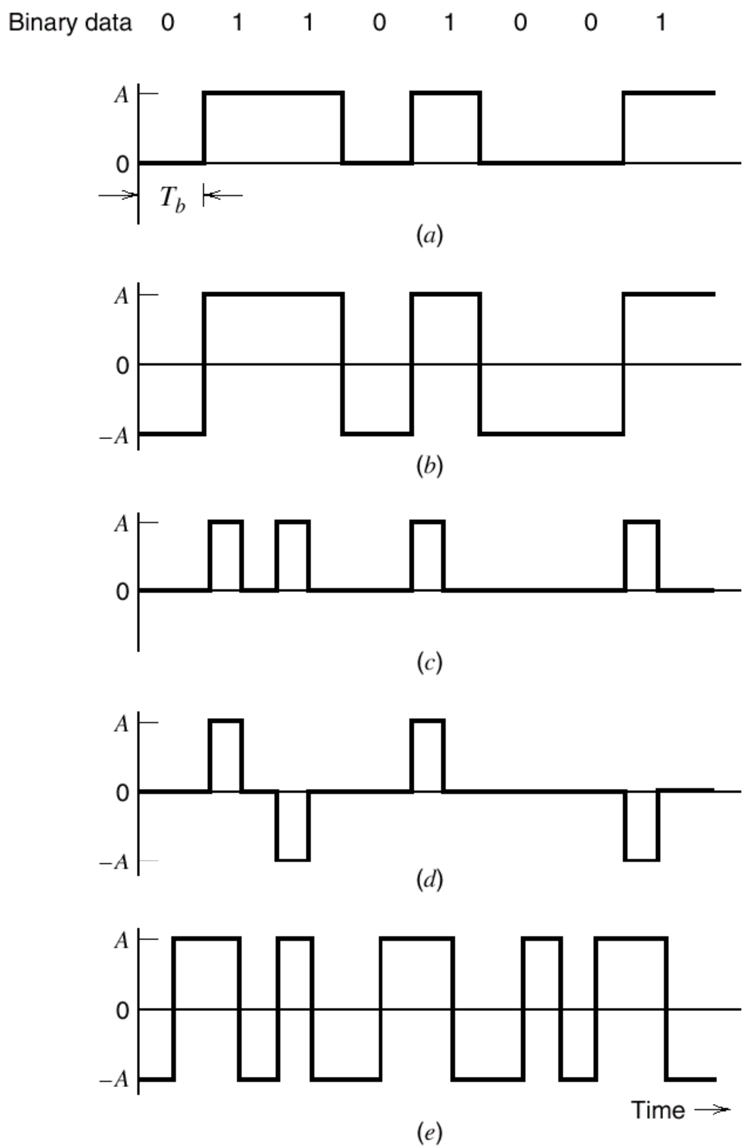
Unipolar nonreturn-to-zero (NRZ) signaling (on-off signaling)

Polar NRZ signaling

Unipolar Return-to-zero (RZ) signaling

Bipolar RZ signaling

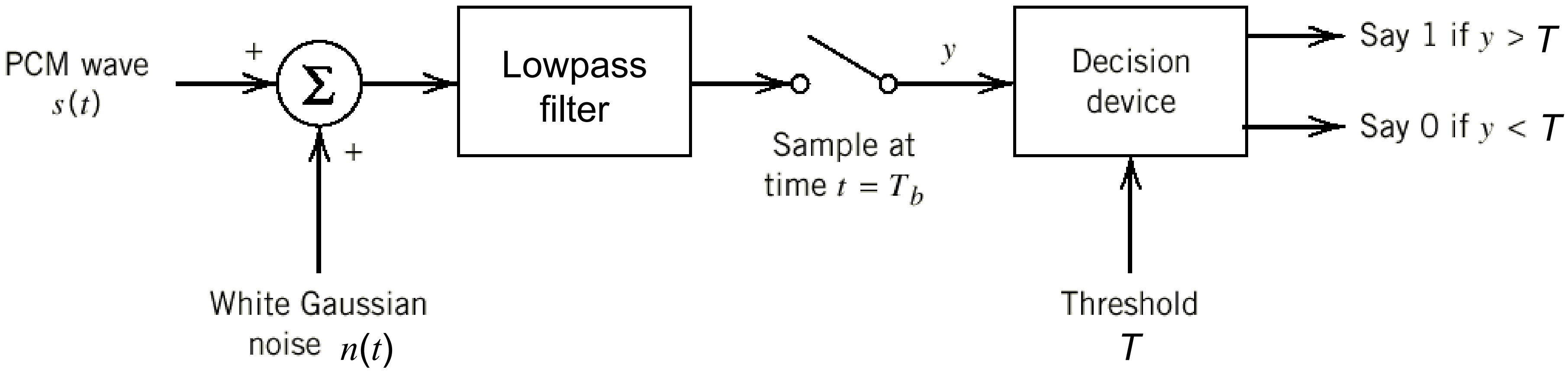
Manchester code



Analog and Digital Communications

- Different goals between analog and digital communication systems:
  - Analog communication systems: to reproduce the transmitted waveform accurately. ⇒ Use signal to noise ratio to assess the quality of the system
  - Digital communication systems: the transmitted symbol to be identified correctly by the receiver ⇒ Use the probability of error of the receiver to assess the quality of the system

Model of Binary Baseband Communication System



- We only consider binary PCM with on-off signaling: $0 \rightarrow 0$ and $1 \rightarrow A$ with bit duration T_b.
- **Assume:**
  - AWGN channel: The channel noise is additive white Gaussian, with a double-sided PSD of $N_0/2$.
  - The LPF is an ideal filter with unit gain on $[-W, W]$.
  - The signal passes through the LPF without distortion (approximately).

Distribution of Noise

- Effect of additive noise on digital transmission: at the receiver, symbol 1 may be mistaken for 0, and vice versa. ⇒ **bit errors**
- What is the probability of such an error?
- After the LPF, the predetection signal is

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

  - *s*(*t*): the binary-valued function (either 0 or *A* volts)
  - *n*(*t*): additive white Gaussian noise with zero mean and variance

$$\sigma^2 = \int_{-W}^{W} N_0 / 2 df = N_0 W$$

- Reminder: A sample value *N* of *n*(*t*) is a Gaussian random variable drawn from a probability density function (the **normal** distribution):

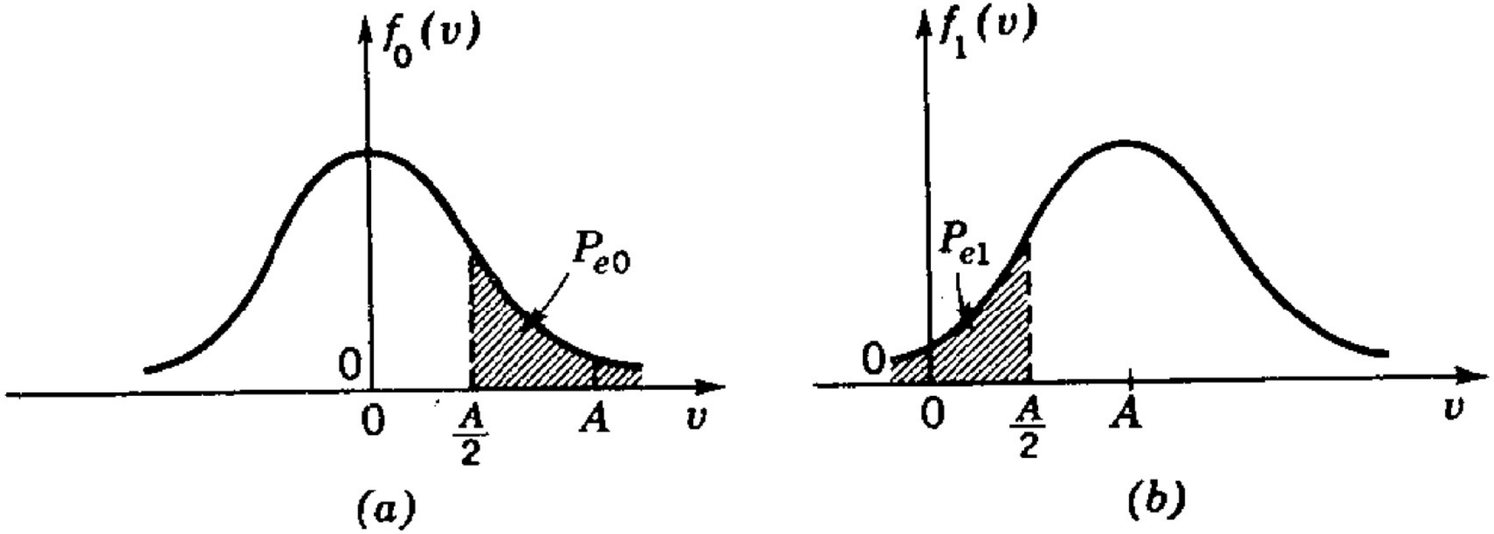
$$p_N(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n^2}{2\sigma^2} \right) = N(0, \sigma^2)$$

Decision

- Y: a sample value of $y(t)$
- If a symbol 0 were transmitted: $y(t) = n(t)$
  - Y will have a PDF of $N(0, \sigma^2)$
- If a symbol 1 were transmitted: $y(t) = A + n(t)$
  - Y will have a PDF of $N(A, \sigma^2)$
- Use as decision threshold T:
  - if $Y < T$, choose symbol 0
  - if $Y > T$, choose symbol 1

Errors

- Two cases of decision error:
  - (i) a symbol 0 was transmitted, but a symbol 1 was chosen
  - (ii) a symbol 1 was transmitted, but a symbol 0 was chosen



Probability density functions for binary data transmission in noise:
(a) symbol 0 transmitted, and (b) symbol 1 transmitted. Here $T = A/2$.

Case (i)

- Probability of (i) occurring = (Probability of an error, given symbol 0 was transmitted) $\times$ (Probability of a 0 to be transmitted in the first place):

$$p(\text{i}) = P_{e0} \times p_0$$

where:

  - p_0: the **a priori** probability of transmitting a symbol 0
  - P_{e0}: the conditional probability of error, given that symbol 0 was transmitted:

$$P_{e0} = \int_T^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n^2}{2\sigma^2}\right) dn$$

Case (ii)

- Probability of (ii) occurring = (Probability of an error, given symbol 1 was transmitted) × (Probability of a 1 to be transmitted in the first place):

$$p(\text{ii}) = P_{e1} \times p_1$$

where:

  - p_1: the **a priori** probability of transmitting a symbol 1
  - P_{e1}: the conditional probability of error, given that symbol 0 was transmitted:

$$P_{e1} = \int_{-\infty}^{T} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(n-A)^2}{2\sigma^2} \right) dn$$

Total Error Probability

- Total error probability:

$$P_e(T) = p_{(i)} + p_{(ii)}$$

$$= p_1 P_{e1} + (1 - p_1) P_{e0}$$

$$= p_1 \int_{-\infty}^{T} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(n-A)^2}{2\sigma^2} \right) dn + (1 - p_1) \int_{T}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n^2}{2\sigma^2} \right) dn$$

- Choose T so that $P_e(T)$ is minimum:

$$\frac{dP_e(T)}{dT} = 0$$

Derivation

- Leibnitz rule of differentiating an integral with respect to a parameter: if

$$I(\lambda)=\int_{a(\lambda)}^{b(\lambda)} f(x;\lambda)dx$$

- then

$$\frac{dI(\lambda)}{d\lambda}=\frac{db(\lambda)}{d\lambda}f(b(\lambda);\lambda)-\frac{da(\lambda)}{d\lambda}f(a(\lambda);\lambda)+\int_{a(\lambda)}^{b(\lambda)}\frac{\partial f(x;\lambda)}{\partial\lambda}dx$$

- Therefore

$$\frac{dP_e(T)}{dT}=p_1\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(T-A)^2}{2\sigma^2}\right)-(1-p_1)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{T^2}{2\sigma^2}\right)=0$$

$$\frac{p_1}{1-p_1}=\exp\left(-\frac{T^2-(T-A)^2}{2\sigma^2}\right)$$

$$=\exp\left(-\frac{T^2-T^2-A^2+2TA}{2\sigma^2}\right)=\exp\left(-\frac{A(2T-A)}{2\sigma^2}\right)$$

Optimum Threshold

$$\ln\frac{p_1}{1-p_1} = -\frac{A(2T-A)}{2\sigma^2} \quad \Rightarrow \quad 2\sigma^2 \ln\frac{p_1}{1-p_1} = -A(2T-A) \quad \Rightarrow$$

$$-\frac{2\sigma^2}{A}\ln\frac{p_1}{1-p_1} = 2T - A \quad \Rightarrow \quad T = -\frac{\sigma^2}{A}\ln\frac{p_1}{1-p_1} + \frac{A}{2}$$

- Equi-probable symbols $(p_1 = p_0 = 1 - p_1) \Rightarrow T = A/2$.
- For equi-probable symbols, it can be shown that $P_{e0} = P_{e1}$.
- Probability of total error:

$$P_e = p_{(i)} + p_{(ii)} = p_0 P_{e0} + p_1 P_{e1} = P_{e0} = P_{e1}$$

since $p_0 = p_1 = 1/2$, and $P_{e0} = P_{e1}$.

Calculation of P_e

- Define a new variable of integration

$$z \equiv \frac{n}{\sigma} \Rightarrow dz = \frac{1}{\sigma} dn \Rightarrow dn = \sigma dz$$

  - When $n = A/2,\ z = A/(2\sigma)$.
  - When $n = \infty,\ z = \infty$.
- Then

$$P_{e0} = \frac{1}{\sigma\sqrt{2\pi}} \int_{A/(2\sigma)}^{\infty} e^{-z^2/2} \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{A/(2\sigma)}^{\infty} e^{-z^2/2} dz$$

- We may express P_{e0} in terms of the Q-function:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

- Then:

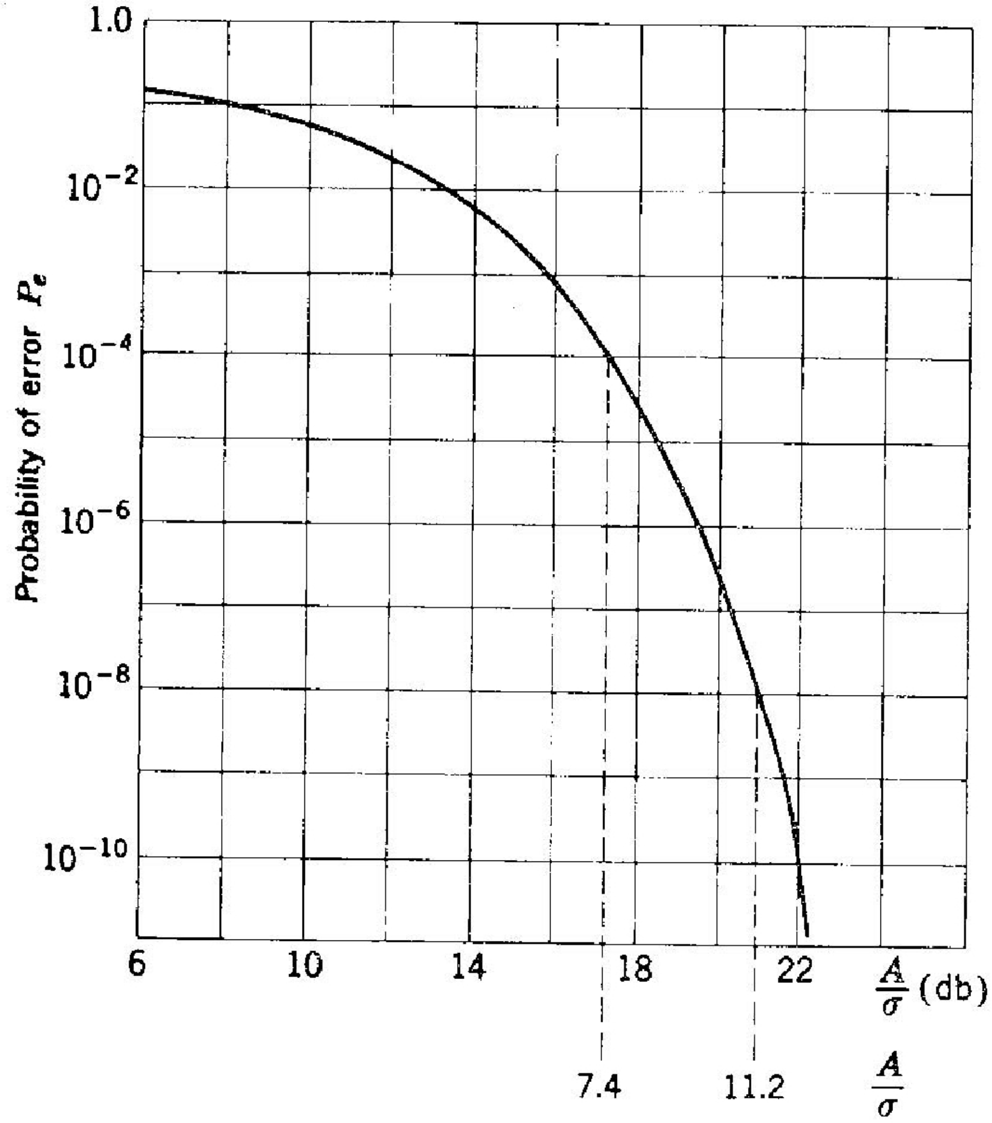
$$\boxed{P_e = Q\left(\frac{A}{2\sigma}\right)}$$

Example

- **Example 1:** $A/\sigma = 7.4 \Rightarrow P_e = 10^{-4}$
- ⇒ For a transmission rate is 10^5 bits/sec, there will be an error every 0.1 seconds
- **Example 2:** $A/\sigma = 11.2 \Rightarrow P_e = 10^{-8}$
- ⇒ For a transmission rate is 10^5 bits/sec, there will be an error every 17 mins

- A/σ = 7.4: Corresponds to 20 $\log_{10}(7.4)$ = 17.4 dB
- A/σ = 11.2: Corresponds to 20 $\log_{10}(11.2)$ = 21 dB
- ⇒ Enormous increase in reliability by a relatively small increase in SNR (if that is affordable).

Probability of Bit Error



Q-function

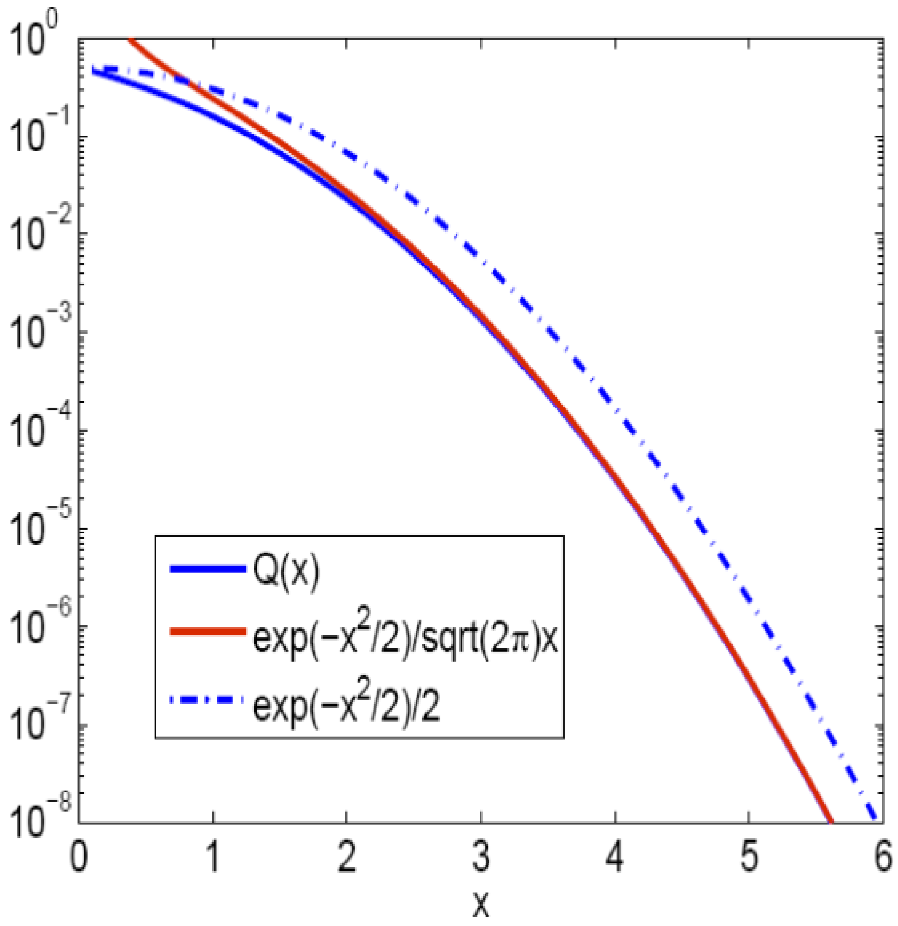
- Upper bounds and good approximations:

$$Q(x) \le \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, x \ge 0$$

  - which becomes tighter for large *x*, and

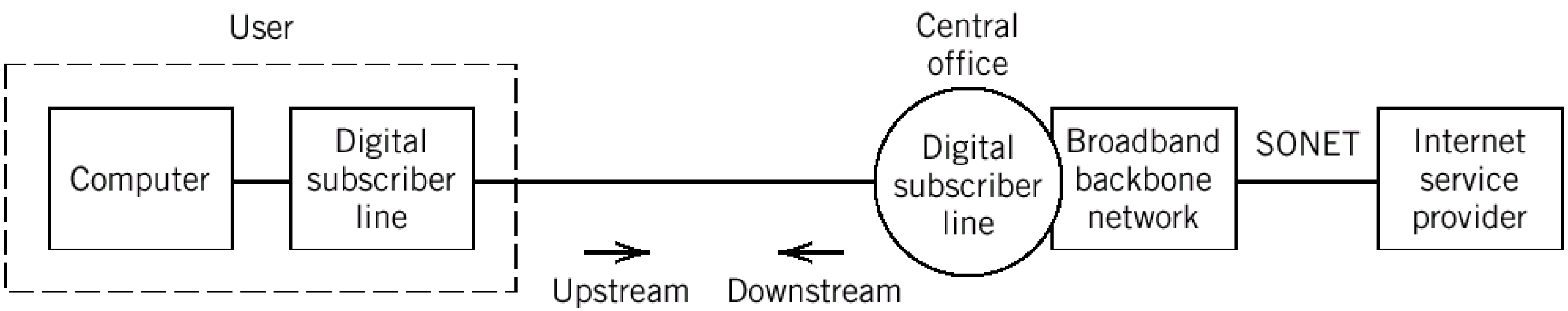
$$Q(x) \le \frac{1}{2} e^{-x^2/2}, x \ge 0$$

  - which is a better upper bound for small *x*.



Applications to Ethernet and DSL

- 100BASE-TX
  - One of the dominant forms of fast Ethernet, transmitting up to 100 Mbps over twisted pair.
  - The link is a maximum of 100 meters in length.
  - Line codes: NRZ → NRZ Invert → MLT-3.
- Digital subscriber line (DSL)
  - Broadband over twist pair.
  - Line code 2B1Q achieved bit error rate of 10^{-7} at 160 kbps.
  - ADSL and VDSL adopt discrete multitone modulation (DMT) for higher data rates.



Imperial College London

EE2-4: Communication Systems

Lecture 10: Digital Modulation

Dr. Cong Ling

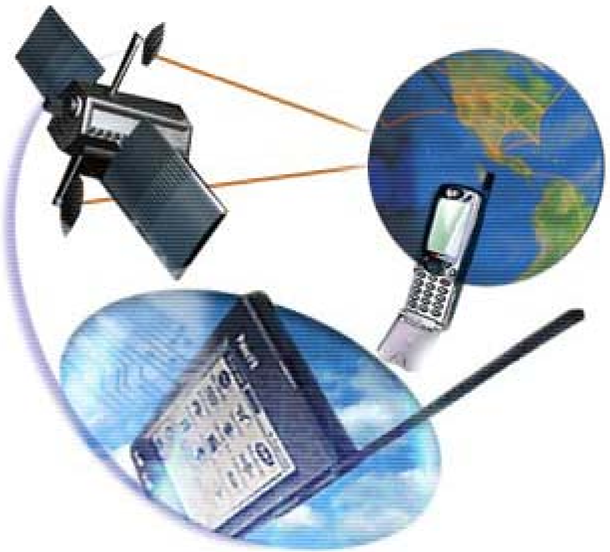
Department of Electrical and Electronic Engineering

Outline

- ASK and bit error rate

- FSK and bit error rate

- PSK and bit error rate



- References
  - Notes of Communication Systems, Chap. 4.5
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 9
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 13

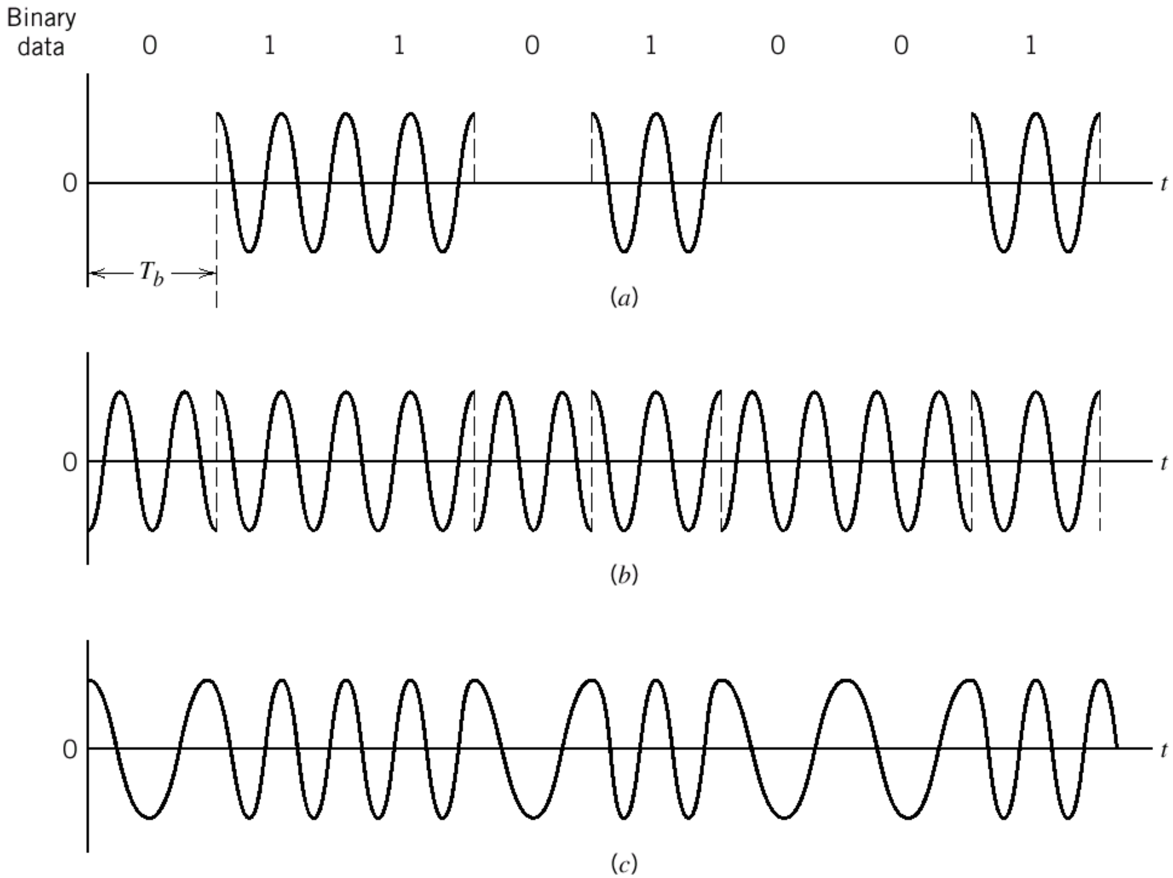
Digital Modulation

- Three Basic Forms of Signaling Binary Information

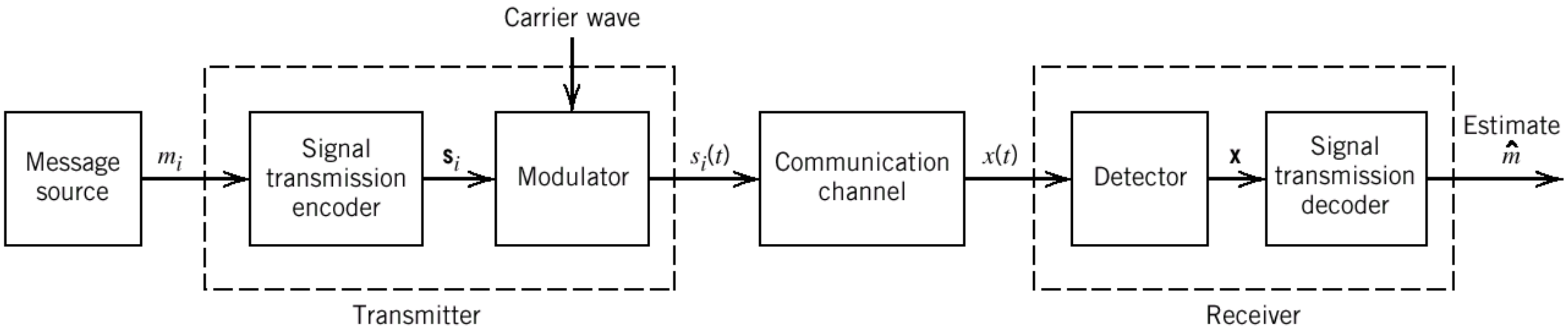
(a) Amplitude-shift keying (ASK).

(b) Phase-shift keying (PSK).

(c) Frequency-shift keying (FSK).



Demodulation



- Coherent (synchronous) demodulation/detection
  - Use a BPF to reject out-of-band noise
  - Multiply the incoming waveform with a cosine of the carrier frequency
  - Use a LPF
  - Requires carrier regeneration (both frequency and phase synchronization by using a phase-lock loop)
- Noncoherent demodulation (envelope detection etc.)
  - Makes no explicit efforts to estimate the phase

ASK

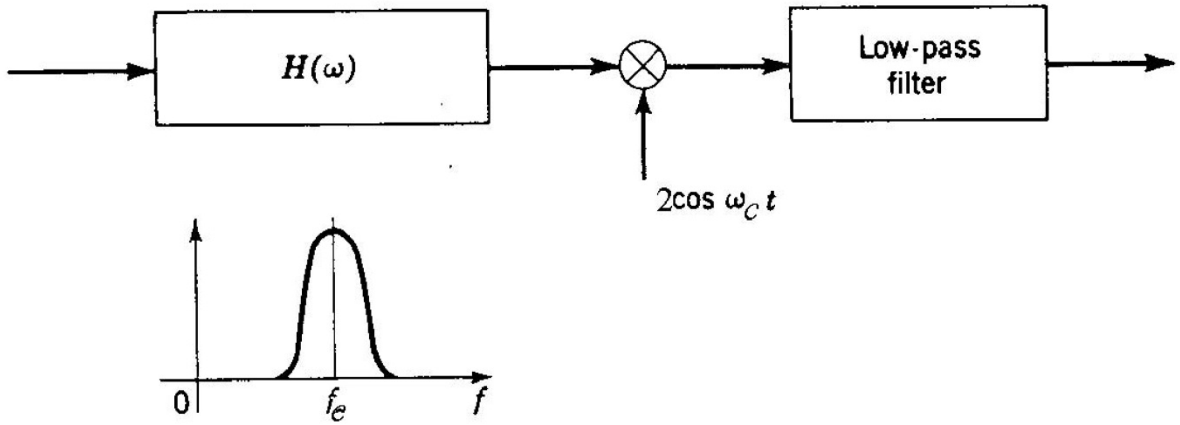
- Amplitude shift keying (ASK) = on-off keying (OOK)

$$s_0(t) = 0$$
$$s_1(t) = A\cos(2\pi f_c t)$$

or

$$s(t) = A(t)\cos(2\pi f_c t), \qquad A(t) \in \{0, A\}$$

- Coherent detection



- Assume an ideal band-pass filter with unit gain on $[f_c - W, f_c + W]$. For a practical band-pass filter, $2W$ should be interpreted as the equivalent bandwidth.

Coherent Demodulation

- Pre-detection signal:

$$x(t) = s(t) + n(t)$$
$$= A\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$
$$= [A(t) + n_c(t)]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

- After multiplication with $2\cos(2\pi f_c t)$:

$$y(t) = [A(t) + n_c(t)]2\cos^2(2\pi f_c t) - n_s(t)2\sin(2\pi f_c t)\cos(2\pi f_c t)$$
$$= [A(t) + n_c(t)](1 + \cos(4\pi f_c t)) - n_s(t)\sin(4\pi f_c t)$$

- After low-pass filtering:

$$\tilde{y}(t) = A(t) + n_c(t)$$

Bit Error Rate

- **Reminder:** The in-phase noise component $n_c(t)$ has the same variance as the original band-pass noise $n(t)$
  - The received signal is identical to that for baseband digital transmission
  - The sample values of $\tilde{y}(t)$ will have PDFs that are identical to those of the baseband case
- For ASK the statistics of the receiver signal are identical to those of a baseband system
- The probability of error for ASK is the same as for the baseband case
- Assume equiprobable transmission of 0s and 1s.
- Then the decision threshold must be *A/*2 and the probability of error is given by:

$$P_{e,ASK} = Q\left(\frac{A}{2\sigma}\right)$$

PSK

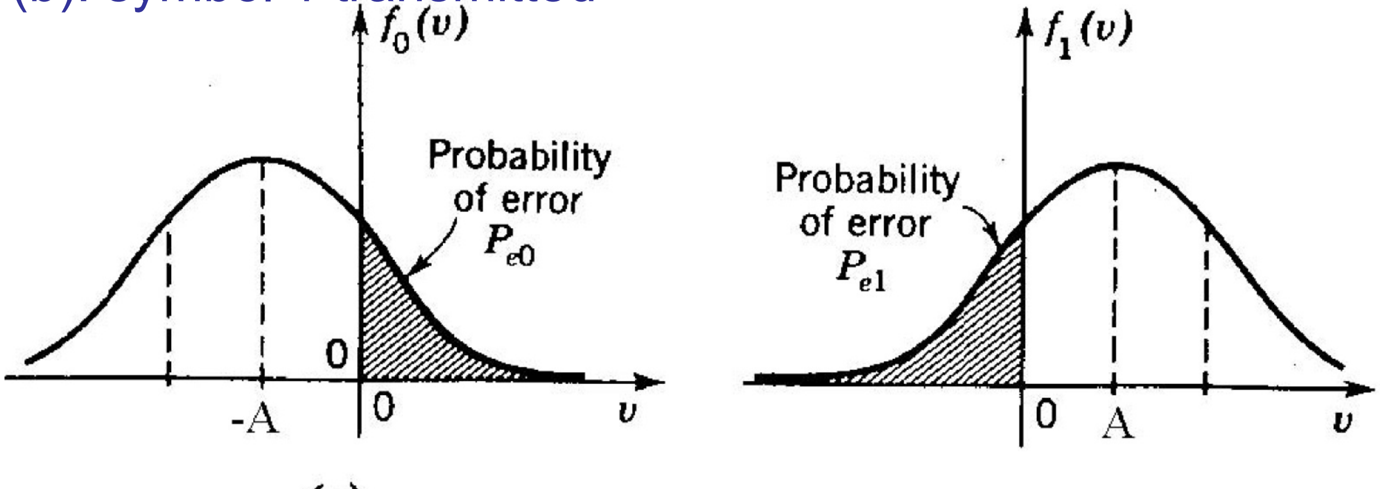
- Phase shift keying (PSK)

$$s(t) = A(t)\cos(2\pi f_c t), \qquad A(t) \in \{-A, A\}$$

- Use coherent detection again, to eventually get the detection signal:

$$\tilde{y}(t) = A(t) + n_c(t)$$

- Probability density functions for PSK for equiprobable 0s and 1s in noise (use threshold 0 for detection):
  - (a): symbol 0 transmitted
  - (b): symbol 1 transmitted



(a) (b)

Analysis

- Conditional error probabilities:

$$P_{e0} = \int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(n+A)^2}{2\sigma^2}\right) dn$$

$$P_{e1} = \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(n-A)^2}{2\sigma^2}\right) dn$$

- In the **first** set $\tilde{n} \equiv n + A \Rightarrow dn = d\tilde{n}$ and when $n = 0, \tilde{n} = A$

$$P_{e0} = \frac{1}{\sigma\sqrt{2\pi}} \int_A^{\infty} \exp\left(-\frac{\tilde{n}^2}{2\sigma^2}\right) d\tilde{n}$$

- In the **second** set $\tilde{n} \equiv -(n - A) = -n + A \Rightarrow dn = -d\tilde{n}$

  when $n = 0, \tilde{n} = A,$ and when $n = -\infty, \tilde{n} = +\infty$:

$$P_{e1} = \frac{1}{\sigma\sqrt{2\pi}} \int_{\infty}^{A} \exp\left(-\frac{\tilde{n}^2}{2\sigma^2}\right)(-1) d\tilde{n} = \frac{1}{\sigma\sqrt{2\pi}} \int_A^{\infty} \exp\left(-\frac{\tilde{n}^2}{2\sigma^2}\right) d\tilde{n}$$

Bit Error Rate

- So:

$$P_{e0} = P_{e1} = P_{e,PSK} = \int_A^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n^2}{2\sigma^2}\right) dn$$

- Change variable of integration to $z \equiv n/\sigma \Rightarrow dn = \sigma dz$ and when $n = A,\ z = A/\sigma$. Then:

$$P_{e,PSK} = \frac{1}{\sqrt{2\pi}} \int_{\frac{A}{\sigma}}^\infty e^{-z^2/2} dz = Q\left(\frac{A}{\sigma}\right)$$

- **Remember** that

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2) dt$$

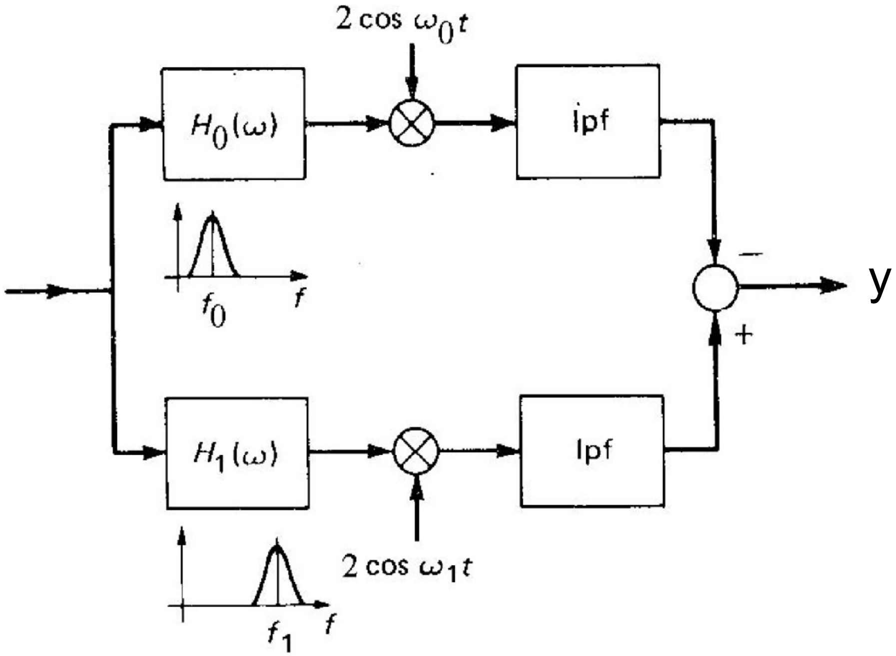
FSK

- Frequency Shift Keying (FSK)

$$s_0(t) = A\cos(2\pi f_0 t), \quad \text{if symbol 0 is transmitted}$$
$$s_1(t) = A\cos(2\pi f_1 t), \quad \text{if symbol 1 is transmitted}$$

- Symbol recovery:
  - Use two sets of coherent detectors, one operating at a frequency f_0 and the other at f_1.

Coherent FSK demodulation. The two BPF's are non-overlapping in frequency spectrum



Output

- Each branch = an ASK detector

$$\text{LPF output on each branch} = \begin{cases} A + \text{noise} & \text{if symbol present} \\ \text{noise} & \text{if symbol not present} \end{cases}$$

- $n_0(t)$: the noise output of the top branch
- $n_1(t)$: the noise output of the bottom branch
- Each of these noise terms has identical statistics to $n_c(t)$.
- Output if a symbol 1 were transmitted

$$y = y_1(t) = A + [n_1(t) - n_0(t)]$$

- Output if a symbol 0 were transmitted

$$y = y_0(t) = -A + [n_1(t) - n_0(t)]$$

Bit Error Rate for FSK

- Set detection threshold to 0
- Difference from PSK: the noise term is now $n_1(t) - n_0(t)$.

- The noises in the two channels are independent because their spectra are non-overlapping.
  - the proof is done in the problem sheet.
  - the variances add.
  - the noise variance has doubled!

- Replace σ^2 in (172) by $2\sigma^2$ (or σ by $\sqrt{2}\,\sigma$)

$$P_{e,FSK} = Q\left(\frac{A}{\sqrt{2}\sigma}\right)$$

The Sum of Two R.V.

- Noise is the sum or difference of two independent zero mean random variables:
  - x_1: a random variable with variance σ_1^2
  - x_2: a random variables with variance σ_2^2
- What is the variance of $y \equiv x_1 \pm x_2$?
- By definition

$$\sigma_y^2 = E\{y^2\} - E\{y\}^2 = E\{(x_1 \pm x_2)^2\}$$

$$= E\{x_1^2 \pm 2x_1x_2 + x_2^2\} = E\{x_1^2\} \pm E\{x_1x_2\} + E\{x_2^2\}$$

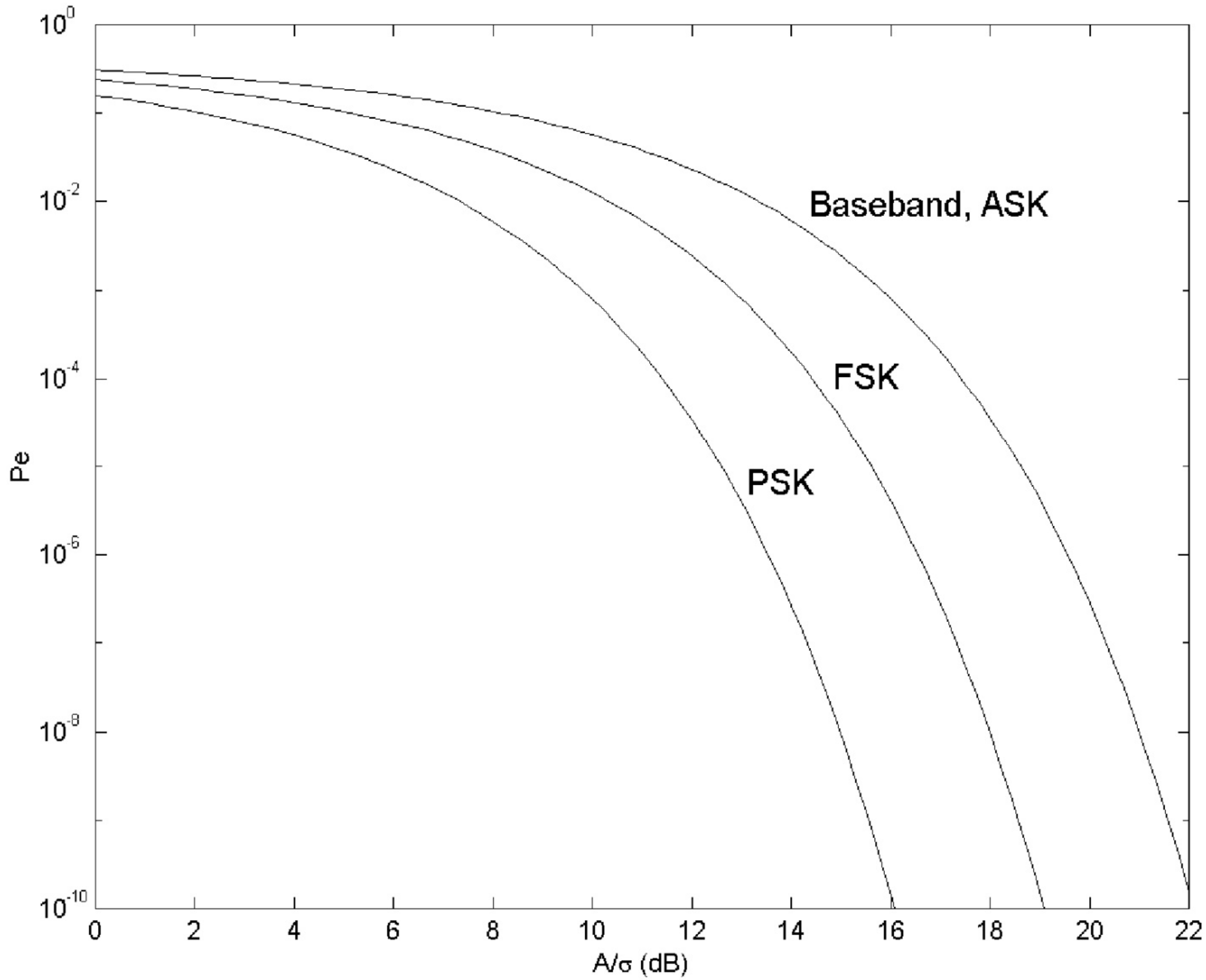
- For independent variables: $E\{x_1x_2\} = E\{x_1\}E\{x_2\}$
- For zero-mean random variables:

$$E\{x_1\} = E\{x_2\} = 0 \Rightarrow E\{x_1x_2\} = 0$$

- So

$$\sigma_y^2 = E\{x_1^2\} + E\{x_2^2\} = \sigma_1^2 + \sigma_2^2$$

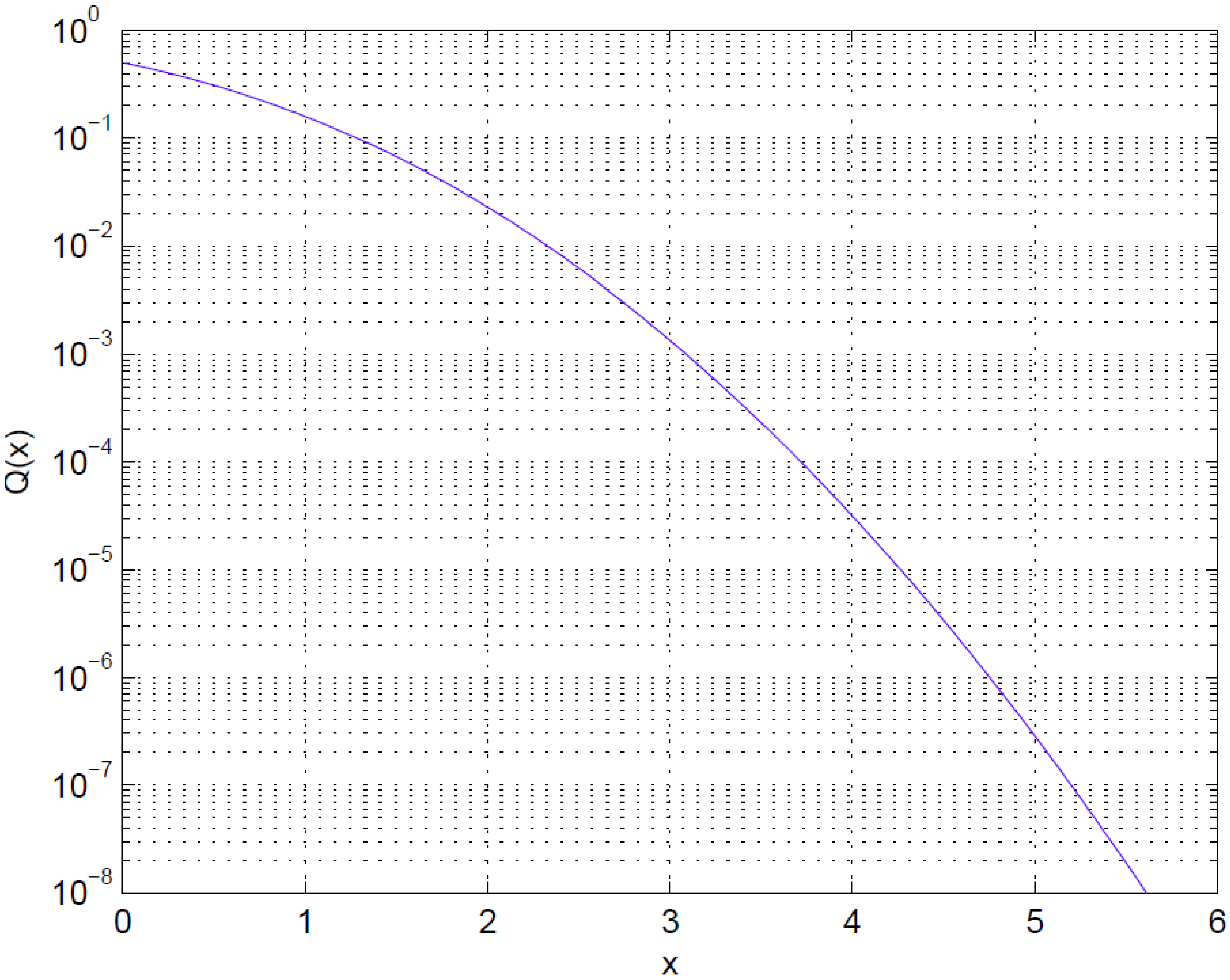
Comparison of Three Schemes



Comment

- To achieve the same error probability (fixed P_e):
- PSK can be reduced by 6 dB compared with a baseband or ASK system (a factor of 2 reduction in amplitude)
- FSK can be reduced by 3 dB compared with a baseband or ASK (a factor of $\sqrt{2}$ reduction in amplitude)
- Caution: The comparison is based on *peak* SNR. In terms of average SNR, PSK only has a 3 dB improvement over ASK, and FSK has the same performance as ASK

Q-function



Examples

- Consider binary PSK modulation. Assume the carrier amplitude A = 1 v, and noise standard deviation $\sigma = 1/3$. Determine the bit error probability.
  – Answer: $P_e = 1.03 \times 10^{-3}$.

- Now suppose the bit error probability is 10^{-5}. Determine the value of A/σ.
  – Answer: $A/\sigma = 4.3$.

Application: GMSK

- Gaussian minimum shift keying (GMSK), a special form of FSK preceded by Gaussian filtering, is used in GSM (Global Systems for Mobile Communications), a leading cellular phone standard in the world.
  - Also known as digital FM, built on some of FM-related advantages of AMPS, the first-generation analog system (30 KHz bandwidth).
  - Binary data are passed through a Gaussian filter to satisfy stringent requirements of out-of-band radiation.
  - Minimum Shift Keying: its spacing between the two frequencies of FSK is minimum in a certain sense (see problem sheet).
  - GMSK is allocated bandwidth of 200 kHz, shared among 32 users. This provides a (30/200)x32=4.8 times improvement over AMPS.

Imperial College London

EE2-4: Communication Systems

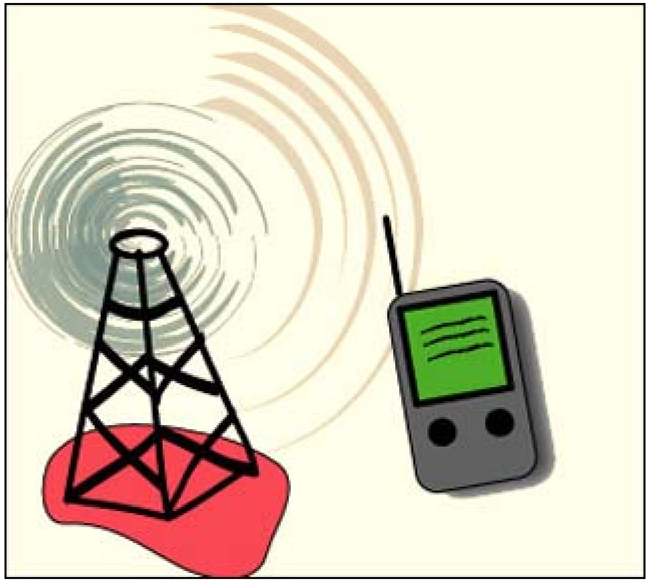
Lecture 11: Noncoherent Demodulation

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

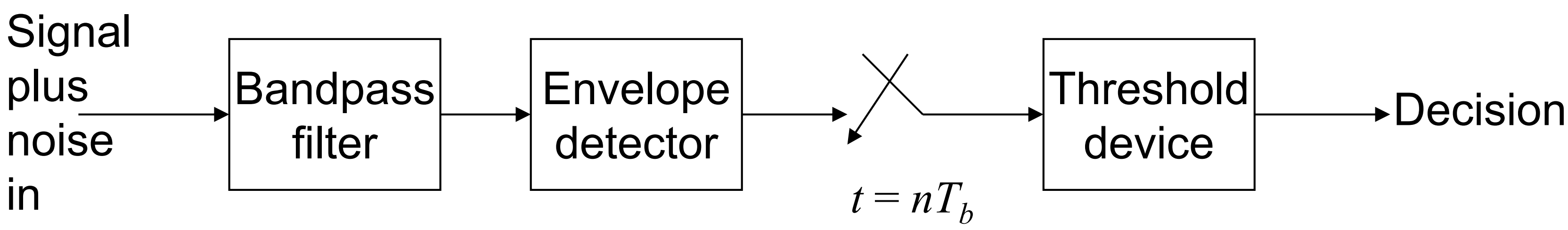
- Noncoherent demodulation of ASK
- Noncoherent demodulation of FSK
- Differential demodulation of DPSK
- References
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 9
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 13



Noncoherent Demodulation

- Coherent demodulation assumes perfect synchronization.
  - Needs a phase lock loop.
- However, accurate phase synchronization may be difficult in a dynamic channel.
  - Phase synchronization error is due to varying propagation delays, frequency drift, instability of the local oscillator, effects of strong noise ...
  - Performance of coherent detection will degrade severely.
- When the carrier phase is unknown, one must rely on non-coherent detection.
  - No provision is made for carrier phase recovery.
- The phase Φ is assumed to be uniformly distributed on $[0, 2\pi]$.
- Circuitry is simpler, but analysis is more difficult!

Noncoherent Demodulation of ASK



- Output of the BPF

$$y(t) = n(t) \quad \text{when 0 is sent}$$
$$y(t) = n(t) + A\cos(2\pi f_c t) \quad \text{when 1 is sent}$$

- Recall

$$n(t) = n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

- Envelope

$$R = \sqrt{n_c^2(t) + n_s^2(t)} \quad \text{when 0 is sent}$$
$$R = \sqrt{(A + n_c(t))^2 + n_s^2(t)} \quad \text{when 1 is sent}$$

Distribution of the Envelope

- When symbol 0 is sent, the envelope (that of the bandpass noise alone) has Rayleigh distribution

$$f(r) = \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}, \quad r \geq 0$$

- When symbol 1 is sent, the envelope (that of a signal + bandpass noise) has Rician distribution

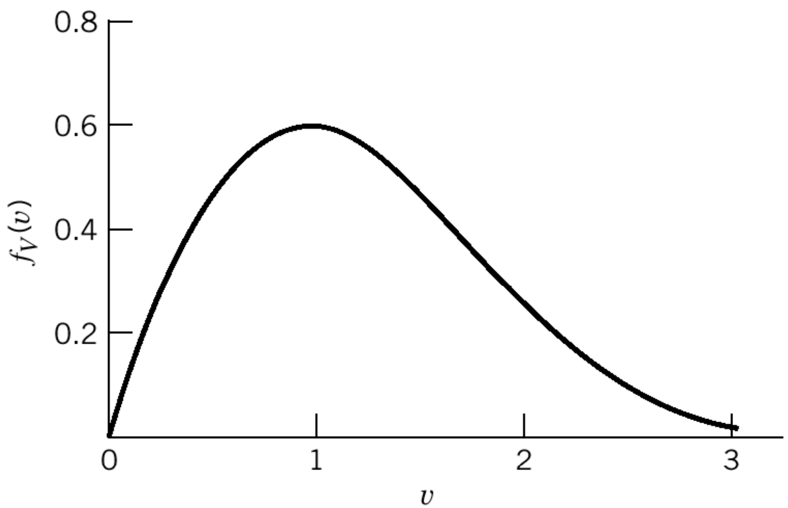
$$f(r) = \frac{r}{\sigma^2} e^{-(r^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar}{\sigma^2}\right), \quad r \geq 0$$

- The first case dominates the error probability when $A/\sigma \gg 1$.

Rayleigh Distribution

- Define a random variable $R = \sqrt{X^2 + Y^2}$ where X and Y are independent Gaussian with zero mean and variance σ^2
- R has Rayleigh distribution:

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}, \quad r \geq 0$$



Normalized Rayleigh distribution

$v = r/\sigma$

$f_V(v) = \sigma f_R(r)$

- Proving it requires change into polar coordinates:

$$R = \sqrt{X^2 + Y^2}, \quad \Theta = \tan^{-1}\frac{Y}{X}$$

Derivation

- Consider a small area $dxdy = rdrd\theta$. One has

$$f_{R,\Theta}(r,\theta)drd\theta = f_{X,Y}(x,y)dxdy = f_{X,Y}(x,y)rdrd\theta$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} rdrd\theta$$

- Hence

$$f_{R,\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

- pdf of R:

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)d\theta = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \qquad r > 0$$

- pdf of Θ

$$f_\Theta(\theta) = \int_0^{\infty} f_{R,\Theta}(r,\theta)dr = \frac{1}{2\pi}, \quad \theta \in [0, 2\pi]$$

Rician Distribution

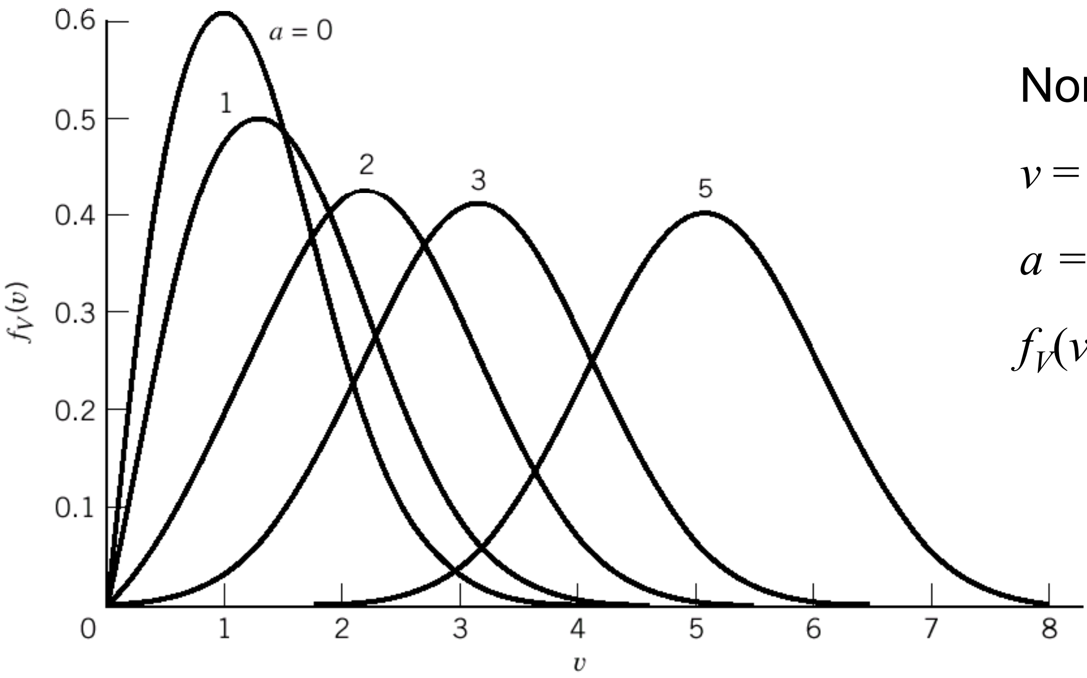
- If X has nonzero mean A, R has Rician distribution:

$$f_R(r) = \frac{r}{\sigma^2} e^{-(r^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar}{\sigma^2}\right), \qquad r \geq 0$$

where

$$I_0(x) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{x\cos\theta} d\theta$$

is the modified zero-order Bessel function of the first kind.



Normalized Rician distribution

$v = r/\sigma$

$a = A/\sigma$

$f_V(v) = \sigma f_R(r)$

Derivation

- Similarly,

$$f_{R,\Theta}(r,\theta)drd\theta = f_{X,Y}(x,y)rdrd\theta$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{(x-A)^2+y^2}{2\sigma^2}} rdrd\theta$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{r^2+A^2-2Ar\cos\theta}{2\sigma^2}} rdrd\theta$$

- Hence

$$f_{R,\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2+A^2-2Ar\cos\theta}{2\sigma^2}}$$

- pdf of R:

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)d\theta = \frac{r}{\sigma^2} e^{-\frac{r^2+A^2}{2\sigma^2}} \overbrace{\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{Ar\cos\theta}{\sigma^2}} d\theta}^{\text{Bessel function}}, \qquad r > 0$$

Error Probability

- Let the threshold be *A*/2 for simplicity.
- The error probability is dominated by symbol 0, and is given by

$$P_e \approx \frac{1}{2}\int_{A/2}^{\infty} \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)} dr$$

- The final expression

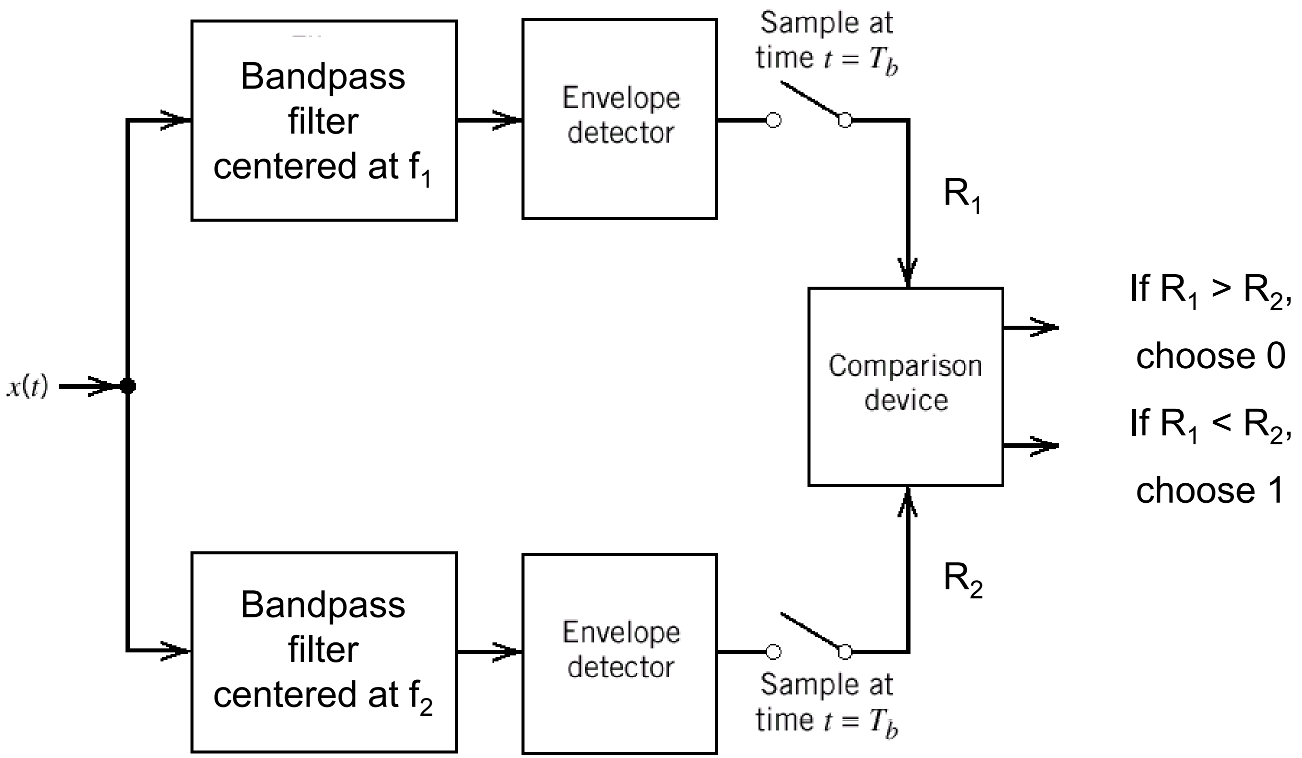
$$P_{e,ASK,noncoherent} \approx \frac{1}{2} e^{-A^2/(8\sigma^2)}$$

- Cf. coherent demodulation

$$P_{e,ASK,Coherent} = Q\left(\frac{A}{2\sigma}\right) \le \frac{1}{2} e^{-A^2/(8\sigma^2)}$$

- Noncoherent demodulation results in some performance degradation. Yet, for a large SNR, the performances of coherent and noncoherent demodulation are similar.

Noncoherent Demodulation of FSK



Distribution of Envelope

- When a symbol 1 is sent, outputs of the BPFs

$$\begin{aligned} y_1(t) &= n_1(t) \\ y_2(t) &= n_2(t) + A\cos(2\pi f_2 t) \end{aligned}$$

- Again, the first branch has Rayleigh distribution

$$f_{R_1}(r_1) = \frac{r_1}{\sigma^2} e^{-r_1^2/(2\sigma^2)}, \quad r_1 \geq 0$$

- while the second has Rice distribution

$$f_{R_2}(r_2) = \frac{r_2}{\sigma^2} e^{-(r_2^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar_2}{\sigma^2}\right), \qquad r_2 \geq 0$$

- Note the envelopes R_1 and R_2 are statistically independent.

Error Probability

- Error occurs if Rice < Rayleigh

$$P_e = P(R_2 < R_1)$$

$$= \int_0^\infty \int_{r_2}^\infty \frac{r_2}{\sigma^2} e^{-(r_2{}^2+A^2)/(2\sigma^2)} I_0(\frac{Ar_2}{\sigma^2}) \frac{r_1}{\sigma^2} e^{-r_1{}^2/(2\sigma^2)} dr_1 dr_2$$

$$= \int_0^\infty \frac{r_2}{\sigma^2} e^{-(r_2{}^2+A^2)/(2\sigma^2)} I_0(\frac{Ar_2}{\sigma^2}) \int_{r_2}^\infty \frac{r_1}{\sigma^2} e^{-r_1{}^2/(2\sigma^2)} dr_1 dr_2$$

$$= \int_0^\infty \frac{r_2}{\sigma^2} e^{-(2r_2{}^2+A^2)/(2\sigma^2)} I_0(\frac{Ar_2}{\sigma^2}) dr_2$$

$$= \frac{1}{2} e^{-A^2/(4\sigma^2)} \int_0^\infty \frac{x}{\sigma^2} e^{-(x^2+\alpha^2)/(2\sigma^2)} I_0(\frac{\alpha x}{\sigma^2}) dx \qquad x = \sqrt{2} r_2,\ \alpha = A/\sqrt{2}$$

- Observe the integrand is a Rician density

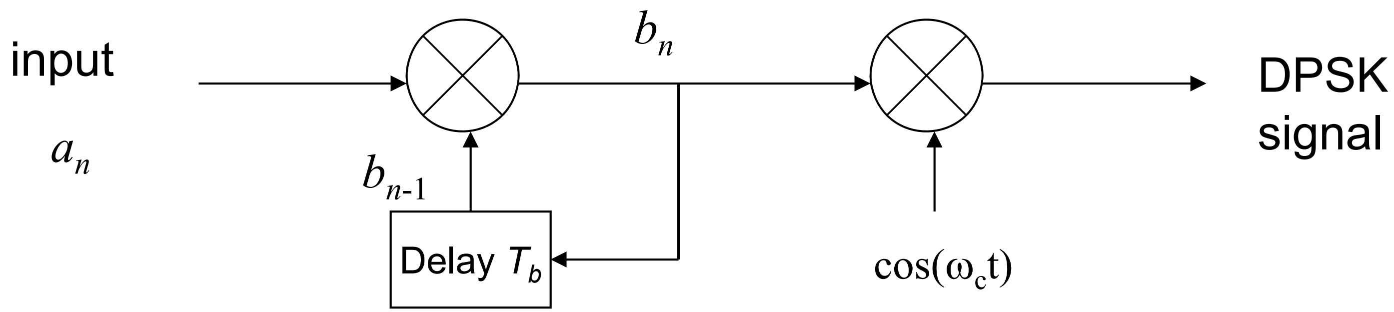
$$P_{e,FSK,noncoherent} = \frac{1}{2} e^{-A^2/(4\sigma^2)}$$

- Cf. coherent demodulation

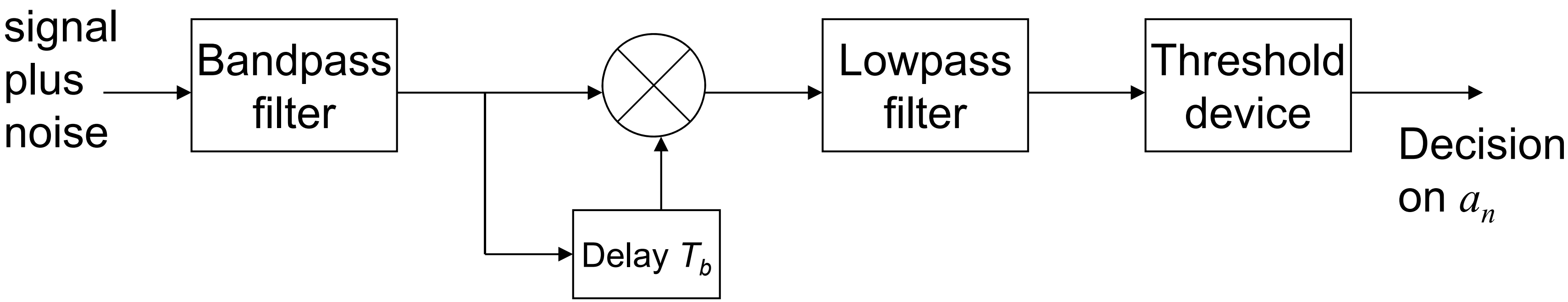
$$P_{e,FSK,Coherent} = Q\left(\frac{A}{\sqrt{2}\sigma}\right) \le \frac{1}{2} e^{-A^2/(4\sigma^2)}$$

DPSK: Differential PSK

- It is impossible to demodulate PSK with an envelop detector, since PSK signals have the same frequency and amplitude.
- We can demodulate PSK differentially, where phase reference is provided by a delayed version of the signal in the previous interval.
- Differential encoding is essential: $b_n = b_{n-1} \times a_n$, where a_n, $b_n \in \pm 1$.



Differential Demodulation



- Computing the error probability is cumbersome but fortunately the final expression is simple:

$$P_{e,DPSK} = \frac{1}{2} e^{-A^2/(2\sigma^2)}$$

  - The derivation can be found in Haykin, Communication Systems, 4th ed., Chap. 6.
  - Performance is degraded in comparison to coherent PSK.

- Cf. coherent demodulation

$$P_{e,PSK} = Q\left(\frac{A}{\sigma}\right) \le \frac{1}{2} e^{-A^2/(2\sigma^2)}$$

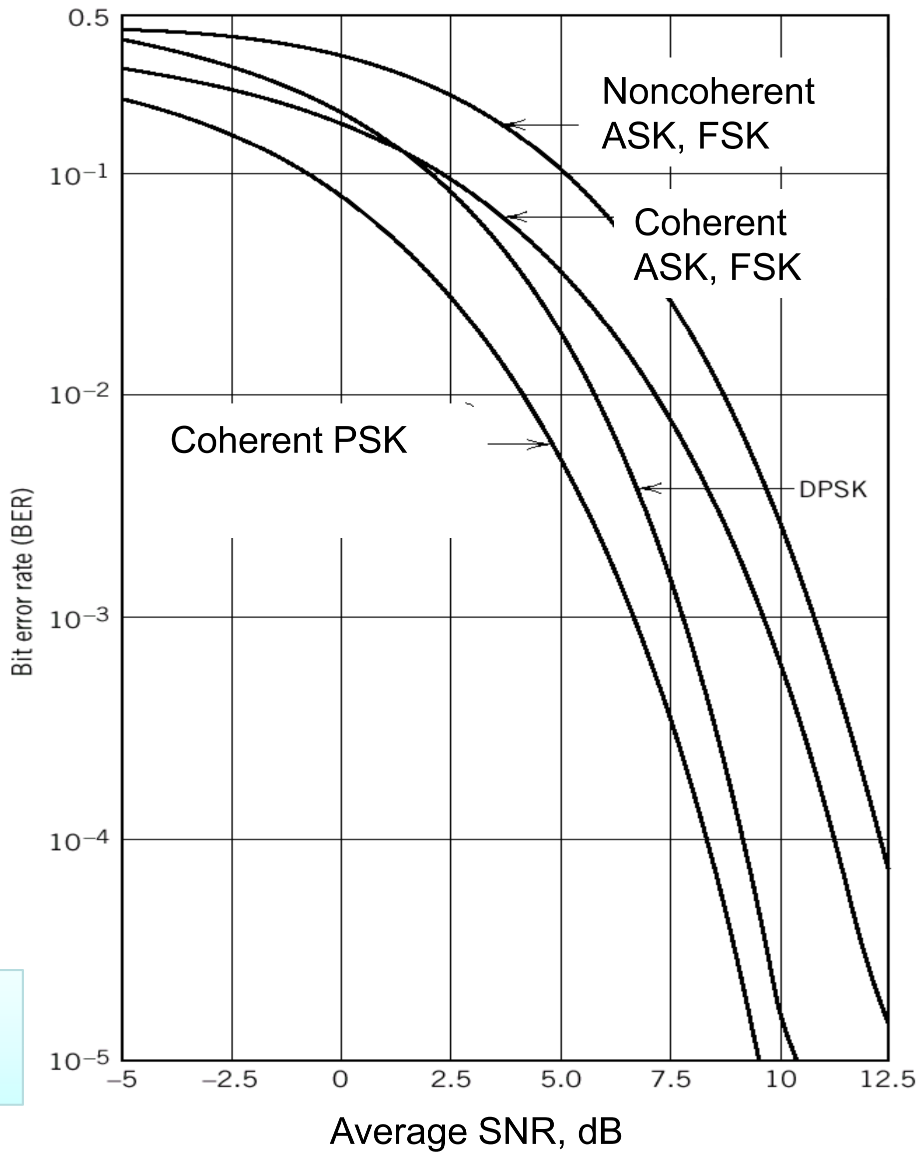
Illustration of DPSK

Information symbols $\{a_n\}$		1	-1	-1	1	-1	-1	1	1
$\{b_{n-1}\}$		1	1	-1	1	1	-1	1	1
Differentially encoded sequence $\{b_n\}$	1	1	-1	1	1	-1	1	1	1
Transmitted phase (radians)	0	0	π	0	0	π	0	0	0
Output of lowpass filter (polarity)		+	–	–	+	–	–	+	+
Decision		1	-1	-1	1	-1	-1	1	1

Note: The symbol 1 is inserted at the beginning of the differentially encoded sequence is the reference symbol.

Summary and Comparison

Scheme	Bit Error Rate
Coherent ASK	$Q(A/2\sigma)$
Coherent FSK	$Q(A/\sqrt{2}\sigma)$
Coherent PSK	$Q(A/\sigma)$
Noncoherent ASK	$\frac{1}{2}\exp(-A^2/8\sigma^2)$
Noncoherent FSK	$\frac{1}{2}\exp(-A^2/4\sigma^2)$
DPSK	$\frac{1}{2}\exp(-A^2/2\sigma^2)$



Caution: ASK and FSK have the same bit error rate if measured by average SNR.

Conclusions

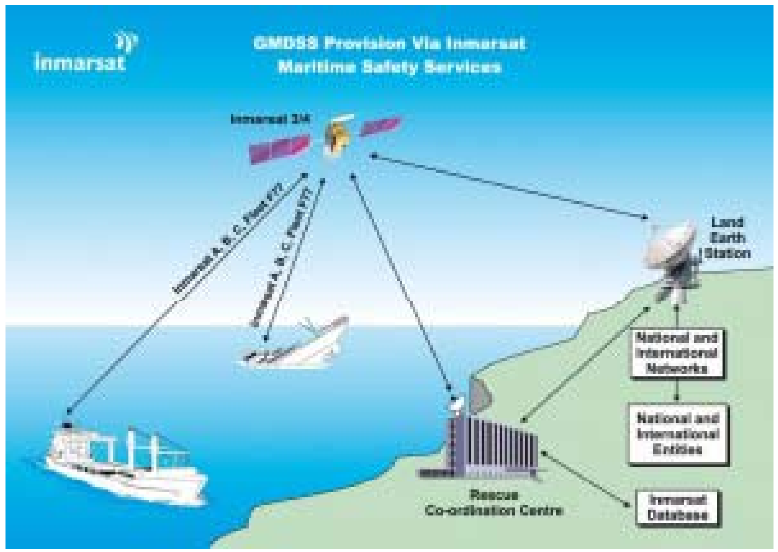
- Non-coherent demodulation retains the hierarchy of performance.

- Non-coherent demodulation has error performance slightly worse than coherent demodulation, but approaches coherent performance at high SNR.

- Non-coherent demodulators are considerably easier to build.

Application: DPSK

- WLAN standard IEEE 802.11b
- Bluetooth2
- Digital audio broadcast (DAB): DPSK + OFDM (orthogonal frequency division multiplexing)
- Inmarsat (International Maritime Satellite Organization): now a London-based mobile satellite company



Imperial College
London

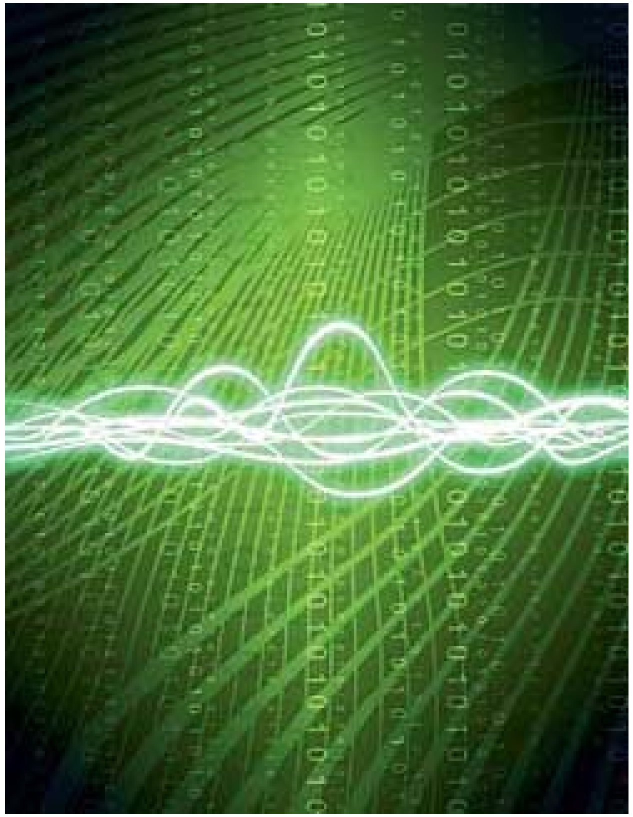
EE2-4: Communication Systems

Lecture 12: Entropy and Source Coding

Dr. Cong Ling

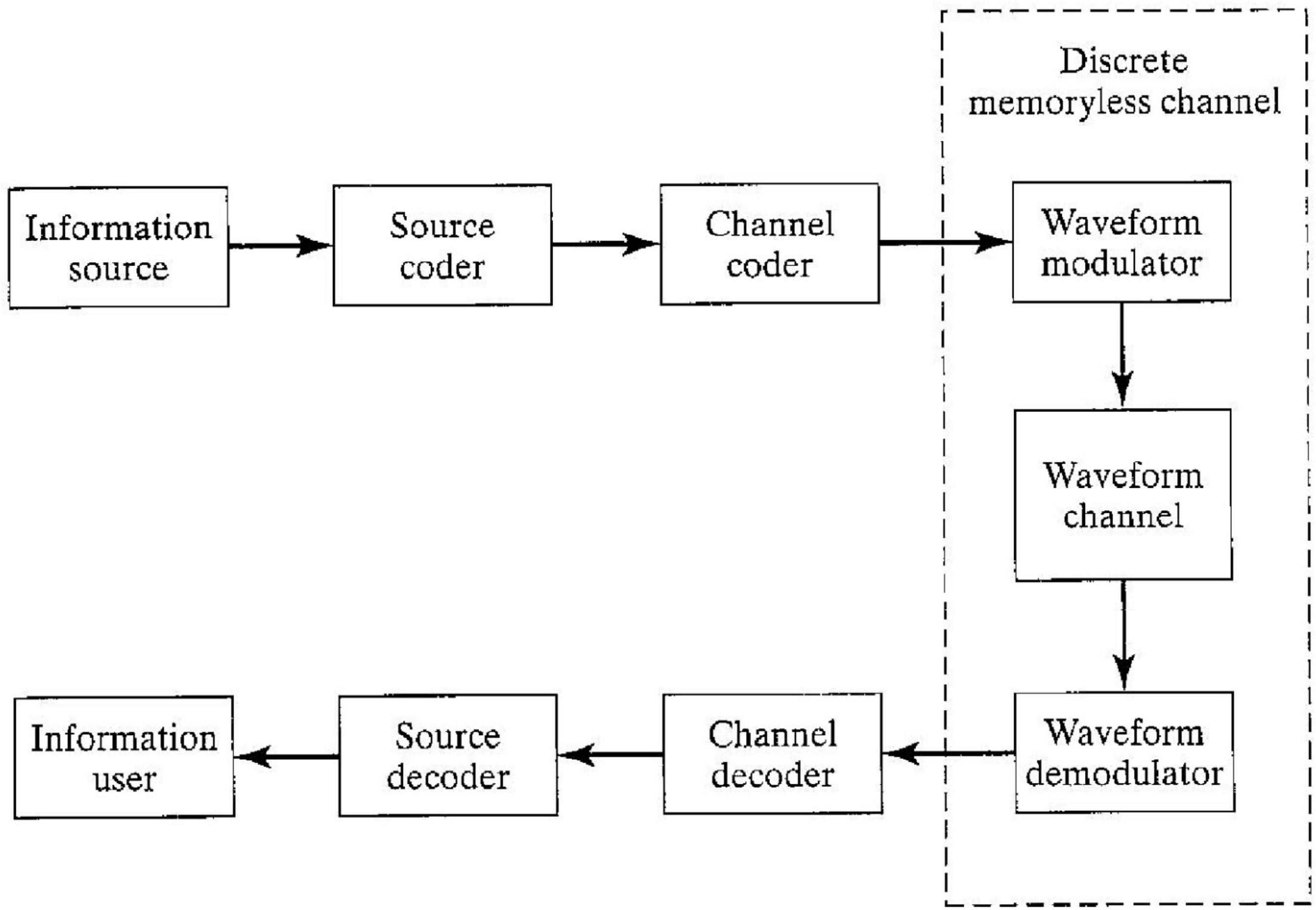
Department of Electrical and Electronic Engineering

Outline



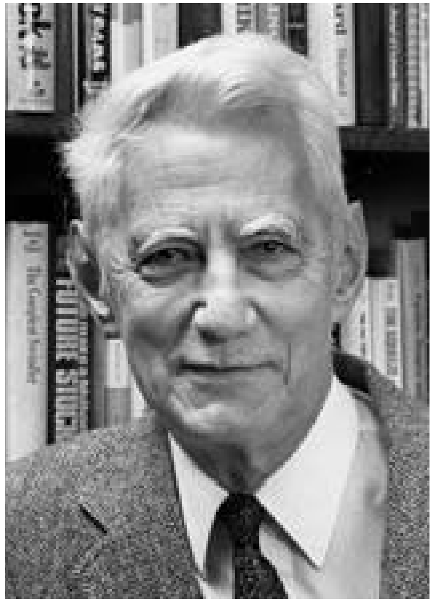
- What is information theory?
- Entropy
  - Definition of information
  - Entropy of a source
- Source coding
  - Source coding theorem
  - Huffman coding
- References
  - Notes of Communication Systems, Chap. 5.1-5.4
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 10
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 15

Model of a Digital Communication System



What is Information Theory

- C. E. Shannon, ”A mathematical theory of communication,” Bell System Technical Journal, 1948.
- Two fundamental questions in communication theory:
  – What is the ultimate limit on data compression?
  – What is the ultimate transmission rate of reliable communication over noisy channels?
- Shannon showed reliable (i.e., error-free) communication is possible for all rates below channel capacity (using channel coding).
- Any source can be represented in bits at any rate above entropy (using source coding).
  – Rise of digital information technology



What is Information?

- Information: any new knowledge about something
- How can we measure it?
- Messages containing knowledge of a **high** probability of occurrence ⇒ Not very informative
- Messages containing knowledge of **low** probability of occurrence ⇒ More informative
- A small change in the probability of a certain output should not change the information delivered by that output by a large amount.

Definition

- The amount of information of a symbol *s* with probability p:

$$I(s) = \log \frac{1}{p}$$

- Properties:
  - $p = 1 \Rightarrow I(s) = 0$: a symbol that is certain to occur contains no information.
  - $0 \leq p \leq 1 \Rightarrow 0 \leq I(s) \leq \infty$: the information measure is monotonic and non-negative.
  - $p = p_1 \times p_2 \Rightarrow I(s) = I(p_1) + I(p_2)$: information is additive for statistically independent events.

Example

- In communications, log base 2 is commonly used, resulting in unit **bit**
- **Example:** Two symbols with equal probability $p_1 = p_2 = 0.5$
- Each symbol represents $I(s) = \log_2(1/0.5) = 1$ bit of information.
- In summary: $I(s) = \log_2 \frac{1}{p}$ bits
- Reminder: From logs of base 10 to logs of base 2:

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2} = 3.32 \log_{10} x$$

Discrete Memoryless Source

- Suppose we have an information source emitting a sequence of symbols from a finite alphabet:

$$S = \{s_1, s_2, \ldots, s_K\}$$

- **Discrete memoryless source**: The successive symbols are statistically independent (i.i.d.)

- Assume that each symbol has a probability of occurrence

$$p_k,\ k = 1,\ldots,K,\ \text{such that} \sum_{k=1}^{K} p_k = 1$$

Source Entropy

- If symbol s_k has occurred, then, by definition, we have received

$$I(s_k) = \log_2 \frac{1}{p_k} = -\log_2 p_k$$

bits of information.

- The expected value of $I(s_k)$ over the source alphabet

$$E\{I(s_k)\} = \sum_{k=1}^{K} p_k I(s_k) = -\sum_{k=1}^{K} p_k \log_2 p_k$$

- **Source entropy:** the average amount of information per source symbol:

$$H(S) = -\sum_{k=1}^{K} p_k \log_2 p_k$$

- Units: bits / symbol.

Meaning of Entropy

- What information about a source does its entropy give us?
- It is the amount of uncertainty before we receive it.
- It tells us how many bits of information per symbol we expect to get on the average.

- Relation with thermodynamic entropy
  - In **thermodynamics**: entropy measures disorder and randomness;
  - In **information theory**: entropy measures uncertainty.
  - Some argue that to obtain 1 bit information, the minimum energy needed is 10^{-23} Joules/Kelvin degree (extremely cheap!).
  - This may have implications on green computation.

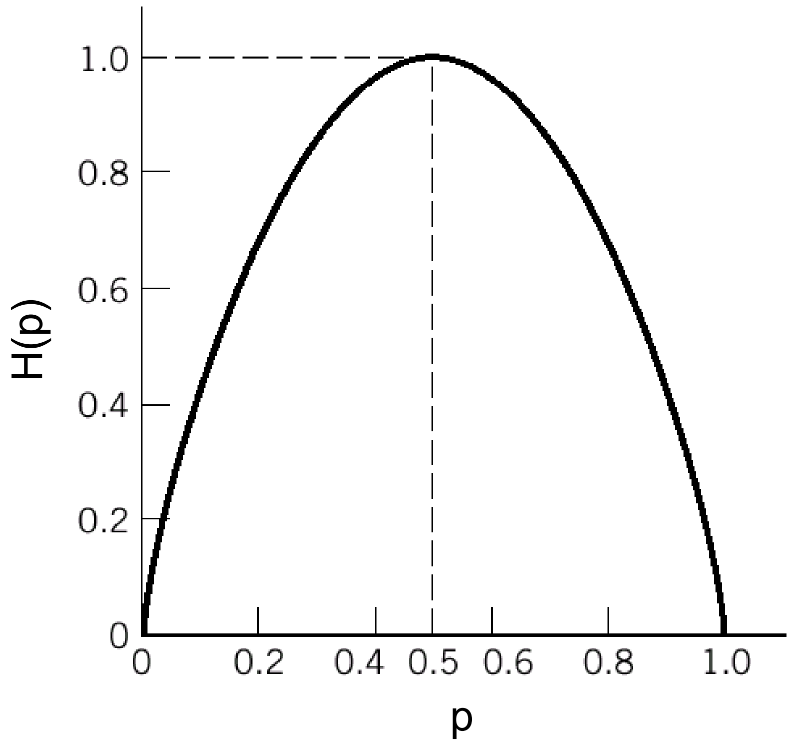
Example: Entropy of a Binary Source

- s_1: occurs with probability p
- s_0: occurs with probability $1 - p$.
- Entropy of the source:

$$H(S) = -(1-p)\log_2(1-p) - p\log_2 p = H(p)$$

  – $H(p)$ is referred to as the entropy function.

Maximum uncertainty when $p = 1/2$.



Example: A Three-Symbol Alphabet

- *A*: occurs with probability 0.7
- *B*: occurs with probability 0.2
- *C*: occurs with probability 0.1

• Source entropy:

$$H(S) = -0.7\log_2(0.7) - 0.2\log_2(0.2) - 0.1\log_2(0.1)$$
$$= 0.7 \times 0.515 + 0.2 \times 2.322 + 0.1 \times 3.322$$
$$= 1.157 \text{ bits/symbol}$$

• How can we encode these symbols in order to transmit them?

• We need 2 bits/symbol if encoded as

_A_ = 00, _B_ = 01, _C_ = 10

• Entropy prediction: the average amount of information is only 1.157 bits per symbol.

• We are wasting bits!

Source Coding Theory

- What is the minimum number of bits that are required to transmit a particular symbol?

- How can we encode symbols so that we achieve (or at least come arbitrarily close to) this limit?

- **Source encoding:** concerned with minimizing the actual number of source bits that are transmitted to the user

- **Channel encoding:** concerned with introducing redundant bits to enable the receiver to detect and possibly correct errors that are introduced by the channel.

Average Codeword Length

- l_k: the number of bits used to code the *k*-th symbol
- K: total number of symbols
- p_k: the probability of occurrence of symbol *k*
- Define the **average codeword length**:

$$\overline{L} = \sum_{k=1}^{K} p_k l_k$$

- It represents the average number of bits per symbol in the source alphabet.
- An idea to reduce average codeword length: symbols that occur often should be encoded with short codewords; symbols that occur rarely may be encoded using the long codewords.

Minimum Codeword Length

- What is the minimum codeword length for a particular alphabet of source symbols?
- In a system with 2 symbols that are equally likely:
  - Probability of each symbol to occur: $p = 1/2$
  - Best one can do: encode each with 1 bit only: 0 or 1

- In a system with *n* symbols that are equally likely: Probability of each symbol to occur: $p = 1/n$
- One needs $L = \log_2 n = \log_2 1/p = -\log_2 p$ bits to represent the symbols.

The General Case

- $S = \{s_1, \ldots, s_K\}$: an alphabet
- p_k: the probability of symbol s_k to occur
- N : the number of symbols generated
- We expect Np_k occurrences of s_k
- Assume:
  - The source is memoryless
  - all symbols are independent
  - the probability of occurrence of a **typical** sequence S_N is:

$$p(S_N) = p_1^{Np_1} \times p_2^{Np_2} \times \ldots \times p_K^{Np_K}$$

- All such typical sequences of N symbols are equally likely
- All other compositions are extremely unlikely to occur as $N \to \infty$ (Shannon, 1948).

Source Coding Theorem

- The number of bits required to represent a typical sequence S_N is

$$L_N = \log_2 \frac{1}{p(S_N)} = -\log_2 (p_1^{Np_1} \times p_2^{Np_2} \times \ldots \times p_K^{Np_K})$$
$$= -Np_1 \log_2 p_1 - Np_2 \log_2 p_2 - \ldots - Np_K \log_2 p_K$$
$$= -N \sum_{k=1}^{K} p_k \log_2 p_k = NH(S)$$

- Average length for one symbol: $\overline{L} = \frac{L_N}{N} = H(S)$ bits / symbol

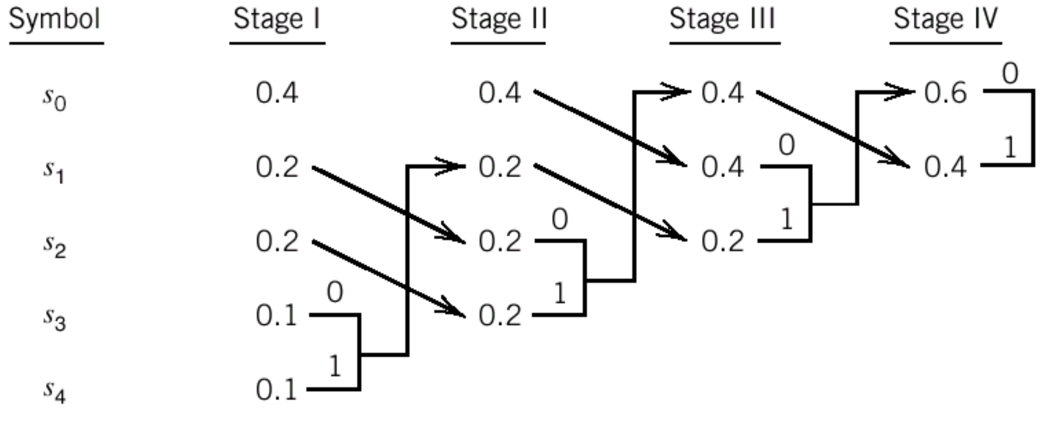
- **Given a discrete memoryless source of entropy $H(S)$, the average codeword length for any source coding scheme is bounded by $H(S)$:**

$$\overline{L} \geq H(S)$$

Huffman Coding

- How one can design an efficient source coding algorithm?
- Use **Huffman Coding** (among other algorithms)
- It yields the shortest average codeword length.
- Basic idea: **choose codeword lengths so that more-probable sequences have shorter codewords**
- Huffman Code construction
  - Sort the source symbols in order of decreasing probability.
  - Take the two smallest $p(x_i)$ and assign each a different bit (i.e., 0 or 1). Then merge into a single symbol.
  - Repeat until only one symbol remains.
- It's very easy to implement this algorithm in a microprocessor or computer.
- Used in JPEG, MP3…

Example



(a)

Symbol	Probability	Code word
s_0	0.4	00
s_1	0.2	10
s_2	0.2	11
s_3	0.1	010
s_4	0.1	011

(b)

Read diagram backwards for codewords

- The average codeword length:

$$\bar{L} = (2\times 0.4) + (2\times 0.2) + (2\times 0.2) + (3\times 0.1) + (3\times 0.1) = 2.2$$

- More than the entropy $H(S)$ = 2.12 bits per symbol.
  ⇒ Room for further improvement.

Application: File Compression

- A drawback of Huffman coding is that it requires knowledge of a probabilistic model, which is not always available a prior.
- Lempel-Ziv coding overcomes this practical limitation and has become the standard algorithm for file compression.
  - compress, gzip, GIF, TIFF, PDF, modem…
  - A text file can typically be compressed to half of its original size.

Summary

- The entropy of a discrete memoryless information source

$$H(S) = -\sum_{k=1}^{K} p_k \log_2 p_k$$

- Entropy function (entropy of a binary memoryless source)

$$H(S) = -(1-p)\log_2(1-p) - p\log_2 p = H(p)$$

- Source coding theorem: The minimum average codeword length for any source coding scheme is $H(S)$ for a discrete memoryless source.

- Huffman coding: An efficient source coding algorithm.

Imperial College
London

EE2-4: Communication Systems

Lecture 13: Channel Capacity

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Channel capacity
  - Channel coding theorem
  - Shannon formula

- Comparison with practical systems

- References
  - Notes of Communication Systems, Chap. 5.5-5.6
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 10
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 15

Channel Coding Theorem

- Shannon, “A mathematical theory of communication,” 1948
- For any channel, there exists an information capacity C (whose calculation is beyond the scope of this course).
- **If the transmission rate $R \leq C$, then there exists a coding scheme such that the output of the source can be transmitted over a noisy channel with an arbitrarily small probability of error. Conversely, it is not possible to transmit messages without error if $R > C$.**
- Important implication:
  - The basic limitation due to noise in a communication channel is not on the *reliability* of communication, but rather, on the *speed* of communication.

Shannon Capacity Formula

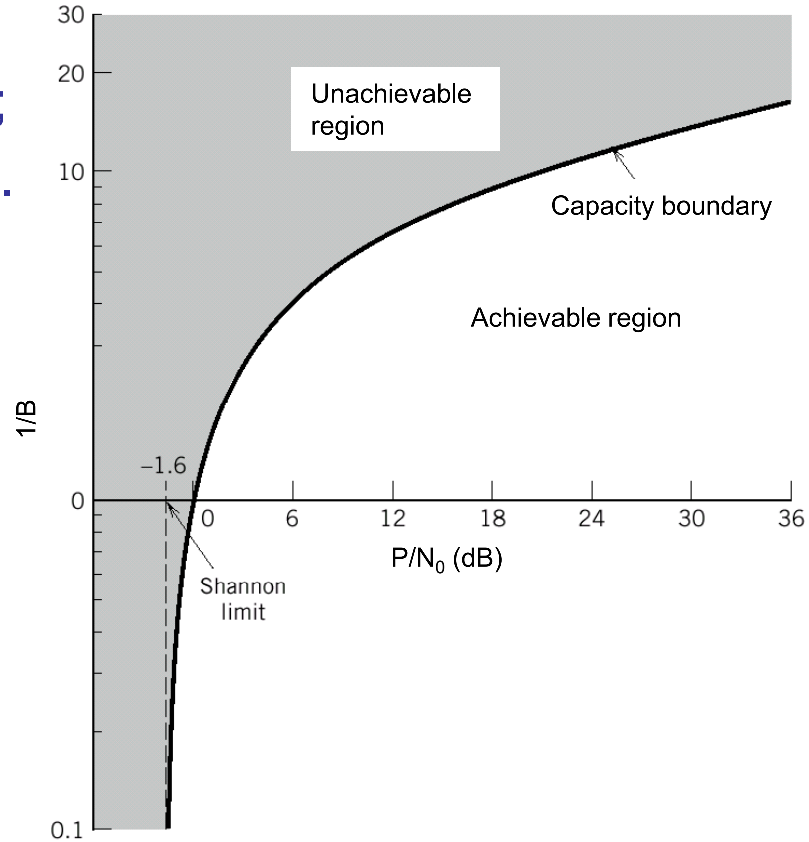
- For an **additive white Gaussian noise (AWGN) channel**, the channel capacity is

$$C = B\log_2(1+\mathrm{SNR}) = B\log_2\left(1+\frac{P}{N_0 B}\right) \quad \text{bps}$$

  - *B*: the bandwidth of the channel
  - *P*: the average signal power at the receiver
  - N_0: the single-sided PSD of noise
- Important implication: **We can communicate error free up to *C* bits per second.**
- How can we achieve this rate?
  - Design power error correcting codes to correct as many errors as possible (the next two lectures).
  - Use the ideal modulation scheme that does not lose information in the detection process.

Shannon Limit

- Tradeoff between power and bandwidth: to get the same target capacity C, one may
  - Increase power, decrease bandwidth;
  - Decrease power, increase bandwidth.
- What's the minimum power to send 1 bps (i.e., C = 1 bps)?



$$\frac{P}{N_0} = B(2^{1/B} - 1) = \frac{2^{1/B} - 1}{1/B}$$

$$\Rightarrow \lim_{B\to\infty} \frac{P}{N_0} = \lim_{B\to\infty} \frac{2^{1/B} \ln 2(-B^{-2})}{-B^{-2}} = \ln 2$$

$$= 0.693 = -1.6 \text{ dB}$$

- This is the ultimate limit of **green communications**.
  - For more information, see http://www.greentouch.org

Example

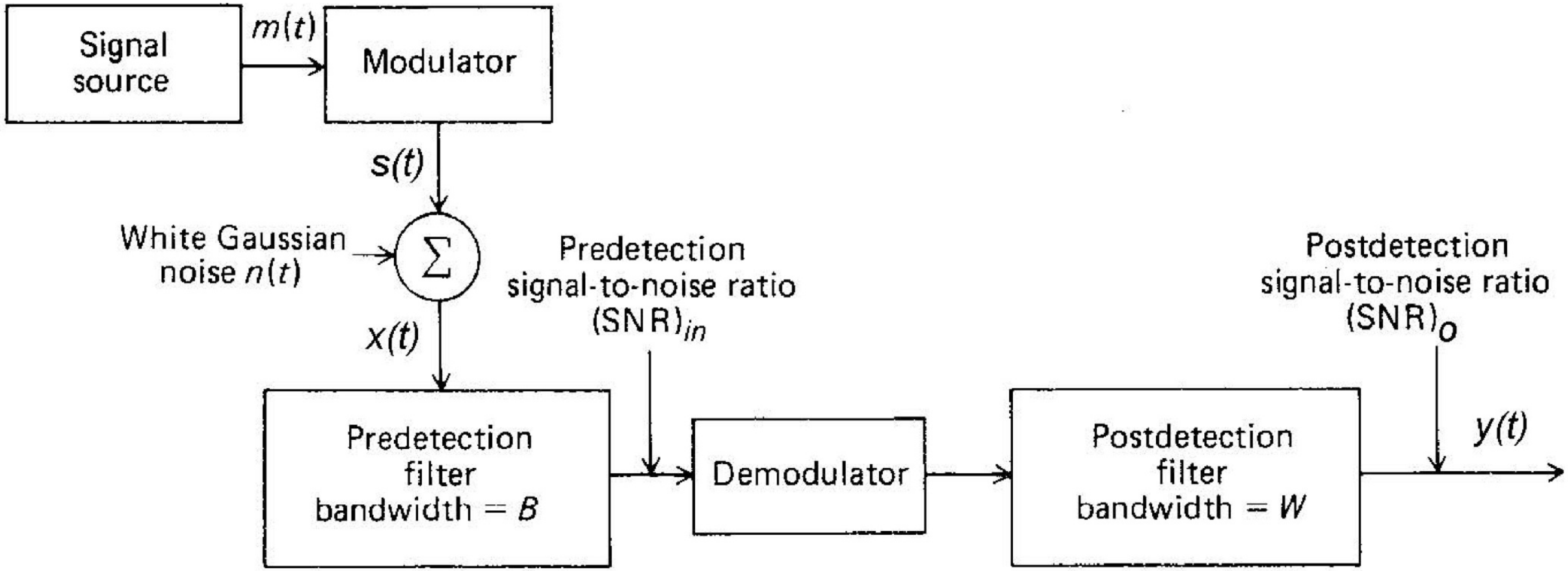
- A voice-grade channel of the telephone network has a bandwidth of 3.4 kHz.
  - (a) Calculate the capacity for a SNR of 30 dB.
  - (b) Calculate the SNR required to support a rate of 4800 bps.
- Answer:
- (a) 30 dB $\Rightarrow$ SNR = 1000

$$\begin{aligned} C &= B\log_2(1+SNR) \\ &= 3.4\times\log_2(1+1000) \\ &= 33.9 \text{ kbps} \end{aligned}$$

- (b)

$$\begin{aligned} SNR &= 2^{C/B}-1 \\ &= 2^{4.8/3.4}-1 \\ &= 1.66 = 2.2 \text{ dB} \end{aligned}$$

General Model of a Modulated Communication System



Input/Output SNR

- The maximum rate at which information may arrive at the receiver is

$$C_{in} = B\log_2(1 + SNR_{in})$$

  - SNR_{in}: the predetection signal-to-noise ratio at the input to the demodulator.
- The maximum rate at which information can leave the receiver is

$$C_o = W\log_2(1 + SNR_o)$$

  - SNR_o: the SNR at the output of the postdetection filter.
- For the ideal modulation scheme:

$$C_{in} = C_o \Rightarrow 1 + SNR_o = [1 + SNR_{in}]^{B/W}$$

- For high SNR,

$$SNR_o \approx SNR_{in}{}^{B/W}$$

- It seems that spreading the bandwidth would make the output SNR increase exponentially (not true).

A More Elaborate Analysis

- If the channel noise has a double-sided white PSD of $N_0/2$ $\Rightarrow$ the average noise power at the demodulator will be N_0B.
- If the signal power is P:

$$SNR_{in} = \frac{P}{N_0B} = \frac{W}{B}\frac{P}{N_0W}$$

$$\frac{P}{N_0W} = SNR_{baseband}$$: the baseband SNR

- Increasing B will reduce SNR_{in}, and thus

$$W\log_2(1+SNR_o) = B\log_2\left(1+\frac{SNR_{baseband}}{B/W}\right)$$

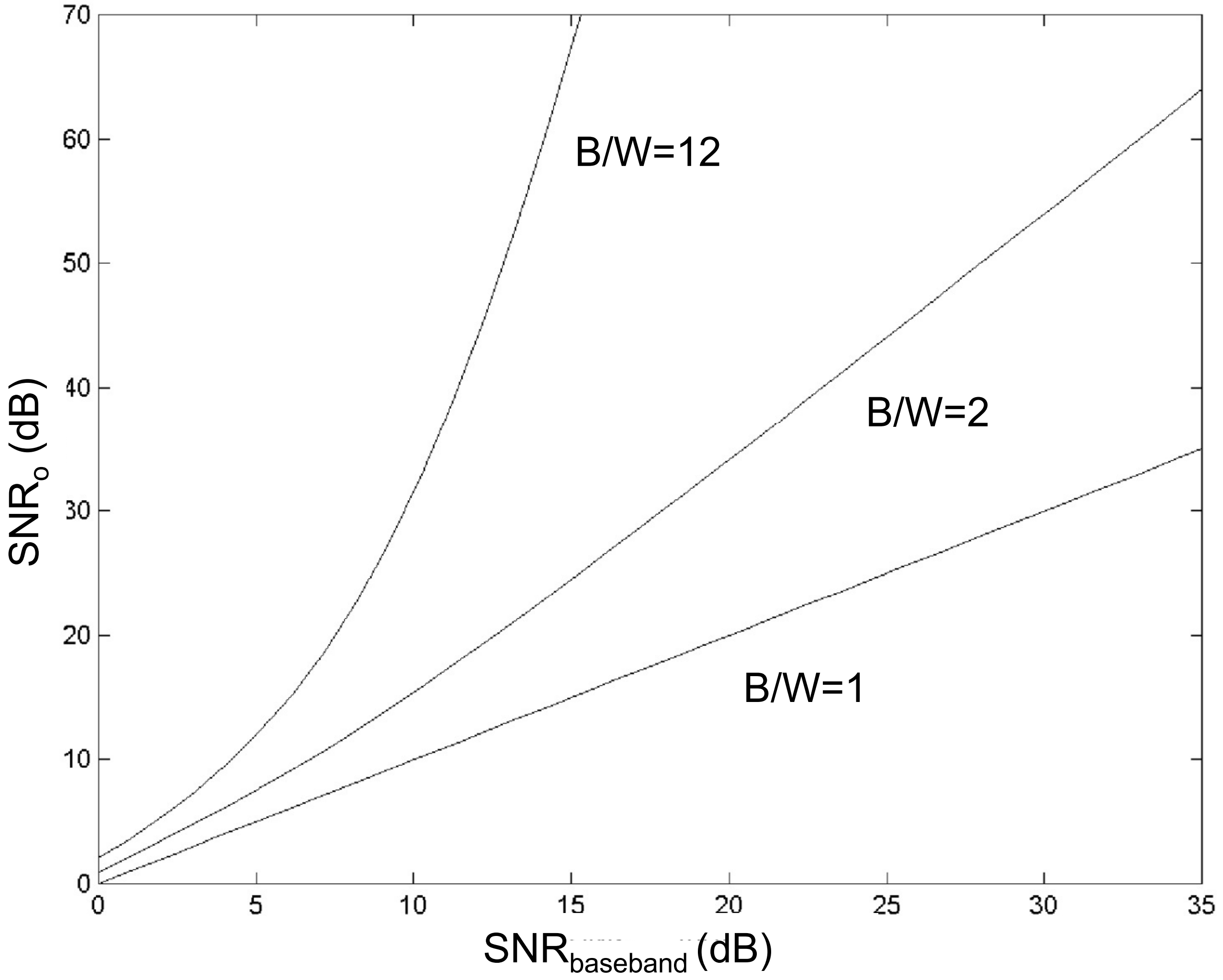
- Output SNR of an **ideal** communication system:

$$SNR_o = \left(1+\frac{SNR_{baseband}}{B/W}\right)^{B/W} - 1$$

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e$$

$$\to e^{SNR_{baseband}} \quad \text{as} \quad B/W \to \infty$$

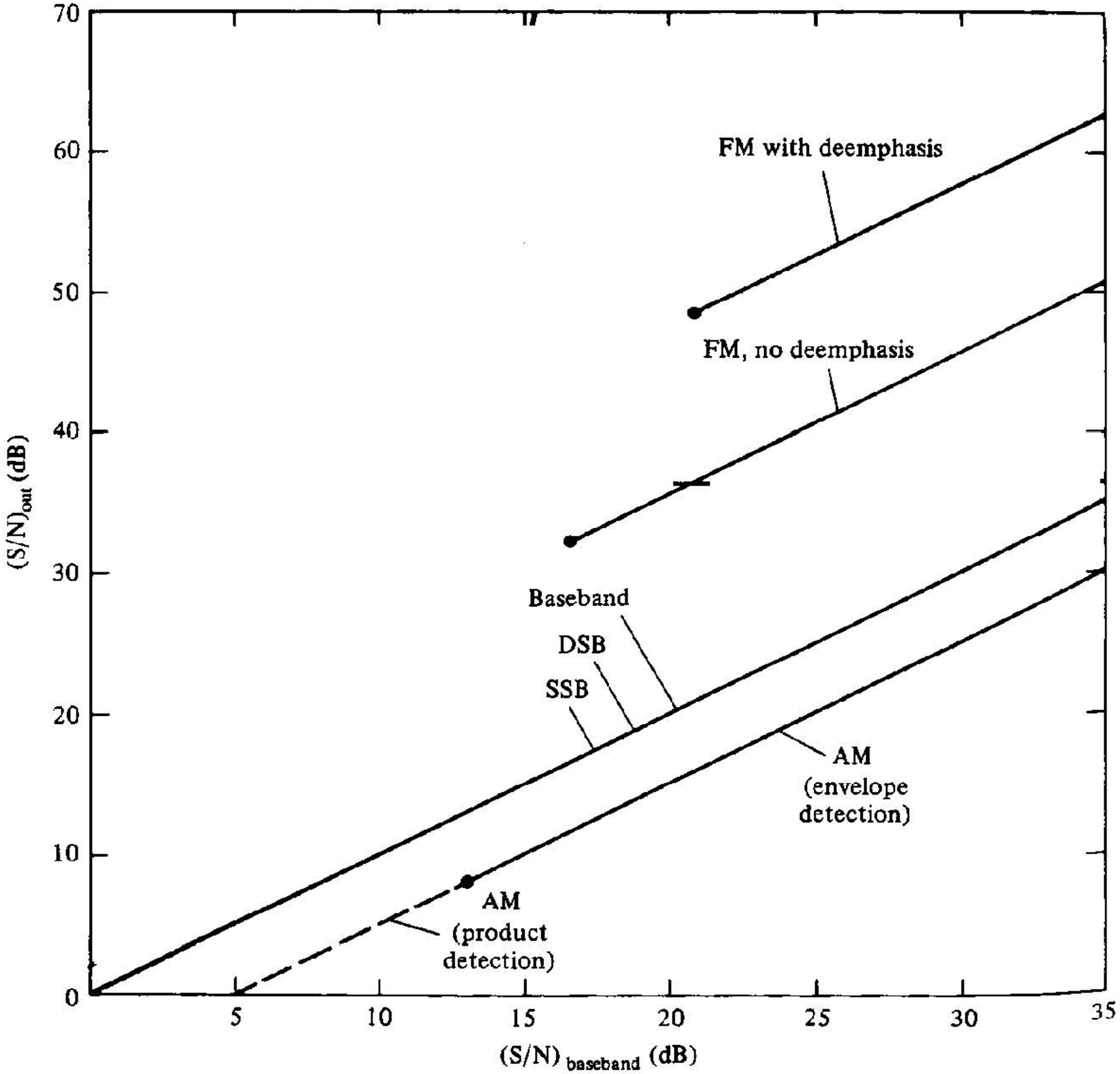
SNR vs. Bandwidth



Implications to Analog Systems

- A bandwidth spreading ratio of *B/W*=1 corresponds to both SSB and baseband.
- A bandwidth spreading ratio of *B/W*=2 corresponds to DSB and AM (the ideal system would have $SNR_o = SNR_{base}^2/4$).
- A bandwidth spreading ratio of *B/W*=12 corresponds to commercial FM broadcasting and AM (the ideal system would have $SNR_o \approx e^{SNR_{base}}$).
- SSB and baseband systems provide noise performance identical to the ideal (which is trivial).
- DSB has a worse performance because of its additional bandwidth requirements.
- FM systems only come close to the ideal system near threshold.

Noise Performance of Analog Communication Systems



Discussion

- Known analogue communication systems do not achieve the ideal performance in general.
- Something must be missing with analogue communications.
- Similarly, it can be shown that simple digital modulation schemes cannot achieve capacity too.
- One must resort to digital communications and **coding** to approach the capacity.
- The idea of tradeoff between bandwidth and SNR is useful in **spread-spectrum** communications (CDMA, 3G mobile communications) and **ultra wideband**.

Imperial College London

EE2-4: Communication Systems

Lecture 14: Block Codes

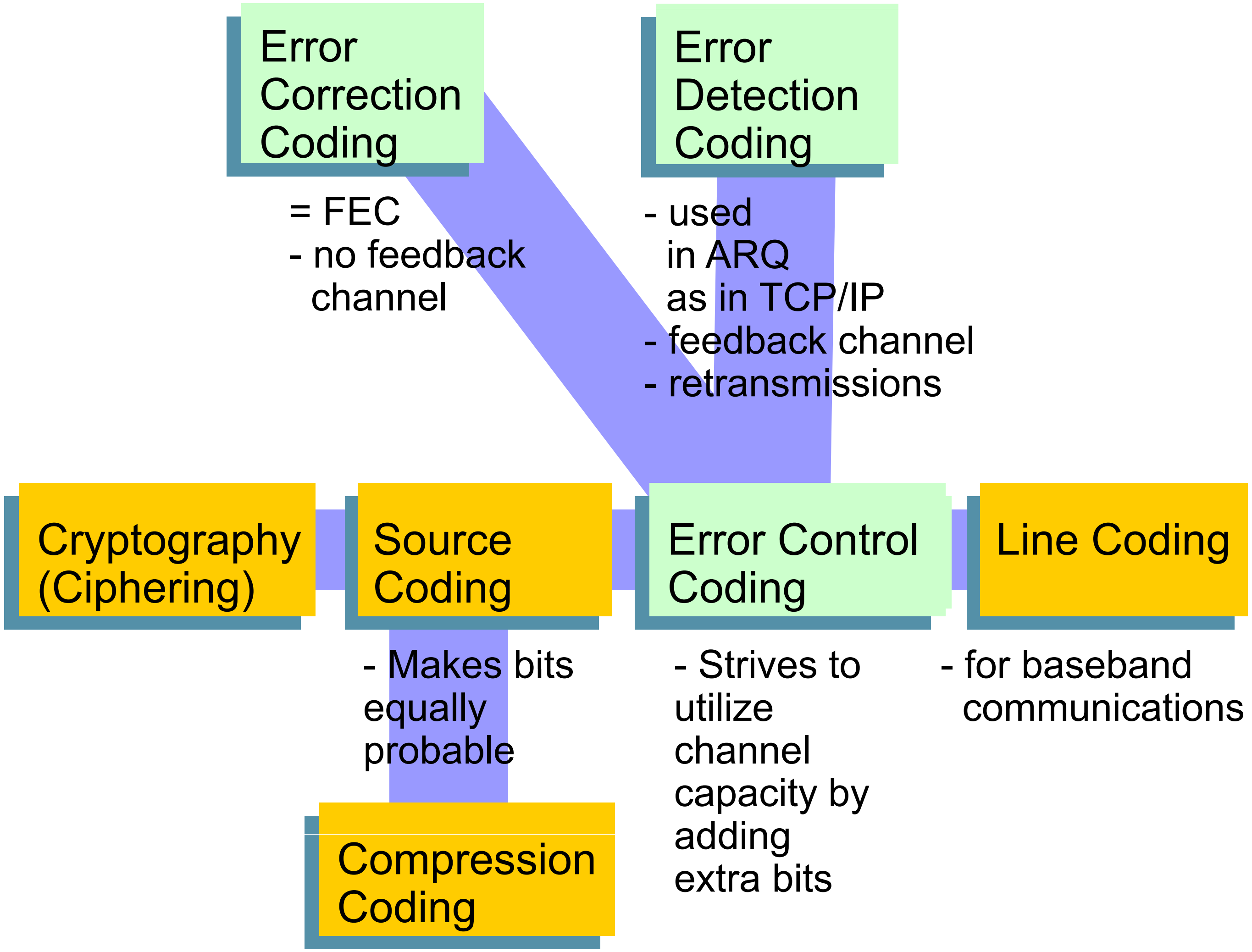
Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Introduction
- Block codes
- Error detection and correction
- Generator matrix and parity-check matrix
- References
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 10
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 16

Taxonomy of Coding



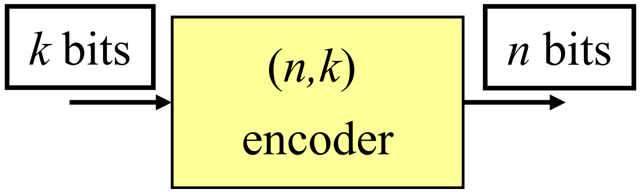
- Redundancy removal:
  - Destructive (jpeg, mpeg)
  - Non-destructive (zip)

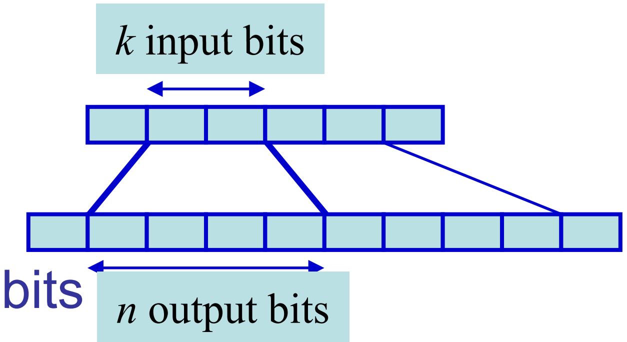
FEC: Forward Error Correction

ARQ: Automatic Repeat Request

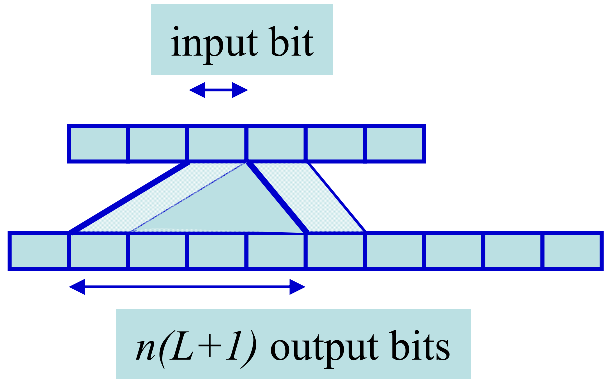
Block vs. Convolutional Codes

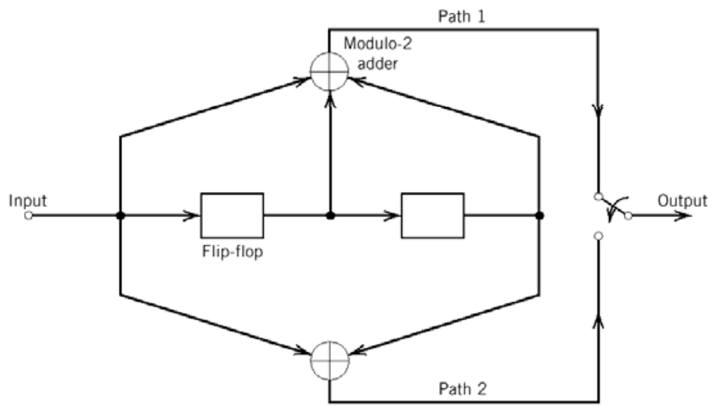
- Block codes





  - Output of n bits depends only on the k input bits
- Convolutional codes
  - Each source bit influences $n(L+1)$ output bits
  - L is the memory length
  - Like convolution in a linear filter





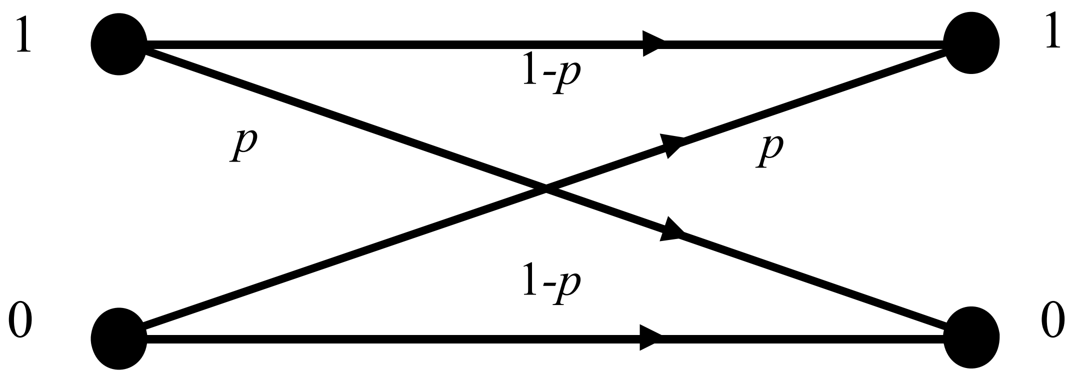
$k = 1,\ n = 2,\ L = 2$

Noise and Errors

- Noise can corrupt the information that we wish to transmit.
- Generally corruption of a signal is a bad thing that should be avoided, if possible.
- Different systems will generally require different levels of protection against errors due to noise.
- Consequently, a number of different techniques have been developed to detect and correct different types and numbers of errors.

Channel Model

- We can measure the effect of noise in different ways. The most common is to specify an error probability, *p*. Consider the case of a **Binary Symmetric Channel** with noise.



- All examples considered here will be for error probabilities that are symmetric, stationary and statistically independent.

Simple Error Checks

- If the error probability is small and the information is fairly fault tolerant, it is possible to use simple methods to detect errors.
- **Repetition** – Repeating each bit in the message
  - If the two symbols in an adjacent pair are different, it is likely that an error has occurred.
  - However, this is not very efficient (efficiency is halved).
  - Repetition provides a means for error checking, but not for error correction.
- **Parity bit** – Use of a 'parity bit' at the end of the message
  - A parity bit is a single bit that corresponds to the sum of the other message bits (modulo 2).
  - This allows any odd number of errors to be detected, but not even numbers.
  - As with repetition, this technique only allows error checking, not error correction.
  - It is more efficient than simple repetition.

Block Codes

- An important class of codes that can detect and correct some errors are block codes
- The first error-correcting block code was devised by Hamming around the same time as Shannon was working on the foundation of information theory
  - Hamming codes are a particular class of linear block code
- Block codes
  - Encode a series of symbols from the source, a ‘block’, into a longer string: codeword or code block
  - Errors can be detected as the received coded block will not be one of the recognized, valid coded blocks
  - Error correction: To “decode” and associate a corrupted block to a valid coded block by its proximity (as measured by the “Hamming distance”)

Binary Fields and Vectors

- We need to discuss some of mathematics that will be needed.
- Fortunately, we have restricted things to binary sources, so the mathematics is relatively simple.
- The binary alphabet **A={0,1}** is properly referred to as a Galois field with two elements, denoted **GF(2)**.
  - Addition (XOR)

    $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$, $1 + 1 = 0$
  - Multiplication (AND)

    $0 \cdot 1 = 0 \cdot 0 = 1 \cdot 0 = 0$, $1 \cdot 1 = 1$
  - This is also referred to as Boolean arithmetic in Digital Electronics or modulo-2 arithmetic.
- A message is built up from a number of binary fields, and forms a binary vector, rather than a larger binary number.
  - Hence, $101 \neq 5$ $101=\{1\}\{0\}\{1\}$

Example

- Calculate the following examples of binary field arithmetic:

i. 01001 + 01110

ii. 10010 • 01110

iii. (1111+0011) • 0011

- Answers

i. 00111

ii. 00010

iii. 1100 • 0011 = 0000

Hamming Distance

- **Hamming Weight**
  - The Hamming weight of a binary vector, a (written as $w_H(a)$), is the number of non-zero elements that it contains.
  - Hence,

    001110011 has a Hamming weight of 5.

    000000000 has a Hamming weight of 0.
- **Hamming Distance**
  - The Hamming Distance between two binary vectors, a and b, is written $d_H(a,b)$, and is equal to the Hamming weight of their (Boolean) sum.

$$d_H(a,b) = w_H(a+b)$$

  - Hence, 01110011 and 10001011 have a Hamming distance of

$$\begin{aligned} d_H &= w_H\,(01110011+10001011) \\ &= w_H\,(11111000) = 5 \end{aligned}$$

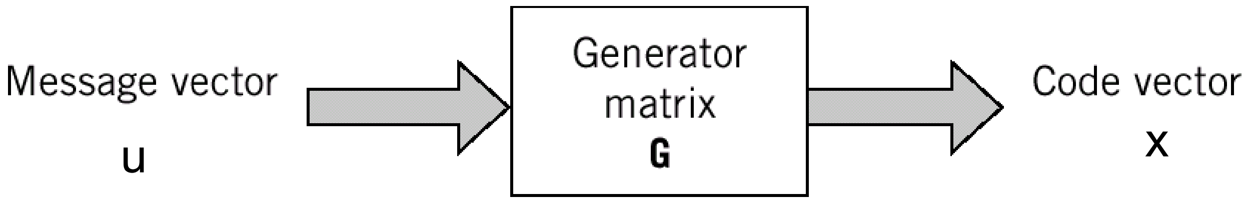
Linear Block Codes

- A binary linear block code that takes block of *k* bits of source data and encodes them using *n* bits, is referred to as a (*n*, *k*) binary linear block code.
  - The ratio between the number of source bits and the number of bits used in the code, $R=k/n$, is referred to as the **code rate**.
- The most important feature of a linear block code
  - **Linearity**: the Boolean sum of any codewords must be another codeword.
  - This means that the set of code words forms a vector space, within which mathematical operations can be defined and performed.

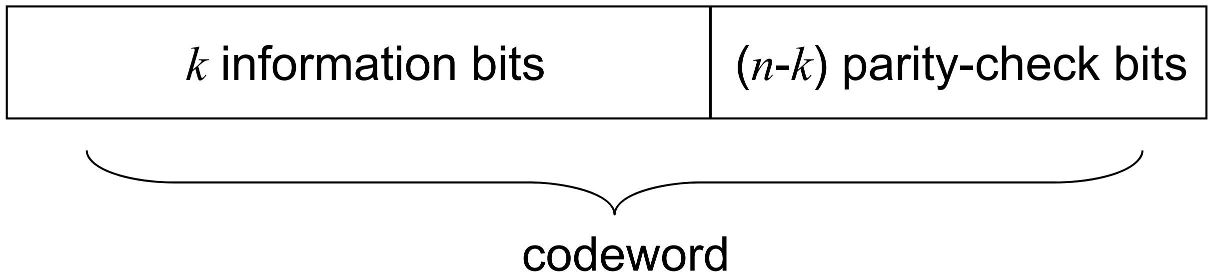
Generator Matrix

- To construct a linear block code we define a matrix, the **generator matrix** G, that converts blocks of source symbols into longer blocks corresponding to code words.
- G is a $k \times n$ matrix (k rows, n columns), that takes a source block u (a binary vector of length k), to a code word x (a binary vector of length n),

$$x = u \cdot G$$



Systematic Codes



- **Systematic codes:** The first k bits will be the original source block.
- From linear algebra, a $k \times n$ matrix of linearly independent rows can always be written into the systematic form:

$$G = [I_{k \times k} \vdots P_{k \times (n-k)}]$$

  where $I_{k \times k}$ is the $k \times k$ identity matrix, and $P_{k \times (n-k)}$ is a $k \times (n-k)$ matrix of parity bits.
- Every linear code is equivalent to a systematic code.

Example

Given the generator matrix for a (7,4) systematic code

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

Calculate the code words for the following source blocks:

i. 1010 [Answer: x = 1010101]

ii. 1101 [Answer: x = 1101001]

iii. 0010 [Answer: x = 0010110]

iv. Calculate the Hamming distances between each pair of code words generated in parts (i) to (iii), and compare them to the Hamming distances for the original source blocks. [Answer: 4, 7, 3, larger than 3, 4, 1]

Error Detection

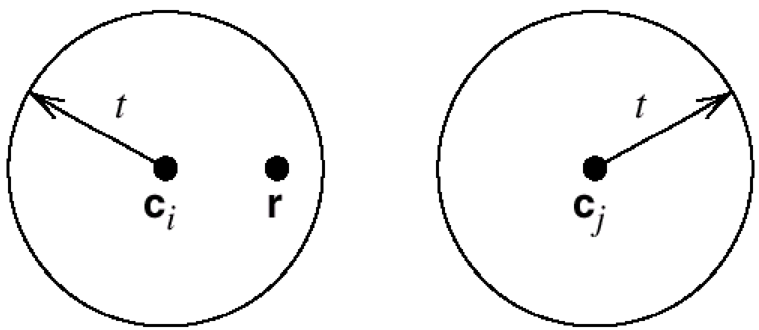
- To determine the number of errors a particular code can detect and correct, we look at the minimum Hamming distance between any two code words.
- From linearity the zero vector must be a code word.
- If we define the minimum distance between any two code words to be

$$d_{min} = \min\{d_H(a,b),\ a,b \in C\} = \min\{d_H(0,a+b),\ a,b \in C\}$$
$$= \min\{w_H(c),\ c \in C,\ c \neq 0\}$$

  where C is the set of code words.
- The number of errors that can be detected is then $(d_{min}-1)$, since d_{min} errors can turn an input code word into a different but valid code word. Less than d_{min} errors will turn an input code word into a vector that is not a valid code word.

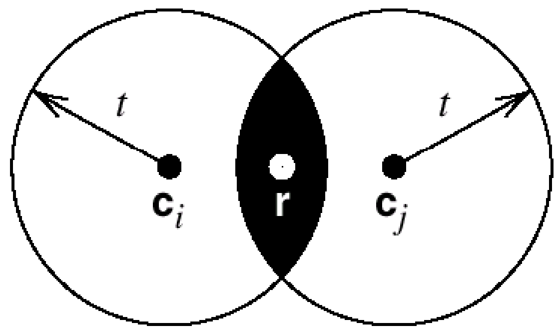
Error Correction

- The number *t* of errors that can be corrected is simply the number of errors that can be detected divided by two and rounded down to the nearest integer, since any output vector with less than this number of errors will ‘nearer’ to the input code word.



(a)

$d_{min} \geq 2t + 1$



(b)

$d_{min} < 2t$

Parity Check Matrix

- To decode a block coded vector, it is more complicated.
- If the generator matrix is of the form

$$G = [I_{k\times k} \vdots P_{k\times(n-k)}]$$

To check for errors, we define a new matrix, the **parity check matrix**, H. The parity check matrix is a $(n\text{-}k)\times n$ matrix that is defined so that

  – It will produce a zero vector when no error in the received code vector x,

$$x \cdot H^T = 0 \qquad (14.1)$$

  where H^T is the transpose of H.

- To satisfy this condition, it is sufficient to write the parity check matrix in the form

$$H = [(P_{k\times(n\text{-}k)})^T \vdots I_{(n-k)\times(n-k)}]$$

- <u>The minimum Hamming distance is equal to the smallest number of columns of H that are linearly dependent</u>.
  – This follows from the condition (14.1). See Haykin, Chap. 10 for proof.

Syndrome

- If the received coded block y contains errors, then the product of the received block with the transpose of the parity check matrix will not be zero,

$$y \cdot H^T \neq 0$$

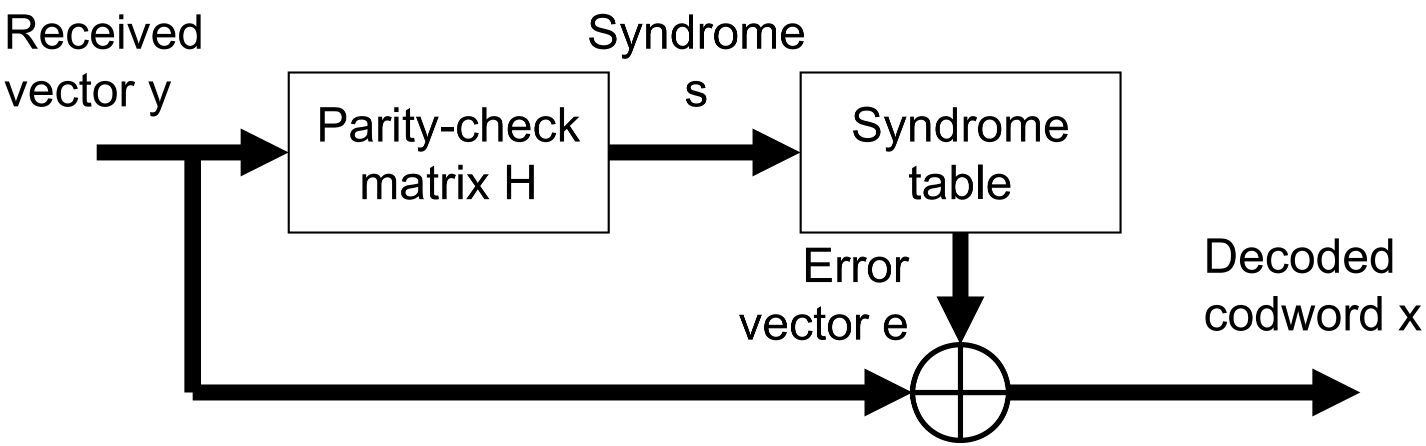
- Writing $y = x + e$ which is the sum of the original coded block x and the error e, we find that

$$y \cdot H^T = e \cdot H^T$$

- This value is referred to as the **syndrome**, $s = y \cdot H^T = e \cdot H^T$.
- The syndrome is a function of the error only, and contains the information required to isolate the position (or positions) of the error (or errors).

Syndrome Decoding

- s is an (n-k) row vector, taking $2^{(n-k)}$ -1 possible values (the zero vector corresponding to no error).
- This means that it is necessary to calculate and store $2^{(n-k)}$ -1 syndromes as a look-up table to be able to pin-point the positions of the errors exactly.
- Problem: this is impractical for large values of (n – k).



Summary

- A linear block code can be defined by a generator matrix G and the associated parity-check matrix H.
- Every linear block code is equivalent to a systematic code.
- The key parameter of a linear block code is the minimum Hamming distance $d_{\min}$:
  - Up to $d_{\min} - 1$ errors can be detected;
  - Up to $\lfloor (d_{\min} - 1)/2 \rfloor$ errors can be corrected.
- Syndrome decoding: compute the syndrome and then find the error pattern.
  - Only practical for short codes.

Imperial College London

EE2-4: Communication Systems

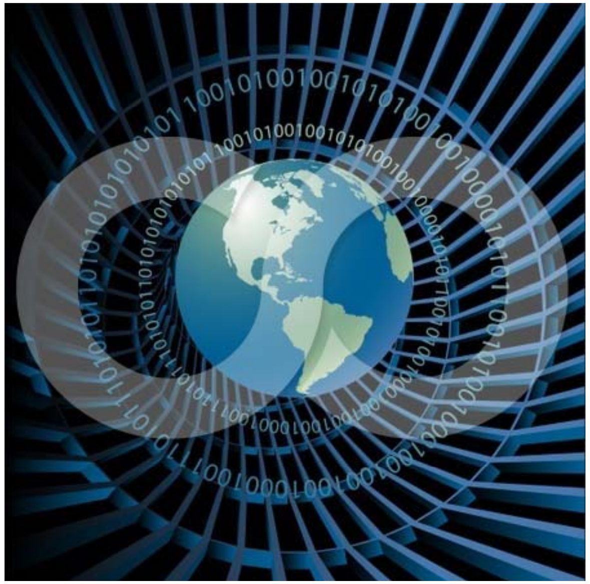
Lecture 15: Cyclic Codes

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Outline

- Hamming codes
  - A special type of cyclic codes that correct a single error
- Cyclic codes
  - Can correct more errors
  - The most important class of block codes
  - Implementation takes advantage of polynomial multiplication/division



- References
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 10
  - Lathi, Modern Digital and Analog Communication Systems, 3rd ed., Chap. 16

Hamming Codes

- Hamming codes are a class of linear block codes that can correct a single error. They satisfy the condition $r = n - k = \log_2(n+1) \Rightarrow n = 2^r - 1,\ k = 2^r - 1 - r.$
- From this expression, it is easy to see that the first few Hamming codes correspond to

$$(n, k) = (7,4), (15,11), (31,26), \ldots$$

- They are easy to construct and are simple to use.
  - For all Hamming codes, $d_{min} = 3$.
  - All (correctable) error vectors have unit Hamming weight, and the syndrome associated with an error in the i'th column of the vector is the i'th row of H^T.
- Columns of H are binary representations of 1, …, n.

Syndrome Table

- For the (7,4) Hamming code, the parity check matrix is

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

- Generator matrix

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

- The corresponding syndrome table is:

$$s = eH^T$$

s	e
000	0000000
001	0000001
010	0000010
100	0000100
111	0001000
110	0010000
101	0100000
011	1000000

- When a coded vector is received, the syndrome is calculated and any single error identified, and corrected by exchanging the relevant bit with the other binary value – <u>However</u>, problems can occur if there is more than one error.

Example

- Consider the (7,4) Hamming code. If the code vector 1000011 is sent while 1000001 and 1001100 are received, decode the information bits.
- Answer:
  - The first vector

    $s=(1000001)H^T=(010)$

    $\Rightarrow e=(0000010)$

    $\Rightarrow x=(1000001)+(0000010)=(1000011)$

    $\Rightarrow u=(1000)$ correct (as there is one error)

  - The second vector

    $s=(1001100)H^T=(000)$

    $\Rightarrow$error-free

    $\Rightarrow x=(1001100)$

    $\Rightarrow u=(1001)$ wrong (as there are 4 errors)

Cyclic Codes

- Cyclic codes are a subclass of linear block codes offering larger Hamming distances, thereby stronger error-correction capability.
- Whereas the principal property of the simple linear block code is that the sum of any two code words is also a code word, the cyclic codes have an additional property: a **cyclic shift** of any code word is also a code word.
- A cyclic shift of a binary vector is defined by taking a vector of length n,

$$a = [a_0, a_1, \ldots, a_{n-1}]$$

and rearranging the elements,

$$a = [a_{n-1}, a_0, a_1, \ldots, a_{n-2}]$$

- **A code is cyclic if:**

$$(c_0, c_1, \ldots, c_{n-1}) \in C \Rightarrow (c_{n-1}, c_0, \ldots, c_{n-2}) \in C$$

Generator Matrices

- A cyclic code is still a linear block code, so all of the properties previously discussed hold for cyclic codes.
- They are constructed by defining a generator matrix, and an associated parity check matrix and are decoded using syndromes in exactly the same way as the other linear block codes that we have discussed.
- A generator matrix is defined in the same way as before, except that the rows are now cyclic shifts of one *n*-dimensional basis vector.

$$G = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-k} & 0 & \cdots & 0 \\ 0 & g_0 & g_1 & \cdots & g_{n-k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_0 & g_1 & \cdots & g_{n-k} \end{bmatrix}$$

Encoding

- The cyclic property leads to a very useful property: they can be represented simply (mathematically and in hardware) by polynomial operations.
- We start by looking at the code word *x* generated by a source block *u*,

$$x = u \cdot G = [x_0, x_1, \ldots, x_{n-1}]$$

- With the generator matrix in the cyclic (non-systematic) form given in the previous slide, the elements of the code words are,

$$x_0 = u_0 g_0$$

$$x_1 = u_0 g_1 + u_1 g_0$$

$$\ldots$$

$$x_{n-1} = u_{k-1} g_{n-k}$$

These elements take the general form,

$$x_l = \sum_{i=0}^{k-1} u_i g_{l-i}$$

Polynomial Representation

- Taking a binary vector x, it is possible to represent this as a polynomial in z (over the binary field)

  $x = [\, x_0 \,, x_1 \,, \ldots \,, x_{n-1} \,] \rightarrow x\,(z) = x_0\, z^{n-1} + x_1\, z^{n-2} + \ldots + x_{n-1}$

- Representing the original code word and the n-dimensional basis vector for the cyclic code as polynomials,

  $u = [u_0 \,, u_1 \,, \ldots \,, u_{k-1} \,] \rightarrow u\,(z) = u_0\, z^{k-1} + u_1\, z^{k-2} + \ldots + u_{k-1}$

  $\mathrm{g} = [g_0 \,, g_1, \ldots \,, g_{n-k} \,, 0, \ldots, 0]$

  $\rightarrow g\,(z) = g_0\, z^{n-k} + g_1 z^{n-k-1} + \ldots + g_{n-k}$

- We notice that the product of these two polynomials is exactly the same as the polynomial representation of the corresponding code vector.

  Nonsystematic Encoding: $x(z) = u(z) \cdot g(z)$

- This means that the problem of matrix-vector multiplication can be reduced to a problem of polynomial multiplication.

Advantage

- The main advantage is that it simplifies the cyclic shift operation, and thereby simplifies the hardware implementation of the code considerably.
- Multiplication of the code word by z shifts all of the coefficients along by one, and replacing the term $(x_0 z^n)$ by a term $(x_0 z^0)$, gives a cyclic shift.
- A simpler way to achieve the same result is to multiply the polynomial by z, divide by z^n-1, and take the remainder term, which can be written as

$$(z\,x(z)) \bmod (z^n - 1)$$

Parity Check Polynomials

- One property important to cyclic codes is the ability to factorise polynomials. Given a polynomial of the form $a(z) = z^n - 1 = z^n + 1$ ($n > 1$), it is always possible to find two polynomials, such that,

$$z^n + 1 = g(z)\, h(z)$$

- Taking $g(z)$ to be a generator polynomial, this condition is **sufficient** for a resultant code to be an (n,k) cyclic code, where:
  - $g(z)$ is a polynomial of degree $n-k = r$.
  - $h(z)$ is a polynomial of degree k.
- The other factor, the polynomial $h(z)$, turns out to be a **parity check polynomial**, playing the same role as the parity check matrix.
- If we have a valid code word, $x(z)$, the product of the code word with $h(z)$ is zero (modulo z^n+1), and if we have a code word contained an error,

$$\begin{aligned}(y(z)h(z))\text{mod}(z^n+1) &= [(x(z)+e(z))h(z)]\text{mod}(z^n+1) \\ &= [e(z)h(z)]\text{mod}(z^n+1)\end{aligned}$$

  which is called the **syndrome polynomial**, $s(z)$.

Encoding/Decoding

- Encoding: $x(z)=u(z)g(z)$;
- **Parity check polynomial:**

$$h(z) = [z^n+1] / g(z)$$

- Decoding: ($y(z)$ is the received vector)
  - Calculate the **syndrome polynomial:**

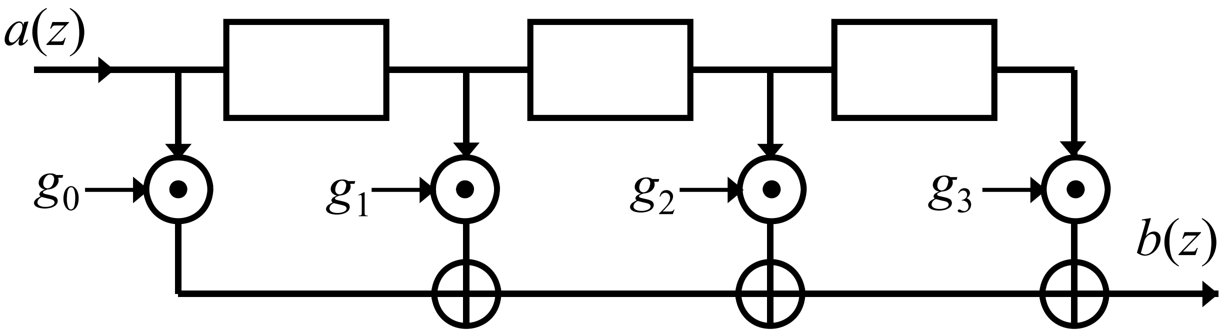
$$s(z)=(y(z)h(z))\mathrm{mod}(z^n+1)$$

  - Look up the syndrome table to get $e(z)$ from $s(z)$;
  - $x(z)=y(z)+e(z)$;
  - $u(z)=x(z)/g(z)$;

- This is the en/decoding procedure for non-systematic codes. A modified version works for systematic codes (non-examinable).

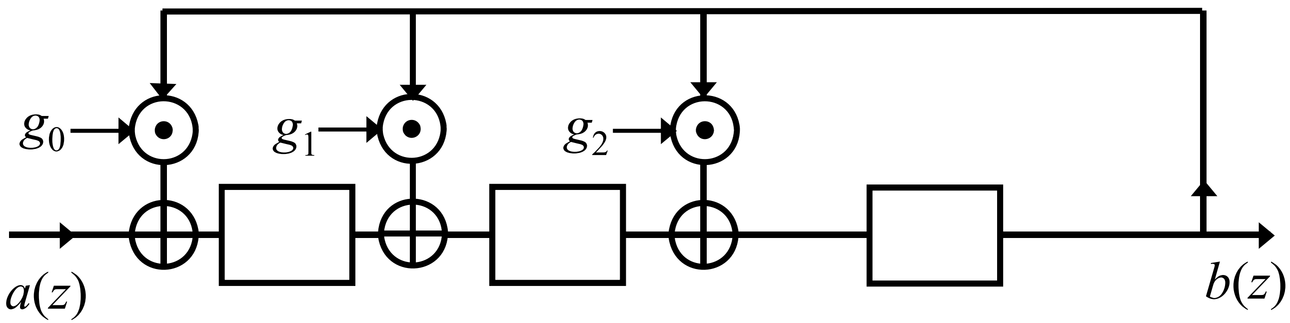
Hardware Implementation

- Binary polynomial multiplication: multiply-by-$g(z)$



$$g(z)=g_0+g_1z+g_2z^2+g_3z^3, \quad g_0=g_3=1;$$

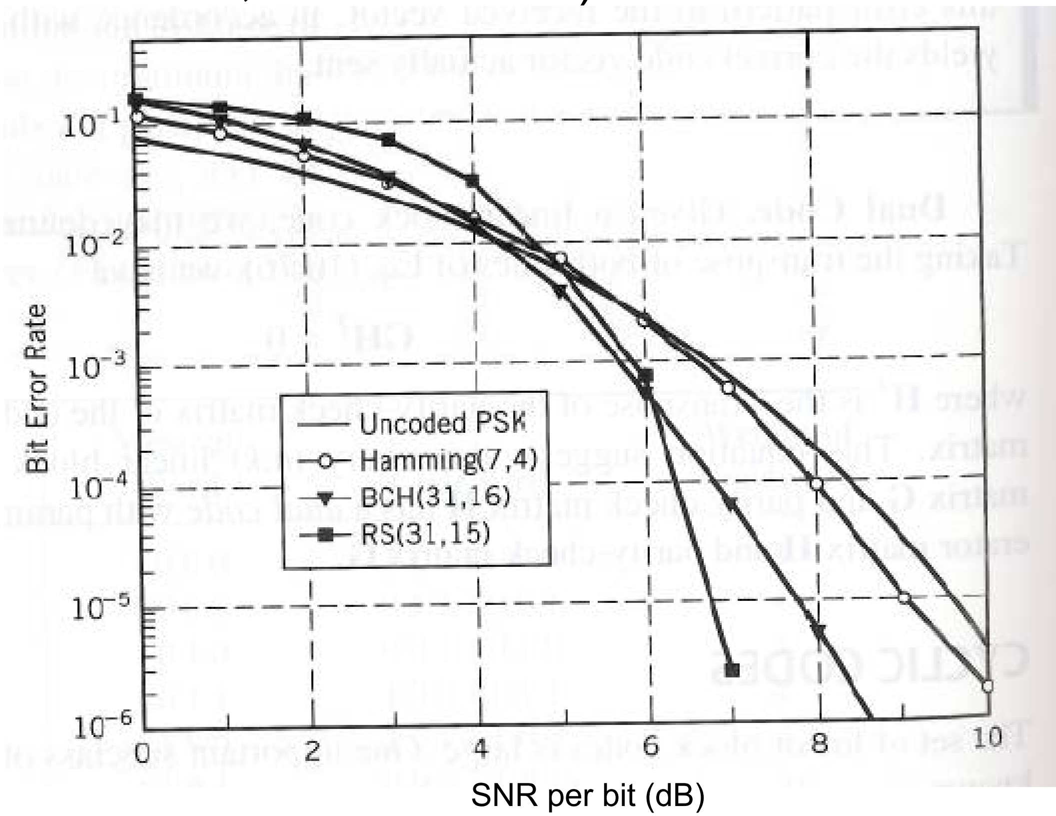
- Binary polynomial division: divide-by-$g(z)$



$$g(z)=g_0+g_1z+g_2z^2+g_3z^3, \quad g_0=g_3=1;$$

Examples of Cyclic Codes

- Hamming code (used in computer memory)
- Cyclic redundancy check (CRC) code (used in Ethernet, ARQ etc.)
- Bose-Chaudhuri-Hocquenghem (BCH) code
- Reed-Solomon (RS) code (widely used in CD, DVD, hard disks, wireless, satellites etc.)



Applications of Coding

- The first success was the application of convolutional codes in deep space probes 1960’s-70’s.
  - Mariner Mars, Viking, Pioneer missions by NASA
- Voyager, Galileo missions were further enhanced by concatenated codes (RS + convolutional).
- The next chapter was trellis coded modulation (TCM) for voice-band modems in 1980’s.
- 1990’s saw turbo codes approached capacity limit (now used in 3G).
- Followed by another breakthrough – space-time codes in 2000’s (used in WiMax, 4G)
- The current frontier is network coding which may widen the bottleneck of Internet.

Fourier Transform Theorems[a]

Name of Theorem		
1. Superposition (a_1 and a_2 arbitrary constants)	$a_1x_1(t) + a_2x_2(t)$	$a_1X_1(f) + a_2X_2(f)$
2. Time delay	$x(t - t_0)$	$X(f)e^{-j2\pi ft_0}$
3a. Scale change	$x(at)$	$\lvert a\rvert^{-1}X\left(\frac{f}{a}\right)$
b. Time reversal	$x(-t)$	$X(-f) = X*(f)$
4. Duality	$X(t)$	$x(-f)$
5a. Frequency translation	$x(t)e^{j\omega_0 t}$	$X(f - f_0)$
b. Modulation	$x(t)\cos \omega_0 t$	$\frac{1}{2}X(f - f_0) + \frac{1}{2}X(f + f_0)$
6. Differentiation	$\frac{d^n x(t)}{dt^n}$	$(j2\pi f)^n X(f)$
7. Integration	$\int_{-\infty}^{t} x(t')\,dt'$	$(j2\pi f)^{-1}X(f) + \frac{1}{2}X(0)\delta(f)$
8. Convolution	$\int_{-\infty}^{\infty} x_1(t - t')x_2(t')\,dt' = \int_{-\infty}^{\infty} x_1(t')x_2(t - t')\,dt'$	$X_1(f)X_2(f)$
9. Multiplication	$x_1(t)x_2(t)$	$\int_{-\infty}^{\infty} X_1(f - f')X_2(f')\,df' = \int_{-\infty}^{\infty} X_1(f')X_2(f - f')\,df'$

[a] $\omega_0 = 2\pi f_0$; $x(t)$ is assumed to be real in 3b.

Fourier Transform Pairs

Pair Number	$x(t)$	$X(f)$
1.	$\Pi\left(\frac{t}{\tau}\right)$	$\tau \operatorname{sinc} \tau f$
2.	$2W \operatorname{sinc} 2Wt$	$\Pi\left(\frac{f}{2W}\right)$
3.	$\Lambda\left(\frac{t}{\tau}\right)$	$\tau \operatorname{sinc}^2 \tau f$
4.	$\exp(-\alpha t)u(t),\ \alpha > 0$	$\frac{1}{\alpha + j2\pi f}$
5.	$t\exp(-\alpha t)u(t),\ \alpha > 0$	$\frac{1}{(\alpha + j2\pi f)^2}$
6.	$\exp(-\alpha\|t\|),\ \alpha > 0$	$\frac{2\alpha}{\alpha^2 + (2\pi f)^2}$
7.	$e^{-\pi(t/\tau)^2}$	$\tau e^{-\pi(f\tau)^2}$
8.	$\delta(t)$	1
9.	1	$\delta(f)$
10.	$\delta(t - t_0)$	$\exp(-j2\pi f t_0)$
11.	$\exp(j2\pi f_0 t)$	$\delta(f - f_0)$
12.	$\cos 2\pi f_0 t$	$\frac{1}{2}\delta(f - f_0) + \frac{1}{2}\delta(f + f_0)$
13.	$\sin 2\pi f_0 t$	$\frac{1}{2j}\delta(f - f_0) - \frac{1}{2j}\delta(f + f_0)$
14.	$u(t)$	$(j2\pi f)^{-1} + \frac{1}{2}\delta(f)$
15.	$\operatorname{sgn} t$	$(j\pi f)^{-1}$
16.	$\frac{1}{\pi t}$	$-j\operatorname{sgn}(f)$
17.	$\hat{x}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\lambda)}{t - \lambda}\,d\lambda$	$-j\operatorname{sgn}(f)X(f)$
18.	$\sum_{m=-\infty}^{\infty}\delta(t - mT_s)$	$f_s\sum_{m=-\infty}^{\infty}\delta(f - mf_s)$, $f_s = T_s^{-1}$

Trigonometic Identities

Sum and Difference Identities:

$$\sin(x+y) = \sin x \cos y + \cos x \sin y$$

$$\sin(x-y) = \sin x \cos y - \cos x \sin y$$

$$\cos(x+y) = \cos x \cos y - \sin x \sin y$$

$$\cos(x-y) = \cos x \cos y + \sin x \sin y$$

$$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x \tan y}$$

$$\tan(x-y) = \frac{\tan x - \tan y}{1 + \tan x \tan y}$$

Double Angle Identities:

$$\sin 2x = 2 \sin x \cos x$$

$$\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x$$

$$\tan 2x = \frac{2\tan x}{1 - \tan^2 x} = \frac{2\cot x}{\cot^2 x - 1} = \frac{2}{\cot x - \tan x}$$

Half Angle Identities:

$$\sin \frac{x}{2} = \pm\sqrt{\frac{1 - \cos x}{2}}$$

$$\cos \frac{x}{2} = \pm\sqrt{\frac{1 + \cos x}{2}}$$

$$\tan \frac{x}{2} = \pm\sqrt{\frac{1 - \cos x}{1 + \cos x}} = \frac{\sin x}{1 + \cos x} = \frac{1 - \cos x}{\sin x}$$

Product-Sum Identities:

$$\sin x \cos y = \frac{1}{2}[\sin(x+y) + \sin(x-y)]$$

$$\cos x \sin y = \frac{1}{2}[\sin(x+y) - \sin(x-y)]$$

$$\sin x \sin y = \frac{1}{2}[\cos(x-y) - \cos(x+y)]$$

$$\cos x \cos y = \frac{1}{2}[\cos(x+y) + \cos(x-y)]$$

Sum-Product Identities:

$$\sin x + \sin y = 2\sin\frac{x+y}{2}\cos\frac{x-y}{2}$$

$$\sin x - \sin y = 2\cos\frac{x+y}{2}\sin\frac{x-y}{2}$$

$$\cos x + \cos y = 2\cos\frac{x+y}{2}\cos\frac{x-y}{2}$$

$$\cos x - \cos y = -2\sin\frac{x+y}{2}\sin\frac{x-y}{2}$$

EE2-4: Communication Systems

Revision Lecture

Dr. Cong Ling

Department of Electrical and Electronic Engineering

Lectures

Introduction and background

1. Introduction
2. Probability and random processes
3. Noise

Effects of noise on analog communications

4. Noise performance of DSB
5. Noise performance of SSB and AM
6. Noise performance of FM
7. Pre/de-emphasis for FM and comparison of analog systems

Digital communications

8. Digital representation of signals
9. Baseband digital transmission
10. Digital modulation
11. Noncoherent demodulation

Information theory

12. Entropy and source coding
13. Channel capacity
14. Block codes
15. Cyclic codes

The Exam

- The exam paper contains 3 questions. All questions are compulsory.
- For example, the questions may look like
  - Question 1 (40 marks): Basic knowledge of communication systems, elements of information theory and/or coding, mostly bookwork
  - Question 2 (30 marks): Analog or digital communications, bookwork, new example/application/theory
  - Question 3 (30 marks): Digital or analog communications or information theory/coding, bookwork, new example/application/theory
- Sample questions:
  - Past papers
  - Problems in classes

Introduction, Probability and Random Processes

- Primary resources in communications: power, bandwidth, cost
- Objectives of system design: reliability and efficiency
- Performance measures: SNR or bit error probability
- Probability distribution: Uniform distribution, Gaussian distribution, Rayleigh distribution, Ricean distribution
- Random process: stationary random process, auto-correlation and power spectral density, Wiener-Khinchine relation

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau$$

Noise

- Why is noise important in communications? How does noise affect the performance? What types of noise exist?
- White noise: PSD is constant over an infinite bandwidth.
- Gaussian noise: PDF is Gaussian.
- Additive white Gaussian noise
- Bandlimited noise, bandpass representation, baseband noise $n_c(t)$ and $n_s(t)$, power spectral density

$$n(t) = n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

$$S_c(f) = S_s(f) = \begin{cases} S_N(f - f_c) + S_N(f + f_c), & |f| \leq B \\ 0, & \text{otherwise} \end{cases}$$

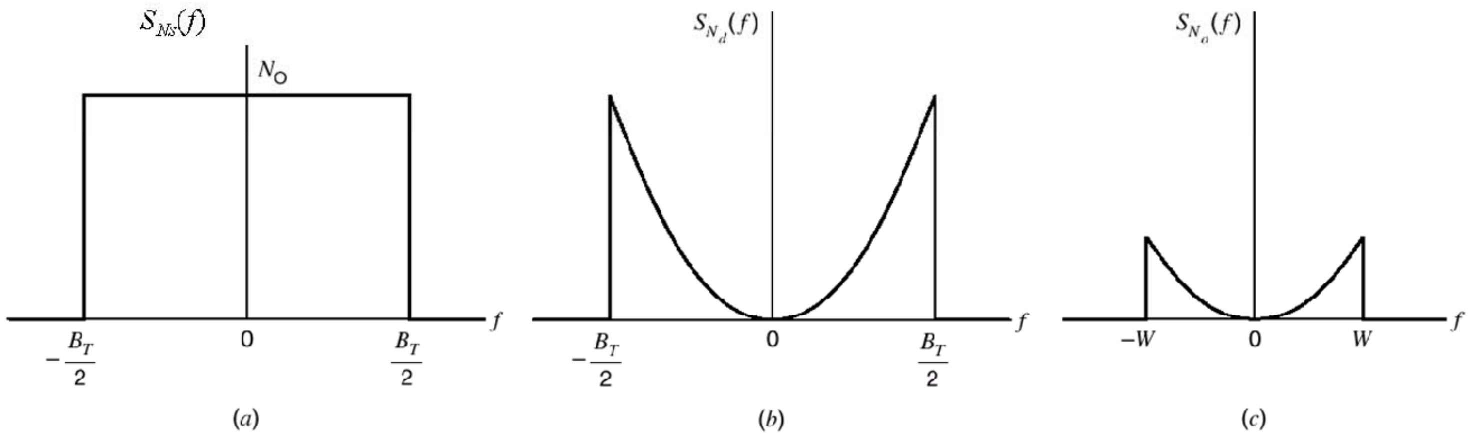
Noise performance of AM

- Signal-to-noise ratio $$SNR = \frac{P_S}{P_N}$$
- Baseband communication model $$SNR_{\text{baseband}} = \frac{P_T}{N_0 W}$$
- AM, DSB-SC, SSB, synchronous detection, envelope detection
- Output SNR

(De-) Modulation Format	Output SNR	Transmitted Power	Baseband Reference SNR	Figure of Merit (= Output SNR / Reference SNR)
AM Coherent Detection	$\frac{P}{2N_0W}$	$\frac{A^2+P}{2}$	$\frac{A^2+P}{2N_0W}$	$\frac{P}{A^2+P}<1$
DSB-SC Coherent Detection	$\frac{A^2P}{2N_0W}$	$\frac{A^2P}{2}$	$\frac{A^2P}{2N_0W}$	1
SSB Coherent Detection	$\frac{A^2P}{4N_0W}$	$\frac{A^2P}{4}$	$\frac{A^2P}{4N_0W}$	1
AM Envelope Detection (Small Noise)	$\frac{P}{2N_0W}$	$\frac{A^2+P}{2}$	$\frac{A^2+P}{2N_0W}$	$\frac{P}{A^2+P}<1$
AM Envelope Detection (Large Noise)	Poor	$\frac{A^2+P}{2}$	$\frac{A^2+P}{2N_0W}$	Poor

Noise performance of FM

- FM modulator and demodulator
- Method to deal with noise in FM: linear argument at high SNR
- Derivation of the output SNR, threshold effect



- Pre-emphasis and de-emphasis, how they increase the output SNR

Digital communications

- PCM: sample, quantize, and encode
- Quantization noise and SNR $SNR_o(\text{dB}) = 6n + 10\log_{10}\left(\frac{3P}{m_p^2}\right)$ (dB)
- Companding (A/μ-law) and line coding
- Baseband data transmission, effects of noise, and probability of error

Noise performance of bandpass digital communications

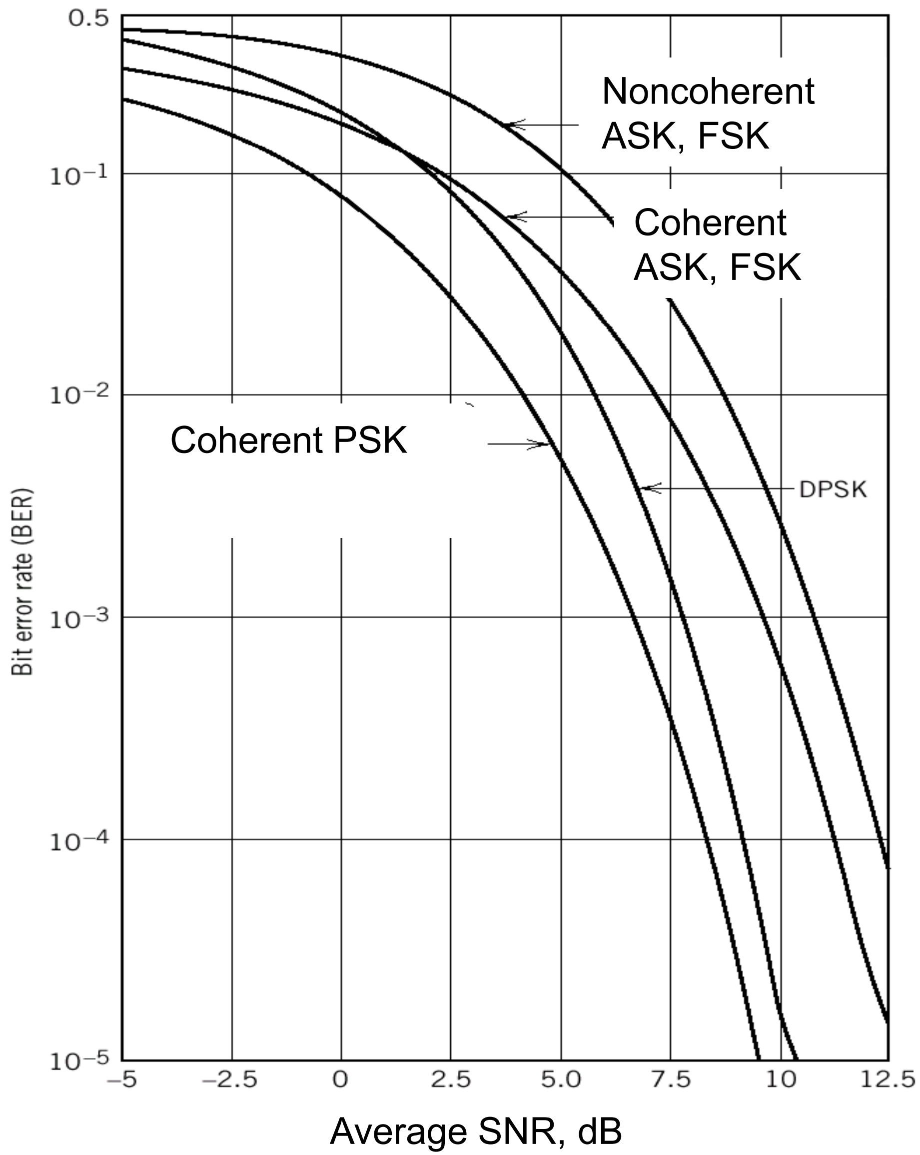
- Modulation formats: ASK, FSK, PSK
- Coherent detection and its error probability
- Noncoherent detection and its error probability (including differential detection for DPSK)
- Q-function Q(x): computation by using the graph or approximation

$$Q(x) \leq \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, x \geq 0$$

Performance of digital modulation

Scheme	Bit Error Rate
Coherent ASK	$Q(A/2\sigma)$
Coherent FSK	$Q(A/\sqrt{2}\sigma)$
Coherent PSK	$Q(A/\sigma)$
Noncoherent ASK	$\frac{1}{2}\exp(-A^2/8\sigma^2)$
Noncoherent FSK	$\frac{1}{2}\exp(-A^2/4\sigma^2)$
DPSK	$\frac{1}{2}\exp(-A^2/2\sigma^2)$

Caution: ASK and FSK have the same bit error rate if measured by average SNR.



Information theory

- The entropy of a discrete memoryless information source

$$H(S) = -\sum_{k=1}^{K} p_k \log_2 p_k$$

- Entropy function (entropy of a binary memoryless source)

$$H(S) = -(1-p)\log_2(1-p) - p\log_2 p = H(p)$$

- Source coding theorem: The minimum average codeword length for any source coding scheme is $H(S)$ for a discrete memoryless source.

- Huffman coding: An efficient source coding algorithm.

Channel coding theorem

- **If the transmission rate $R \leq C$, then there exists a coding scheme such that the output of the source can be transmitted over a noisy channel with an arbitrarily small probability of error. Conversely, it is not possible to transmit messages without error if $R > C$.**
- **Shannon formula**

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

Channel coding

- Block vs. convolutional codes
- Binary fields and vector space, Hamming distance/weight
- A linear block code can be defined by a generator matrix G and the associated parity-check matrix H.
- Every linear block code is equivalent to a systematic code.
- The key parameter of a linear block code is the minimum Hamming distance $d_{\min}$:
  - Up to $d_{\min} - 1$ errors can be detected;
  - Up to $\lfloor (d_{\min} - 1)/2 \rfloor$ errors can be corrected.
- Syndrome decoding: compute the syndrome and then find the error pattern.
- Hamming codes
- Cyclic codes: (polynomial representation is not examinable)

Appendix:

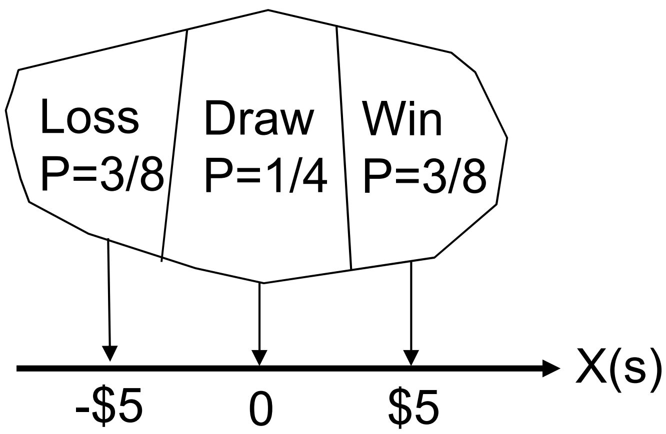
More Background on Probability

Probability

- Sample Space
  - S : Set of all random experiment outcomes
  - S = { s : s is an outcome }
- Examples
  - For tossing a coin, S = { H, T }
  - Roll a die, S = { 1, 2, …, 6 }
- Event $E \subseteq S$
  - Roll a die twice: S = { (H,H), (H,T), (T,H), (T,T) }
  - Event *E* = { (H,H), (T,T) }
  - A collection of events, *F*. Obviously, $F \subseteq S$
- A probability measure on *F* is a function *P* : *F* → [0,1] satisfying the probability axioms:
  - *P*(S) = 1
  - *P*(*F*) ≥ 0
  - For events A, B belonging to *F*, if A ∩ B=0, *P*(A U B) = *P*(A) + *P*(B)

Random Variables

- A random variable $X(s)$ is a variable whose value depends on the outcome s of a random experiment with the defined probability measure.
- For convenience, we simply denote the random variable by X.
- Consider a gamble in Macau (an example of discrete random variables):



$$X(s) = \begin{cases} +5 & s \in W \\ 0 & s \in D \\ -5 & s \in L \end{cases} \qquad S = W \cup D \cup L$$

$$P(X(s) = 5) = P(s : s \in W) = P(W) = \tfrac{3}{8}$$

$$P(X(s) = 0) = P(s : s \in D) = P(D) = \tfrac{1}{4}$$

$$P(X(s) = -5) = P(s : s \in L) = P(L) = \tfrac{3}{8}$$

- Continuous random variables: take values that vary continuously, e.g., water flow of River Thames through London Bridge

CDF and pdf

- Cumulative Distribution Function (CDF), also known as Probability Distribution Function
- Probability Density Function (pdf)

$$\text{CDF} : F_X(x) = P(X \le x)$$

$$\text{pdf} : f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$

$$F_X(\infty) = \int_{-\infty}^{\infty} f_X(y)dy = 1$$

$$P(a < X \le b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(y)dy$$

Since $F_X(x)$ is non - decreasing

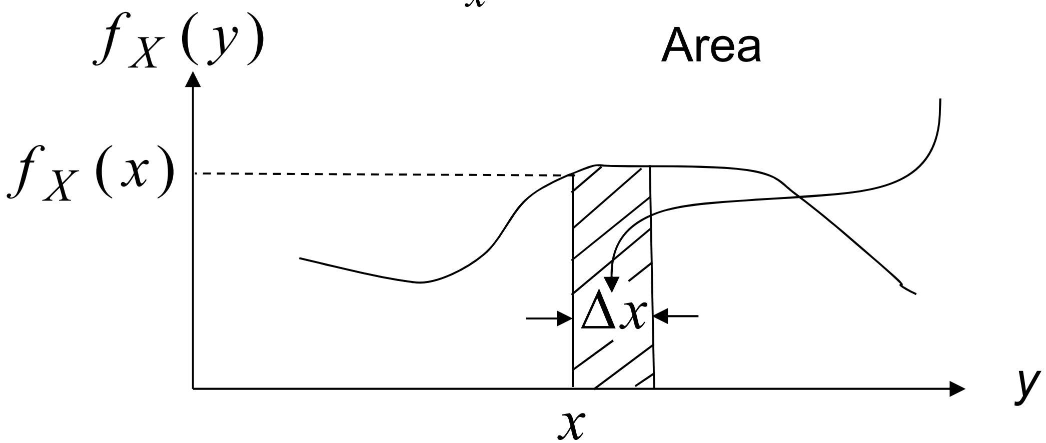
$$f_X(x) = \frac{dF_X(x)}{dx} \ge 0$$

Interpretation of pdf

- If Δx is sufficiently small,

$$P(x < X \le x + \Delta x) = \int_{x}^{x+\Delta x} f_X(y)dy \cong \underline{f_X(x)\Delta x}$$

Area



- Expectation operator (expected value or average):

$$E[X] \equiv \mu_X = \int_{-\infty}^{\infty} y\, f_X(y)dy$$

- Variance:

$$\sigma_X^2 \equiv E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (y-\mu_X)^2 f_X(y)dy = E[X^2] - \mu_X^2$$

Moments of a Random Variable

- Moments: $\mu_r = E[X^r]$ for $r = 1,2,3,...$

  Average (mean) of *X*:

  $$\mu_1 = E[X] \text{ or } m_X$$

- Central moments, centered on the mean

  $$\mu_r' = E[|X - m_X|^r] \quad \text{for } r = 1,2,3,...$$

- Comments

  $$\mu_1' = E[X - m_X] = E[X] - m_X = 0$$

  $$\text{Variance}: \quad \mu_2' = E[|X - m_X|^2]$$

  $$\text{Standard deviation}: \quad \sigma_X = \sqrt{\mu_2'}$$

  which gives a measure of "dispersion" of *X* about its mean

Examples of Discrete Distributions

- Let $p + q = 1$

- Bernoulli distribution

$$P(X = 1) = p \qquad P(X = 0) = q$$

with p representing the probability of success in an experiment

- Binomial distribution

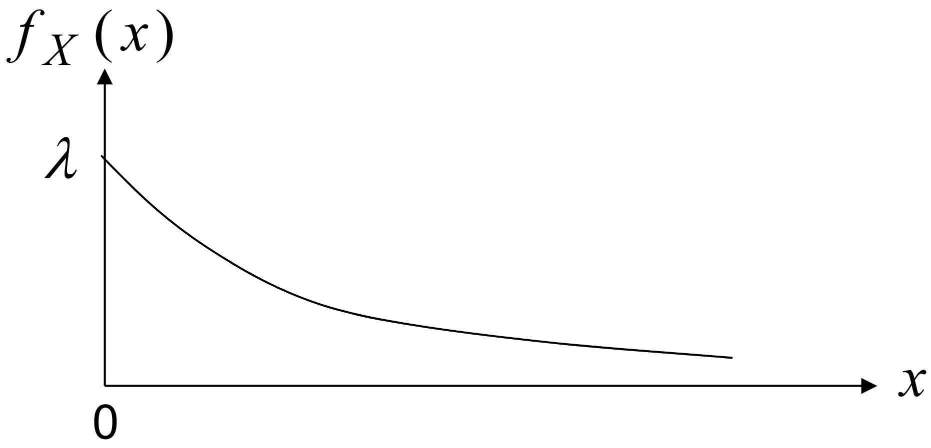
$$P(Y = k) = \binom{n}{k} p^k q^{n-k} \qquad k = 0, 1, 2, \ldots, n$$

Y represents the total number of successes, in an independent trial of n Bernoulli experiments

- Exercise: Verify $E[X] = p, \ \sigma_X^2 = pq$

$$E[Y] = np, \ \sigma_Y^2 = npq$$

Examples of Continuous Distributions: Exponential Distribution

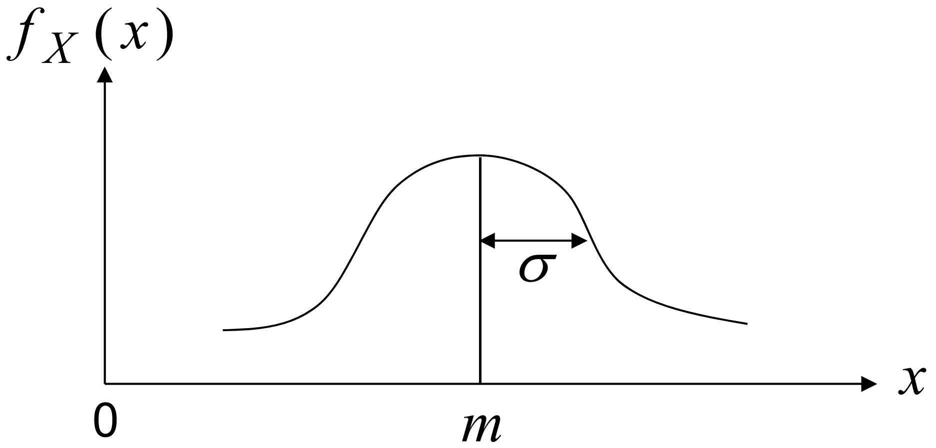


$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

- Exercise: Verify $E(X) = \frac{1}{\lambda}$

  $\sigma_X^2 = \frac{1}{\lambda^2}$

Normal (Gaussian) Distribution



$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

$$F_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}} dy$$

- Exercise: Verify $E(X) = m$

$$\sigma_X^{\,2} = \sigma^2$$

Rayleigh and Rice Distributions

- Define a random variable $R = \sqrt{X^2 + Y^2}$ where *X* and *Y* are

  independent Gaussian with zero mean and variance σ^2

- *R* has Rayleigh distribution:

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)} \qquad r \geq 0$$

- If *X* has nonzero mean *A*, *R* has Rice distribution:

$$f_R(r) = \frac{r}{\sigma^2} e^{-(r^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar}{\sigma^2}\right) \qquad r \geq 0$$

where $I_0(x) \equiv \frac{1}{2\pi}\int_0^{2\pi} e^{x\cos\theta} d\theta$ is the modified zero-order

Bessel function of the first kind

Conditional Probability

- Consider two events *A* and *B*, not necessarily independent of each other
  - Define *P*(*A*|*B*) = Probability of *A* given *B*
- Carry out *N* independent trials (experiments) and count
  - *N*(*A*) = Number of trials with *A* outcome
  - *N*(*A*,*B*) = Number of trials with both *A* and *B* outcome
- By definition

$$P(A) = \frac{N(A)}{N} \qquad \text{as } N \to \infty$$

$$P(A \mid B) = \frac{N(A,B)}{N(B)}$$

$$P(A \cap B) = \frac{N(A,B)}{N} = \frac{N(A,B)}{N(A)} \frac{N(A)}{N} = P(B \mid A) P(A)$$

$$\text{Similarly,} \quad P(A \cap B) = P(A \mid B) P(B)$$

- Statistical independence between *A* and *B*

$$P(A \mid B) = P(A) \quad \Leftrightarrow \quad P(A \cap B) = P(A) P(B)$$

Joint Random Variables

Random variables : X, Y

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)\,du\,dv$$

$f_{XY}(x,y)$: the joint pdf

$F_{XY}(x,y)$: the joint CDF

- Properties of joint distribution:

1) $F_{XY}(\infty,\infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(u,v)\,du\,dv = 1$

2a) $f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x,y)\,dy$

2b) $f_Y(x) = \int_{x=-\infty}^{\infty} f_{XY}(x,y)\,dx$

3) $f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial \mathrm{x} \partial \mathrm{y}}$

4) X, Y are independent $\Leftrightarrow f_{XY}(x,y) = f_X(x) f_Y(y)$

Conditional CDF and pdf

Define $F_Y(y\,|\,x)$ as the conditional CDF for *Y* given *X* = *x.*

By conditional probability,

$$F_Y(y\,|\,x < X \le x+\Delta x) = \frac{P(Y \le y, x < X \le x+\Delta x)}{P(x < X \le x+\Delta x)} = \frac{\int_{-\infty}^{y}\int_{x}^{x+\Delta x} f_{XY}(u,v)\,du\,dv}{\int_{x}^{x+\Delta x} f_X(u)\,du}$$

$$= \frac{\int_{-\infty}^{y} f_{XY}(x,v)\,\Delta x\,dv}{f_X(x)\,\Delta x}$$

As $\Delta x \to 0$, $\quad F_Y(y\,|\,x) = \dfrac{\int_{-\infty}^{y} f_{XY}(x,v)\,dv}{f_X(x)}$

The conditional pdf

$$f_Y(y\,|\,x) = \frac{dF_Y(y|x)}{dy} \quad \Rightarrow \quad f_Y(y\,|\,x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

Joint Distribution Function of Several Random Variables

- The joint PDF of *n* random variables

$$X_1,\ldots,X_n$$

$$F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)=P(X_1\le x_1,X_2\le x_2,\ldots X_n\le x_n)$$

- The joint pdf

$$f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)\equiv\frac{\partial^n F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)}{\partial x_1\partial x_2\ldots\partial x_n}$$

- Independent random variables

$$F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$$

$$f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)=f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$$

- Uncorrelated random variables

$$E[X_iX_j]=E[X_i]E[X_j]\quad\forall i,j,i\neq j$$

Covariance and Correlation Coefficient

Covariance of *X* and *Y*:

$$\operatorname{cov}(X,Y) \equiv E[(X - m_X)(Y - m_Y)]$$

Correlation Coefficient:

$$\rho_{XY} \equiv \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-m_X)(Y-m_Y)]}{\sigma_X \sigma_Y}$$

Property 1: $-1 \leq \rho_{XY} \leq 1$

Property 2: *X* and *Y* are linearly related $(i.e., \; Y = aX + b)$ if and only if $\rho_{XY} = \pm 1$

Property 3: If *X* and *Y* are independent, $\rho_{XY} = 0$

Caution: The converse of Property 3 is not true in general. That is, if $\rho_{XY} = 0$, X and Y are not necessarily independent!

Joint Gaussian distribution

X, Y jointly normally distributed with $m_X = m_Y = 0, \sigma_X = \sigma_Y = \sigma$

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2\sigma^2(1-\rho_{XY}^2)}(x^2 - 2\rho_{XY}xy + y^2)\right\}$$

The marginal density :

$$f_Y(y) = \int_{x=-\infty}^{\infty} f_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{y^2}{2\sigma^2}\}$$

Conditional density function :

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi}\sigma\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2\sigma^2(1-\rho_{XY}^2)}(x - \rho_{XY}y)^2\right\}$$

Note that this conditional pdf is also a normal density function with mean $\rho_{XY}y$ and variance $\sigma^2(1-\rho_{XY}^2)$. The effect of the condition (having *Y*=*y* on *X*) is to change the mean of *X* to $\rho_{XY}y$ and to reduce the variance by $\sigma^2\rho_{XY}^2$.

Independent vs. Uncorrelated

- Independent implies Uncorrelated
- Uncorrelated does not imply Independence
- For jointly Gaussian random variables, Uncorrelated implies Independent (this is the only exceptional case!)

- Exercise: verify the above claim of jointly Gaussian random variables

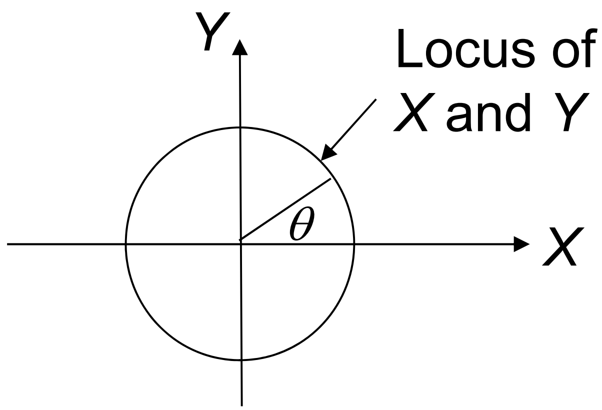
An example of uncorrelated but dependent random variables

Let θ be uniformly distributed in $[0, 2\pi]$

$$f_\theta(x) = \tfrac{1}{2\pi} \quad \text{for} \quad 0 \le x \le 2\pi$$

Define random variables *X* and *Y* as $X = \cos\theta \quad Y = \sin\theta$

Clearly, *X* and *Y* are not independent. In particular, for a given θ , X and Y are dependent.



If *X* and *Y* were independent, we should see possible sample points of (*X*,*Y*) assume all possible values of *X* and *Y* in a unit square.

But *X* and *Y* are uncorrelated as

$$\begin{aligned} \rho_{XY} &= E[(X - m_X)(Y - m_Y)] \\ &= E[XY] \\ &= \frac{1}{2\pi}\int_0^{2\pi} \cos\theta \sin\theta d\theta \\ &= 0! \end{aligned}$$

Central Limit Theorem

- Define $S_n \equiv \sum_{i=1}^{n} X_i$

  where X_i's are i.i.d. with $E[X] = \mu, \quad \sigma_X^2 = \sigma^2 < \infty$

- Define a new random variable, $R_n \equiv \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$

Central Limit Theorem :

$$\lim_{n \to \infty} P(R_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

- Importance: The "shifted and scaled" sum of *a very large number* of i.i.d. random variables has a Gaussian distribution