Imperial College London

Lecture 1: Overview on Communication

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

Imperial College London

1

Course Information

- Lecturer: Dr. Geoffrey Li
  - Office: No office yet
  - Email: Geoffrey.Li@imperial.ac.uk
- Handouts
  - Slides
  - Notes: more details
  - Problem sheets
- Grading
  - Labs
  - Mid-term exam
  - 2-hour exam

2

Overview

Introduction and background

Effects of noise on analog communications

Digital communications

Information theory

ELEC95005 vs. EE1-6

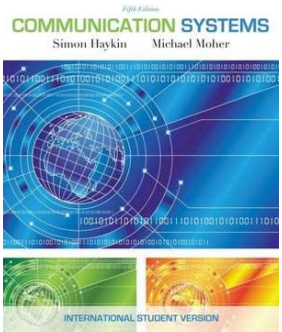
- Introduction to Signals and Communications (EE1-6)
  - How do communication systems work?
  - Modulation, demodulation, signal analysis, ...
  - The main mathematical tool is the Fourier transform for deterministic signal analysis
  - More about analog communications (i.e., signals are continuous)
- Communication Systems (ELEC95005)
  - How do communication systems perform in the presence of noise?
  - About statistical aspects and noise: Essential for a meaningful comparison of communication systems
  - The main mathematical tool is probability and stochastic processes
  - More about digital communications (i.e., signals are discrete)

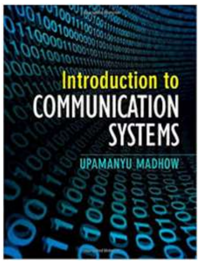
Learning Outcomes

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

References

- Lecture Notes
- S. Haykin & M. Moher, Communication Systems, 5th ed., International Student Version, Wiley, 2009.
- S. Haykin, Communication Systems, 4th ed., Wiley, 2001
  - Owns copyright of many figures in the slides
  - For convenience, the note "© 2000 Wiley, Haykin/Communication System, 4th ed." is not shown for each figure





- U. Madhow, Introduction to Communication Systems, Cambridge University Press, 2015.

Extra Readings for the Curios Mind

- The Mathematical Theory of Communication, C. E. Shannon and W. Weaver
- Fortune's Formula: The Untold Story of the Scientific Betting System that Beat the Casinos and the Wall Street, W. Poundstone
- Symbols, Signals and Noise: The Nature and Process of Communication, J. R. Pierce
- The Signal and the Noise: The Art and Science of Prediction, N. Silver
- A Diary on Information Theory, A. Renyi
- The Idea Factory: Bell Labs and the Great Age of American Innovation, Jon Gertner
- A Mind at Play: How Claude Shannon Invented the Information Age, Jimmy Soni and Rob Goodman

Different Communication Systems

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

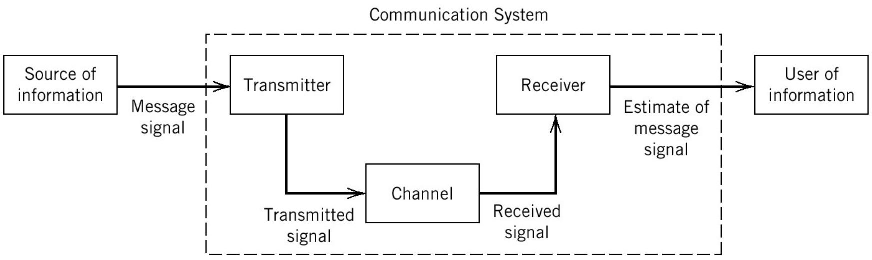
Analog and Digital Communications

- Analog communications
  - AM, FM
  - Analog TV
  - 1G:
  - AMPS (Advanced Mobile Phone System) in North America and Australia
  - TACS (Total Access Commun System) in UK

- Digital communications
  - Transfer of information in bits
  - Digital TV, CDs, DVD
  - Broadband, 2G (GSM), 3G, LTE, 5G, …

What is Communication?

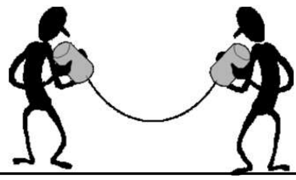
- Communication involves transmission of information from one point to another.
- Four basic elements
  - **Information source:** voice, music, picture, video, …
  - **Transmitter**: converts information in the source into a form suitable for transmission over the channel
  - **Channel**: the physical medium, introduces distortion, noise, interference
  - **Receiver**: reconstruct a recognizable form of the source signal



Communication Channels

- Cable, optical fiber, free space, underwater, hard disk, …
- Propagation loss: Signal strength decays with distance
- Bandwidth: Range of frequencies that can be used for communication. More bandwidth means higher transmission capacity:
  - Copper wire: 1 MHz
  - Coaxial cable: 100 MHz
  - Microwave: GHz
  - Optical fiber: THz

- Time-variation: Channel characteristics change over time (e.g. mobile radio channels)
- Nonlinearity: Some elements (e.g. repeaters) might introduce nonlinearity (e.g. satellite channel)
- Multipath: causing frequency selectivity
- Noise



Noise in Communications

- Unwanted signals present in a communication system
- Derived from Latin word "nausea" meaning seasickness
- External noise: interference from nearby channels, human-made noise, natural noise...
- Internal noise: thermal noise, random motion of electrons
- Noise limits the performance of communication systems
- A widely used metric is the signal-to-noise ratio (SNR)

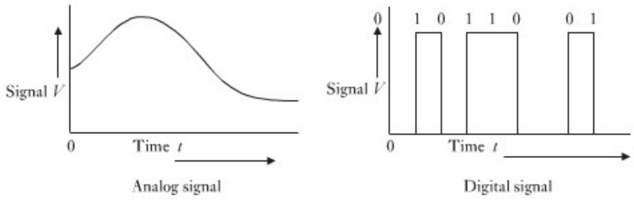
$$\text{SNR} = \frac{\text{signal power}}{\text{noise power}}$$

Transmitter and Receiver

- Transmitter converts the source signal into a form suitable for transmission over the channel
- It includes modulation and up-conversion
- Modulation: Some parameter of a carrier wave is varied based on the source signal
  - Analog: AM, FM, PM
  - Digital: ASK, FSK, PSK (SK: shift keying)
- Up-conversion: Modulated signal converted to final radio frequency (RF)
- Receiver reconstructs original message by down-conversion and demodulation
  - Recovery is not exact due to noise/distortion
  - Resulting degradation depends on the type of modulation and channel

13

Digital vs. Analog Communications



- Analog signals are continuous in time and amplitude (speech, image, video)
- Digital signals are discrete in time and amplitude
- Transmitted signals, including current, voltages, EM waves, are always analog
- Digital vs. analog refer to how parameters of these waveforms are formed
- Digital systems: source signal -> source messages

  -> a finite set of (continuous) signals
- Design of analog communication is conceptually simple
- Digital communication is more efficient and reliable; design is more sophisticated
- Digital design is universal and modular: any signal can be converted to digital format

14

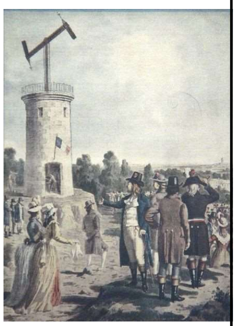
Objectives of System Design

- Two primary resources in communications
  - Transmitted power
  - Channel bandwidth (sometimes very expensive)

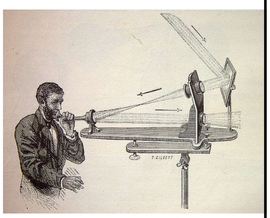
  One resource may be more limited than the other
  - Power-limited (e.g. deep-space communication)
  - Bandwidth-limited (e.g. telephone circuit)

- Objectives of communication system design: Delivering message efficiently and reliably, subject to power, bandwidth, and cost constraints.
  - Efficiency measured by the number of bits transmitted in unit power, unit time, and unit bandwidth.
  - Reliability in terms of SNR or probability of error.

Milestones in Communications

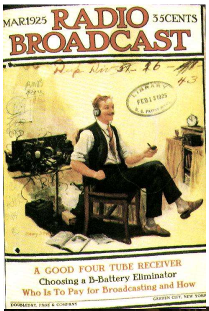


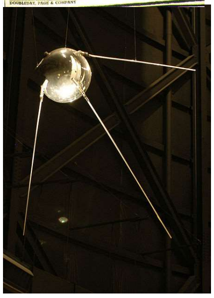
- 1792, Chappe invents “semaphore” (optical telegraph)
- 1837, First commercial telegraph service implemented between Paddington station and West Drayton by William Cooke and Charles Wheatstone in 1839
- 1851, England connected to Europe by a cable between Dover and Calais
- 1864, Maxwell formulated the electromagnetic (EM) theory
- 1875, Bell invented the telephone (Alexander Graham Bell: born in Scotland and Immigrated to USA)
- 1880, Bell invented “photophone” which transmits speech on an optical beam
- 1887, Hertz demonstrated physical evidence of EM waves
- 1890’s-1900’s, Marconi & Popov, long-distance radio telegraph
  - Across Atlantic Ocean, from Cornwall (UK) to Canada



Milestones (2)

- 1906, First radio broadcast
- 1918, Armstrong invented superheterodyne radio receiver (and FM in 1933)
- 1920, First commercial radio station (500 stations by 1923)
- 1922, British Broadcasting Corporation (BBC) founded
- 1925, John Logie Baird demonstrated transmission of moving images in London
- 1928, First TV station by General Electric factory in Schenectady, NY, broadcast for 2 hours/day for several years
- 1928, Nyquist sampling theorem at Bell Lab
- 1935, Robert Watson-Watt developed first practical radar system
- 1948, establishing information theory by Shannon at Bell Labs
- 1948, Invention of transistor at Bell Labs
- 1956, Bell System and British Post Office started transatlantic telephone service
- 1957, USSR placed first artificial satellite Sputnik





17

Milestones (3)

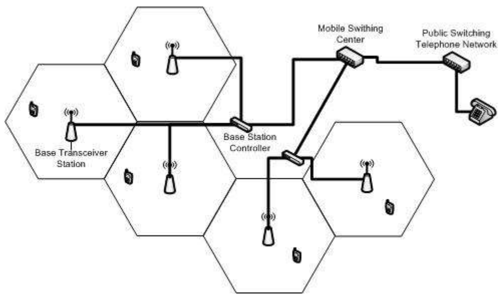
- 1960, AT&T placed first telecommunications satellite Echo I
- 1964, IBM built first commercial computer network SABRE Airline reservation system for American Airlines
- 1966, Kuen Kao pioneered fiber-optical communications (Nobel Prize Winner)
- 1971, First public wireless packet network ALOHANET demonstrated at University of Hawaii
- 1973, Metcalfe at Xerox Palo Alto invents Ethernet
- 1974, Vinton Cerf and Robert Kahn designed TCP/IP
- 1975, First experimental fiber-optic link installed in Atlanta by AT&T Bell

  first non-experimental fiber-optic link installed by the Dorset (UK) police
- 1978, First cellular mobile phone system (1G) developed by AT&T Bell Labs
- 1984, Internet connects 100 universities in the US and Europe (30K by 1987, and 160K by 1989)
- 1991, Digital cellular system (GSM, 2G) debuted in Europe

18

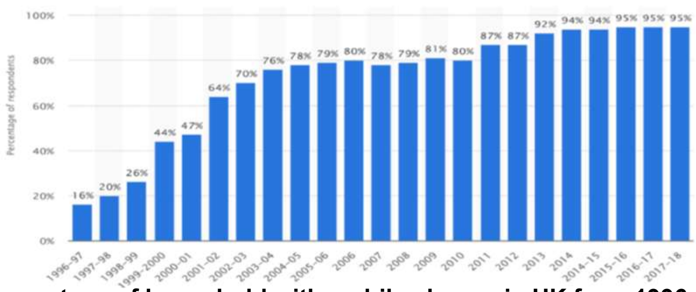
Cellular Mobile Phone Network

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



Growth of Mobile Communications

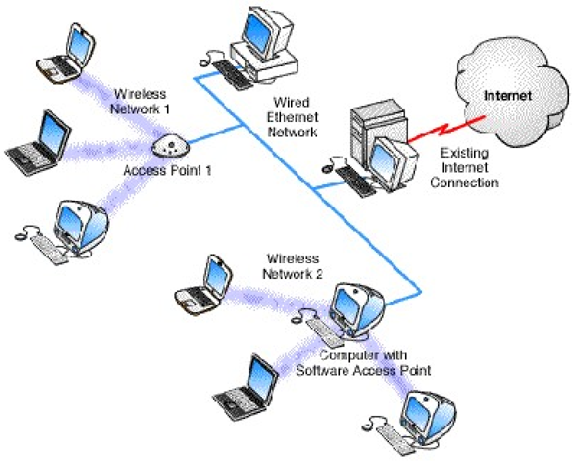
- 1G: analog communications
  - AMPS
- 2G: digital communications
  - GSM
  - IS-95
- 3G: CDMA networks
  - WCDMA
  - CDMA2000
  - TD-SCDMA
- 4G: up to 1 Gbps
  - WIMAX, LTE, LTE-A



Percentage of household with mobile phones in UK from 1996 - 2018

Wi-Fi

- Wi-Fi connects "local" devices (within ~100m range)



21

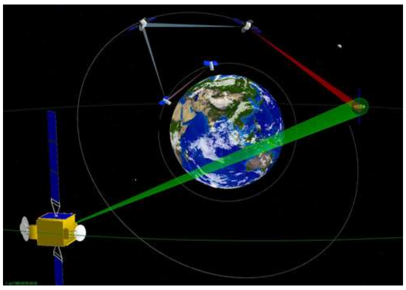
IEEE 802.11 Wi-Fi Standard

- 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)

22

Satellite/Space Communication

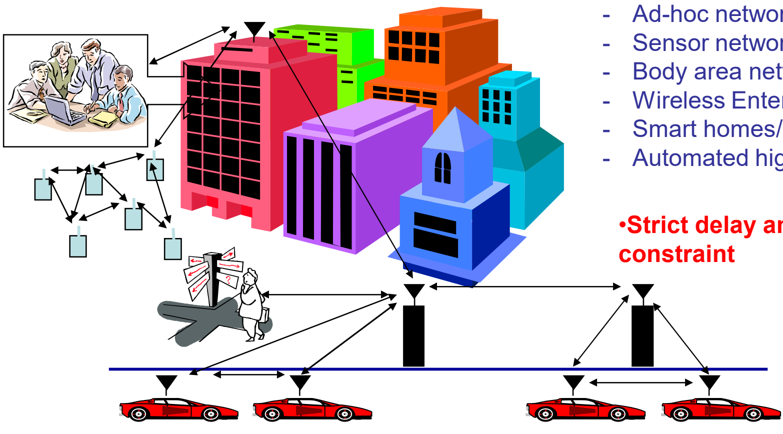
- Satellite communication
  - Cover very large areas
  - Optimized for one-way transmission
    - Radio (DAB) and movie (SatTV) broadcasting
  - Two-way systems
    - The only choice for remote-area and maritime communications
    - Propagation delay (0.25 s) is uncomfortable in voice communications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



23

Future Wireless Networks

Ubiquitous Communication Among People and Devices



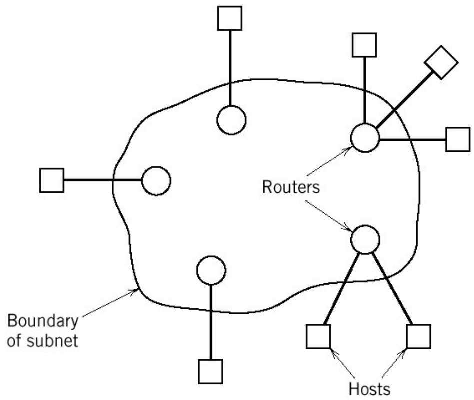
- Wireless Internet access
- Ad-hoc networks
- Sensor networks
- Body area networks
- Wireless Entertainment
- Smart homes/grids
- Automated highways

•**Strict delay and energy constraint**

24

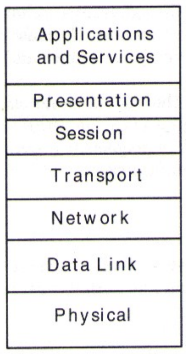
Communication Networks

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



Concept of Layering

- Partitioned into layers, each doing a relatively simple task



OSI Model

Application
Transport
Network
Link
Physical

TCP/IP protocol stack (Internet)

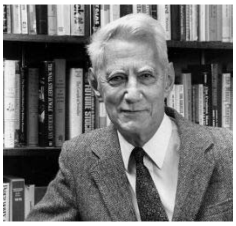
Physical layer: physical mechanism for transmitting bits;
Data-link layer: error protection
Medium access control (MAC): share the media
Network layer: routing
Transport layer: determine QoS, flow control
Application layer: user interface

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

Information Theory

- Is it possible to achieve zero probability of error even though the channel is noisy and power and bandwidth are limited?
- Shannon characterized the maximum rate of reliable transmission
  - Famous **Shannon capacity formula** for a channel with bandwidth W (Hz)

    C = W log(1+SNR) bps (bits/second)
  - (Almost) zero prob. of error is possible as long as actual signaling rate is less than C.
- Many fundamental concepts are developed using information theory, which paved the way for future developments in communication systems.
  - Provides a basis for tradeoff between SNR and bandwidth, and for comparing different communication schemes.



Imperial College London

Lecture 2: Probability and Random Processes

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

Imperial College London

Outline

- Probability
  - definition
  - 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

29

Why Probability/Random Process?

- Probability: the core mathematical tool for communication theory.
- Stochastic model: widely used in the study of communication systems.
- Random factors in a radio communication system:
  - Message is random: No randomness, no information.
  - Interference is random.
  - Noise is random.
  - And many more (delay, phase, fading, ...)
- Other real-world applications of probability and random processes include
  - Stock market modelling, gambling, weather forecast, control systems, …

30

Probability Space

- Sample space (S): Set of all possible outcomes
- Event: any subset of S
- The probability of an event is a non-negative number assigned to each event:
  - The probability of the event that includes all possible outcomes of the experiment is 1,i.e., P(S)=1
  - Probability of an event,

    P(A)≥0 for any subset A of S

  - The probability of two events that do not have any common outcome is the sum of the probabilities of the two events separately,

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$$

Random Variable

- A random variable **X(s)** is a real-valued function defined on the set of all possible outcomes S
- X(s) maps every outcome s into a real number
- {X≤x}: subset of S consisting of all outcomes s such that X(s)≤x

Example: S=all possible outcomes of two coin tosses

S={HH, HT, TH, TT}

Random variable X: number of heads in two tosses

X(HH)=2, X(HT)=1, X(TH)=1, X(TT)=0

CDF and PDF

- Cumulative distribution function (cdf) of a random variable:

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

- Properties

$$F_X(-\infty) = 0, \quad F_X(\infty) = 1$$
$$F_X(x_1) \le F_X(x_2) \quad \text{if } x_1 \le x_2$$

- Probability density function (pdf):

$$f_X(x) = \tfrac{dF_X(x)}{dx}$$
$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$
$$P(a < X \le b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(y)dy$$
$$f_X(x) = \tfrac{dF_X(x)}{dx} \ge 0 \ \text{ since } F_X(x) \text{ is non-decreasing}$$

33

Mean and Variance

- Mean (or expected value ⇔ DC level):

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

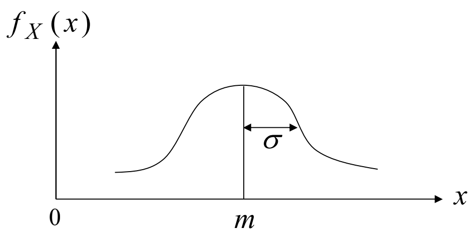
$E[\ ]$: expectation operator

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

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

34

Normal (Gaussian) Distribution



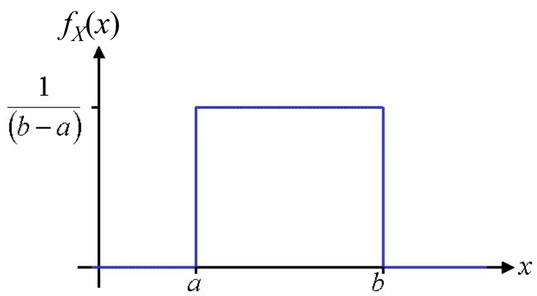
$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

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

$$E[X] = m$$
$$\sigma_X^2 = \sigma^2$$

35

Uniform Distribution



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

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

$$E[X] = \frac{a+b}{2}$$
$$\sigma_X{}^2 = \frac{(b-a)^2}{12}$$

36

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

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

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

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

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 $\underset{\not\leftarrow}{\rightarrow}$ uncorrelated

(c.pdf) (exp)

Independent vs. Uncorrelated

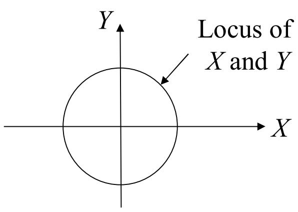
- Independent implies Uncorrelated
- Uncorrelated does not imply Independence
- For normal RVs (jointly Gaussian), Uncorrelated implies Independent (this is the exceptional case!)
- An example of uncorrelated but dependent RV's

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

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



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

But *X* and *Y* are uncorrelated:

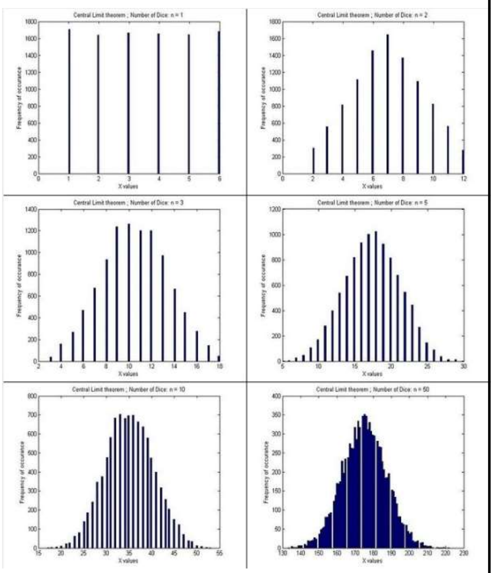
$$E[XY] = \frac{1}{2\pi}\int_0^{2\pi} \cos\theta \sin\theta \, d\theta = 0$$

Joint Distribution of *n* RVs

- Joint cdf $F_{X_1X_2...X_n}(x_1,x_2,...x_n) \equiv P(X_1 \le x_1, X_2 \le x_2,...X_n \le x_n)$
- Joint pdf
$$f_{X_1X_2...X_n}(x_1,x_2,...x_n) \equiv \frac{\partial^n F_{X_1X_2...X_n}(x_1,x_2,...x_n)}{\partial x_1 \partial x_2 ... \partial x_n}$$
- **Independent**
$$F_{X_1X_2...X_n}(x_1,x_2,...x_n) = F_{X_1}(x_1)F_{X_2}(x_2)...F_{X_n}(x_n)$$
$$f_{X_1X_2...X_n}(x_1,x_2,...x_n) = f_{X_1}(x_1)f_{X_2}(x_2)...f_{X_n}(x_n)$$
- **i.i.d.** (independent and identically distributed)
  - Independent random variables with the same distribution.
  - Example: outcomes from repeatedly flipping a coin.

Central Limit Theorem

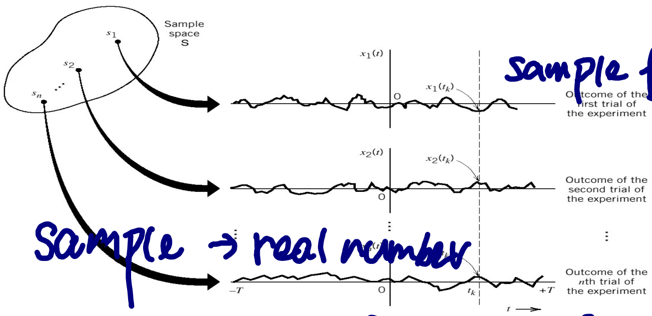
- For i.i.d. random variables,
  $z = x_1 + x_2 + \cdot\cdot\cdot + x_n$
  tends to Gaussian as *n* goes to infinity.
- Extremely useful in communications.
- That's why noise is usually Gaussian. We often say "Gaussian noise"
- Gaussian channel in communications: a channel with only Gaussian noise and without other distortions (fading, nonliniearity)



Random Process

- A random process: a function of time that assigns to each outcome s a function of time $X(t,s)$ for $-T \le t \le T$, where 2T is the total observation interval.
- For a fixed sample point s_j, function $X(t,s_j)$ versus time is called a sample function of the random process.

  It is actually a deterministic function of time
- For fixed t, a random process is a random variable.
- Noise can often be modelled as a **Gaussian random process**.



sample function

Random variable: sample → real number

Random process: sample → real function of time.

Statistics of a Random Process

- Mean of a random process at t:

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

It is *usually* a function of t.

mean: function of t

- Autocorrelation function

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

autocorrelation: between samples at two instants.

  – *In general*, the autocorrelation function is a two-variable function.
  – It measures the correlation between samples at different two different times.

Strict-Sense Stationary (SSS) Random Processes

A process is n^{th}-order Strict-Sense Stationary (SSS) if

$$f_X(x_1,x_2,\cdots x_n,\ t_1,t_2\cdots,t_n) \equiv f_X(x_1,x_2,\cdots x_n,\ t_1+c,t_2+c\cdots,t_n+c)$$

for *any c*, where the left side represents the joint density function of the random variables

$$X_1 = X(t_1),\ X_2 = X(t_2),\ \cdots,\ X_n = X(t_n)$$

and the right side corresponds to the joint density function of the random variables

$$X_1' = X(t_1+c),\ X_2' = X(t_2+c),\ \cdots,\ X_n' = X(t_n+c).$$

everything strict is w.r.t. pdf/cdf.

43

Stationary Random Processes

- A random process is stationary to first order if the distribution function (hence the density function) of X(t) is invariant over time.
- X(t) is stationary to second order if the joint distribution function $f_{X(t_1)X(t_2)}(x_1,x_2)$ depends only on the difference between t_2 and t_1.
- If X(t) is stationary to second order and autocorrelation function exits, then

$$R_X(t_1,t_2) = R_X(t_2-t_1),\ \text{for all } t_1 \text{ and } t_2$$

44

Wide-Sense Stationary (WSS) Random Processes

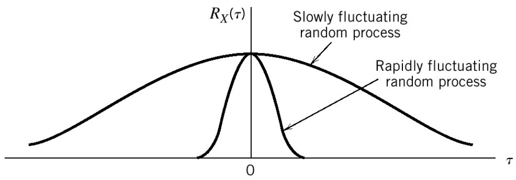
- A random process is **wide-sense stationary (WSS)** if
  - Its mean does not depend on *t*

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

  - Its autocorrelation function depends only on time difference

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

WSS is defined over 1-st and 2-nd moments, instead of 1.2 order SSS.



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

45

second-order SSS $\xrightarrow{\text{mean. autocorrelation exist}}$ WSS

SSS and WSS Random Processes

- Strict-sense stationarity always implies wide-sense stationarity if the autocorrelation function exists

- Converse holds only for Gaussian processes!

- If $X(t)$ is a Gaussian process, then WSS=SSS

46

WSS: $R_X(t, t+\tau) = R_X(\tau)$.

① $R_X(0) = R_X(t,t) = E[X(t)X(t)] = E[X^2(t)]$

② $R_X(-\tau) = R_X(t, t-\tau) = E[X(t)X(t-\tau)] = E[X(t-\tau)X(t)] = R_X(t-\tau, t) = R_X(\tau$

Properties of Autocorrelation Function

Let X(t) be stationary (WSS is enough) with autocorrelation function $R_X(\tau)$

1.We have $R_X(0)= E[X^2(t)]$

2.$R_X(\tau)$ is an even function,

$$R_X(\tau) = R_X(-\tau)$$

3.$R_X(\tau)$ takes its maximum magnitude at $\tau=0$

$$|R_X(\tau)| \le R_X(0)$$

$R_X(\tau)$ can tell how predictable X(t) is based on X(t-τ)

③ $E[(X(t)+X(t+\tau))^2] = E[(X(t))^2 + 2X(t)X(t+\tau) + (X(t+\tau))^2]$

$= 2R_X(0) + 2R_X(\tau) \ge 0$

$E[(X(t)-X(t+\tau))^2] = 2R_X(0) - 2R_X(\tau) \ge 0$

47

$\therefore -R_X(\tau) \le R_X(0) \le R_X(\tau)$. $\Rightarrow |R_X(\tau)| \le R_X(0)$

Example

- Show that sinusoidal wave with random phase

$$X(t) = A\cos(\omega_c t + \Theta) \qquad f_\Theta(\theta) = \frac{1}{2\pi}, \quad \theta \in [0, 2\pi]$$

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

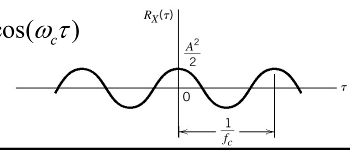
– Mean is a constant:

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

– Autocorrelation function depends only on the time difference:

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

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



Ergodic Processes

- Expectations of a stochastic process X(t): ensemble averages (averages "across the process"). It could change with time.

- Sometimes it is hard/impossible to observe all sample functions of a random process at a given time, while it may be possible to observe a single sample over a long period of time.

- If the time average is equal to the ensemble average, then the process is said to be ergodic.

49

LTI: $\begin{cases} \text{linear: i/o is related by a linear mapping. } f[X(t)] \to f[y(t)] \\ \text{time-invariant: for a given input, the output is irrelevant of time} \end{cases}$

$X(t-\tau) \to y(t-\tau)$

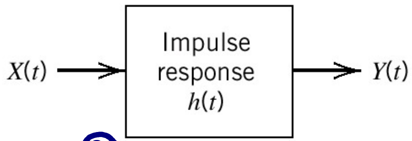
If X(t) is WSS, then $\mu_X(t) = \mu_X$.

$\mu_Y(t) = \int_{-\infty}^{+\infty} \mu_X h(t)\,dt$

$= \mu_X \int_{-\infty}^{+\infty} h(t)\,dt$

$= \mu_X H(0)$

Random Process Through LTI System

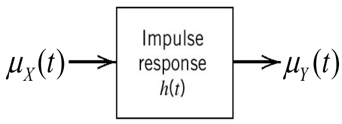


- If ① E[X(t)] is finite for all t and ② the system is stable, we have

$$\mu_Y(t) = E[Y(t)] = E\left[\int_{-\infty}^{+\infty} h(\tau)X(t-\tau)\,d\tau\right]$$

$$= \int_{-\infty}^{+\infty} h(\tau)E[X(t-\tau)]\,d\tau$$

$$= \mu_X(t) * h(t)$$



- Let ③ X(t) be WSS, then

$$\mu_Y(t) = \mu_X H(0)$$

where H(0) is the zero-frequency (dc) response of the system.

50

$$R_Y(t,u) = E[Y(t)Y(u)] = E\left[\int_{-\infty}^{+\infty} h(\tau_1)X(t-\tau_1)d\tau_1 \cdot \int_{-\infty}^{+\infty} h(\tau_2)X(u-\tau_2)d\tau_2\right]$$

$$\overset{①②}{=} \int_{-\infty}^{+\infty} d\tau_1 h(\tau_1)\int_{-\infty}^{+\infty} d\tau_2 h(\tau_2)\, E[X(t-\tau_1)X(u-\tau_2)]$$

$R_X(t-\tau_1, u-\tau_2)$

Random Process Through LTI System (2)

- If ① $E[X^2(t)]$ is finite for all t and the ② system is stable, we have

$$\overset{③}{=} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} R_X(t-u-\tau_1+\tau_2)h(\tau_1)h(\tau_2)\,d\tau_1 d\tau_2$$

$$R_Y(t,u) = \int_{-\infty}^{+\infty} d\tau_1 h(\tau_1)\int_{-\infty}^{+\infty} d\tau_2 h(\tau_2) E[X(t-\tau_1)X(u-\tau_2)]$$

$$= \int_{-\infty}^{+\infty} d\tau_1 h(\tau_1)\int_{-\infty}^{+\infty} d\tau_2 h(\tau_2) R_X(t-\tau_1, u-\tau_2)$$

$$= \int_{-\infty}^{+\infty} d\tau_1 h(\tau_1)[h(u) * R_X(t-\tau_1, u)]$$

$$= h(t) * [h(u) * R_X(t,u)]$$

$$= \int_{-\infty}^{+\infty} d\tau_1 h(\tau_1)\; h(u) * R_X(t-\tau_1, u)$$

$$= h(t) * h(u) * R_X(t,u)$$

$\tau \triangleq t-u$

$$= \int_{-\infty}^{+\infty} d\tau_1 h(\tau_1) \cdot h(-\tau) * R_X(\tau-\tau_1)$$

$$= h(\tau) * h(-\tau) * R_X(\tau)$$

$$R_{XX}(t,u) \longrightarrow \boxed{h(u)} \xrightarrow{R_{XY}(t,u)} \boxed{h(t)} \longrightarrow R_{YY}(t,u)$$

- ③ Let X(t) be WSS, then letting τ=t-u

$$R_Y(\tau) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(\tau_1)h(\tau_2)R_X(\tau-\tau_1+\tau_2)d\tau_1 d\tau_2 = h(\tau) * [h(-\tau) * R_X(\tau)]$$

LTI preserves WSS/SSS/Gaussian

51

Alternative proof:

$$R_{XY}(t,u) = E[X(t)Y(u)] = E\left[X(t)\int_{-\infty}^{+\infty} X(u-\tau)h(\tau)d\tau\right]$$

$$\overset{①②}{=} \int_{-\infty}^{+\infty} E[X(t)X(u-\tau)]h(\tau)d\tau = \int_{-\infty}^{+\infty} R_X(t, u-\tau)h(\tau)d\tau = R_X(t,u) * h(u)$$

Random Process Through LTI Filter (3)

$$R_{YY}(t,u) = E[Y(t)Y(u)] = E\left[\int_{-\infty}^{+\infty} X(t-\tau)h(\tau)d\tau\, Y(u)\right]$$

$$\overset{①②}{=} \int_{-\infty}^{+\infty} E[X(t-\tau)Y(u)]h(\tau)d\tau = \int_{-\infty}^{+\infty} R_{XY}(t-\tau, u)h(\tau)d\tau = R_{XY}(t,u) * h(t)$$

$X(t)$ wide-sense stationary process $\longrightarrow$ LTI system $h(t)$ $\longrightarrow$ $Y(t)$ wide-sense stationary process.

$\therefore R_{YY}(t,u) = h(t) * h(u) * R_{XX}(t,u)$

$X(t)$ strict-sense stationary process $\longrightarrow$ LTI system $h(t)$ $\longrightarrow$ $Y(t)$ strict-sense stationary process.

If $X(t)$ is WSS, $h(t)$ is LTI, then $Y(t)$ is also WSS.

$$R_{XY}(t,u) = \int_{-\infty}^{+\infty} R_{XX}(t, u-\alpha)h(\alpha)d\alpha = \int_{-\infty}^{+\infty} R_{XX}(\tau+\alpha)h(\alpha)d\alpha = R_{XX}(\tau) * h(-\tau)$$

$$R_{YY}(t,u) = \int_{-\infty}^{+\infty} R_{XY}(t-\beta, u)h(\beta)d\beta = \int_{-\infty}^{+\infty} R_{XY}(\tau-\beta)h(\beta)d\beta = R_{XY}(\tau) * h(\tau)$$

52

$\because R_{XY}(t,u) = R_{XY}(\tau)$ since X, Y are jointly WSS

$\therefore R_{YY}(\tau) = h(\tau) * h(-\tau) * R_{XX}(\tau)$.

Power Spectral Density (PSD)

- PSD is a function that measures the distribution of power of a random process over its spectrum.
- PSD is defined only for WSS processes.

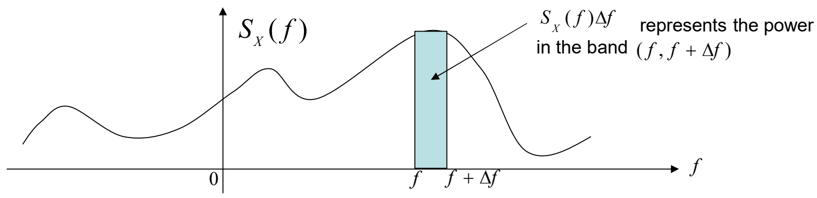
PSD $\underset{DFT}{\overset{IDFT}{\rightleftharpoons}}$ ATC

- **Einstein-Wiener-Khintchine relation:** The PSD of a wide-sense stationary (WSS) process is equal to the Fourier transform of its autocorrelation function:

$$\boxed{S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau \geq 0}$$

- 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.

53



Then the average power can be found as

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

Real WSS
$\Downarrow$

If $X(t)$ is a real WSS process, we have $R_X(\tau) = R_X(-\tau)$. Then

Even . nonnegative . real
PSD

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

$$= \int_{-\infty}^{+\infty} R_X(\tau) \cos 2\pi f\tau \, d\tau$$

$$= 2\int_{0}^{\infty} R_X(\tau) \cos 2\pi f\tau \, d\tau = S_X(-f) \geq 0$$

The power spectrum is an even (real and nonnegative) function.

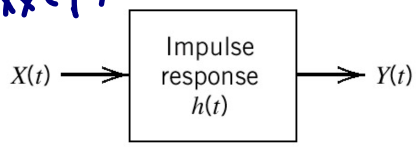
54

$R_{XY}(\tau) = h(-\tau) * R_{XX}(\tau) \Rightarrow S_{XY}(f) = H^*(f)\, S_{XX}(f)$

$R_{YY}(\tau) = h(\tau) * R_{XY}(\tau) \Rightarrow S_{YY}(f) = H(f)\, S_{XY}(f)$

$\therefore S_{YY}(f) = |H(f)|^2 S_{XX}(f)$

Passing Through a Linear System



- If a WSS process X(t) goes through a linear system of transfer function h(t):

$$S_Y(f) = |H(f)|^2 S_X(f)$$

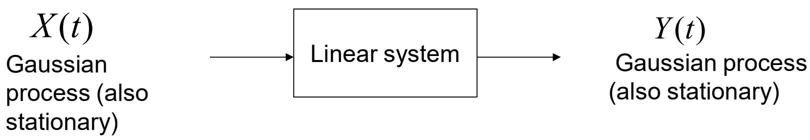
- For a complex process X(t) going through a compex-valued LTI system, it can be shown that

$$R_Y(\tau) = h(\tau) * [h^*(-\tau) * R_X(\tau)]$$

55

Gaussian Process Through LTI System

- If X(t) is a Gaussian process, then Y(t) is also a Gaussian process.
  - Gaussian processes are very important in communications.



56

Imperial College London

Lecture 3: Baseband and Passband Signals

Professor Geoffrey Li
Department of Electrical and Electronic Engineering
Imperial College London

$\sin\alpha\cos\beta = \frac{1}{2}\left(\sin(\alpha+\beta)+\sin(\alpha-\beta)\right)$
$\cos\alpha\sin\beta = \frac{1}{2}\left(\sin(\alpha+\beta)-\sin(\alpha-\beta)\right)$
$\cos\alpha\cos\beta = \frac{1}{2}\left(\cos(\alpha+\beta)+\cos(\alpha-\beta)\right)$
$\sin\alpha\sin\beta = -\frac{1}{2}\left(\cos(\alpha+\beta)-\cos(\alpha-\beta)\right)$

$\sin^2\theta = \frac{1}{2}(1-\cos 2\theta)$
$\cos^2\theta = \frac{1}{2}(1+\cos 2\theta)$

Energy and Power

Energy and power are important concepts in communication
--How much power is needed to transmit a signal?
--How is the signal-to-noise ratio found?
--How much interference do signals create for each other?

Energy:

$$E = \int_{-\infty}^{+\infty} |s(t)|^2\,dt = \int_{-\infty}^{+\infty} |S(f)|^2\,df$$

Power = time average of energy, computed over a large interval

$$P = \frac{1}{T}\int_{-T/2}^{+T/2} |s(t)|^2\,dt$$

Example: Power of a sinusoid $s(t) = A\cos(2\pi f_0 t + \theta)$

power of sin/cos

$$|s(t)|^2 = A^2\cos^2(2\pi f_0 t + \theta) = \frac{A^2}{2} + \frac{A^2}{2}\cos(4\pi f_0 t + 2\theta)$$

$$\boxed{P_s = \frac{A^2}{2}}$$

Average value

Integrates to zero over each period (averages to zero over large interval)

29

Fourier transform of a real signal is conjugate symmetric:

$x(t)$ real $\Rightarrow X(-\omega) = X^*(\omega)$.

Hence, for real signals, we only consider single sideband and one-sided bandwidth (positive half).

Bandwidth

- Bandwidth of a signal quantifies its frequency occupancy
- **One-sided bandwidth:** We only consider positive frequencies when computing bandwidth for *physical* (real) signals
  - For example, a WiFi signal may occupy a 20 MHz bandwidth, between 2.4-2.42 GHz
  - Physical signals are real-valued (in the time domain)
  - Hence they are conjugate symmetric in the frequency domain, so we can specify them completely by their spectrum over positive frequencies
- We shall also consider complex-valued (in the time domain) signals later
  - Complex envelope of a real-valued passband signal
  - The two-sided bandwidth of the complex envelope equals the physical (one-sided) bandwidth of the passband signal

Why Do We Need Negative Frequencies?

- We work with complex exponentials because they are eigenfunctions of LTI systems
  - Need complex exponentials at both positive and negative frequencies to span the space of square integrable signals
  - Real-valued sines and cosines with positive frequencies alone would also work, but they are not eigenfunctions of LTI systems, hence are less convenient
- Physical signals are real-valued (in time domain)
  - They must satisfy conjugate symmetry (all the information resides in either positive or negative frequencies, hence only need spectrum for one of these)
  - Hence physical bandwidth = one-sided bandwidth

Baseband communication is mainly limited to wired transmission.

Passband communication is popular for large bandwidth/long range/wireless.

Baseband and Passband Signals/Channels

- Channels often approximated as LTI systems
  - Signal passes through channel, and then noise is added
- Channels allocated/described typically in terms of frequency bands
  - Proper signal forms have to be designed for the corresponding frequency band

- Baseband channels/signals
  - Energy concentrated in a frequency band around DC
- Passband channels/signals
  - Energy concentrated in a frequency band away from DC
- Unified treatment of baseband and passband systems
  - Complex baseband representation of passband systems

Both baseband and passband signals can be complex ...

but complex passband signals do not have "equivalent" baseband representations (as the USB and LSB are different).

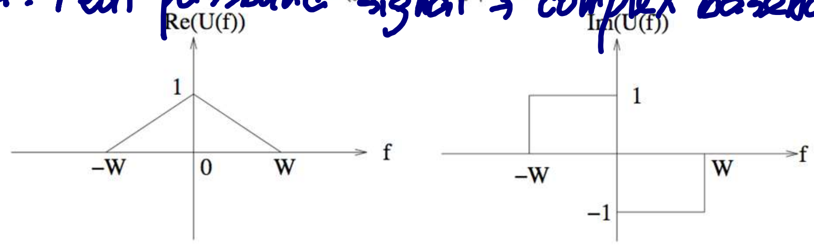
baseband channel: real baseband channel

passband channel: real passband signal → complex baseband channel

Baseband Signals

(Complex) baseband signals have energy/power concentrated in a band around DC.

$$U(f) \approx 0, \quad |f| \geq W$$



Real baseband signal: *U(f)* is conjugate-symmetric; therefore, |*U(f)*| is symmetric

Baseband signals are artificial and are usually complex.
Signals in physical world are always real (in the time domain).
There is an one-to-one correspondence between baseband and passband signals.

Communication over a physical passband channel (discussion coming up): complex-valued baseband signal, a convenient mathematical representation for the corresponding passband signals.
Communication over a physical baseband channel, we consider physical (real-valued) baseband signals.

Passband Signals

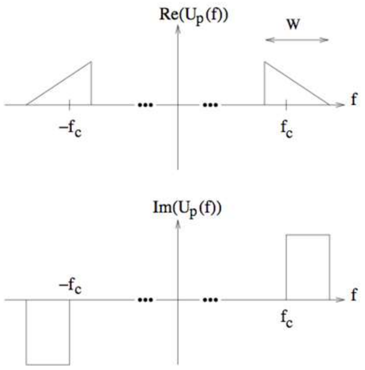
(Real-) passband signals have energy/power concentrated in a band away from DC.

$f_c = 0$ for baseband

$$U(f) \approx 0, \quad |f \pm f_c| > W \qquad f_c > W > 0$$

Carrier frequency

Bandwidth



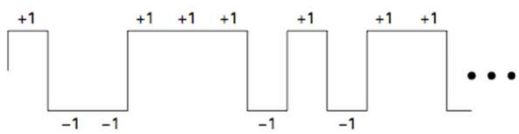
Passband signal of bandwidth W

We only consider physical (real-valued) passband signals, hence their spectra always obey conjugate symmetry

Examples of Baseband Signals (in terms of frequency)

Speech, audio are baseband signals

Two-level digital signal is baseband



We often want to send such signals over a passband channel (e.g., a 20 MHz WiFi channel at 2.4 GHz).

Need to understand how passband signals are structured in order to accomplish this.

Modulation: Baseband to Passband

Consider a real-valued baseband message signal *m(t)*

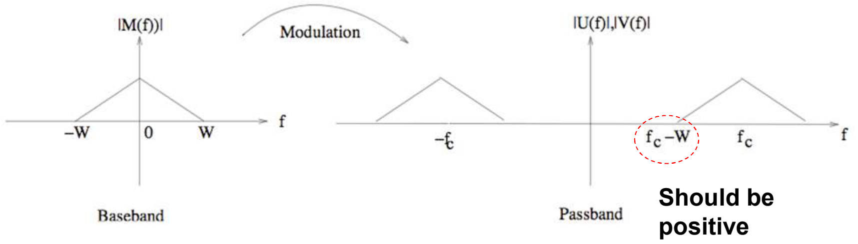
Modulation: Translate to passband by multiplying by a sinusoid $e^{j2\pi f_c t}$

$$u_p(t) = m(t)\cos 2\pi f_c t \leftrightarrow U_p(f) = \frac{1}{2}(M(f-f_c)+M(f+f_c))$$

so we need ×2 in upconversion or downconversion.

or

$$v_p(t) = m(t)\sin 2\pi f_c t \leftrightarrow V_p(f) = \frac{1}{2j}(M(f-f_c)-M(f+f_c))$$



Carrier frequency should be (usually much) bigger than the message bandwidth to keep away from DC!!!

$f_c \gg W$

I and Q Components

Can modulate separately using cosine and sine of carriers? Yes

Sinusoids are rapidly varying but predictable (contain no info)

$$u(t) = u_I(t)\cos(2\pi f_c t) - u_Q(t)\sin(2\pi f_c t)$$

Passband signal

In-phase (I) component

Quadrature (Q) component

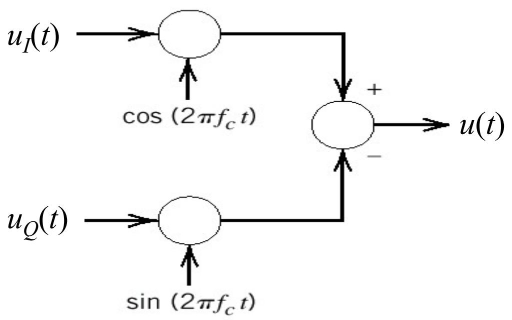
Real baseband signals (contain all the information)

2 real baseband signal
↓
1 complex baseband signal
↓
1 real passband signal

We can start from two real baseband signals and get a passband signal by I/Q modulation.

- How do we get back the I and Q components from the passband signal?
- Can any passband signal be decomposed into I and Q components?

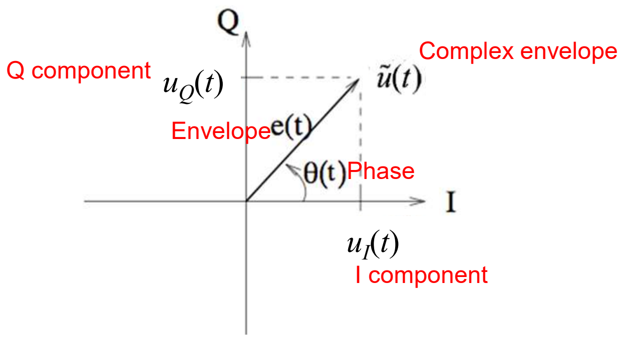
Upconversion: Baseband (low freq.) to Passband (high freq.)



- Block diagram follows directly from the equation defining the modulated signal.
- Happens at the transmitter

Baseband Signals (I)

- Passband signal can be mapped to a pair of real baseband signals (by real and imag parts of downconverted baseband signal)
- That is, passband modulation is *two-dimensional*
- We can also plot it on complex plane

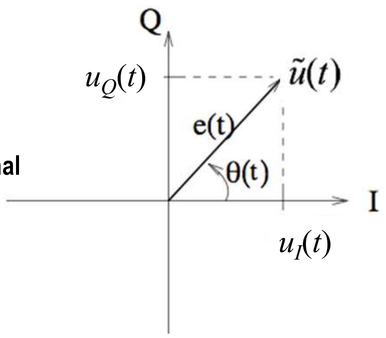


Three equivalent representations of the passband signal

- Rectangular coordinates: I and Q real + imag
- Polar coordinates: Envelope and phase mag ∠ phase
- Complex number: Complex envelope $Re\{\tilde{u}(t) e^{j2\pi f_c t}\}$

Baseband Signals (II)

Each representation of a complex envelope (baseband signal) corresponds to a time domain expression for the passband signal



$$\tilde{u}(t) = u_I(t) + j u_Q(t) = e(t) e^{j\theta(t)}$$ **Complex envelope**

$$e(t) = \sqrt{u_I^2(t) + u_Q^2(t)} \qquad \theta(t) = \tan^{-1}\frac{u_Q(t)}{u_I(t)}$$

$$u_I(t) = e(t)\cos\theta(t), \; u_Q(t) = e(t)\sin\theta(t)$$

Complex Envelope

$$u(t) = \mathrm{Re}\{\tilde{u}(t) e^{j2\pi f_c t}\}$$

Complex Envelope
Baseband Signal

Passband Signal
(Real-valued!)

- All information in a passband signal is contained in its complex envelope.
- Complex baseband representation can be defined for arbitrary f_c, as long as f_c>B

$$\tilde{u}(t) = u_I(t) + j u_Q(t) = e(t) e^{j\theta(t)}$$

$$e(t) = \sqrt{u_I^2(t) + u_Q^2(t)}, \quad \theta(t) = \tan^{-1} \frac{u_Q(t)}{u_I(t)}$$

Time Domain Expressions for a Passband Signal

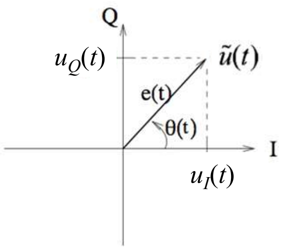
$j^2 = -1$

$$\boxed{u(t) = u_I(t)\cos(2\pi f_c t) - u_Q(t)\sin(2\pi f_c t)}$$ In I and Q components

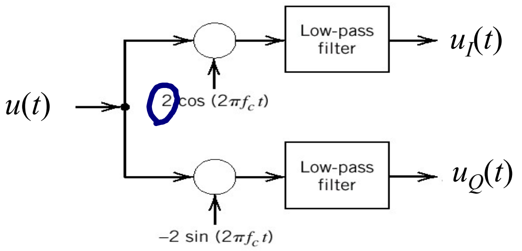
$$\boxed{u(t) = e(t)\cos(2\pi f_c t + \theta(t))}$$ In envelope and phase

$$\boxed{u(t) = \mathrm{Re}\{\tilde{u}(t) e^{j2\pi f_c t}\}}$$ In complex envelope

Starting from one representation, we can derive the rest based on the relations depicted in the figure



Downconversion: Passband to Baseband



Complex baseband signal

$$\tilde{u}(t) = u_I(t) + j u_Q(t) = e(t) e^{j\theta(t)}$$

Receiver needs to be coherent: coherent $\begin{cases} \text{carrier: frequency + phase} \\ \text{timing} \end{cases}$

- phase and frequency of the copy of the carrier at the receiver is same as that of the incoming signal)

Hilbert Transform (HT)

- Hilbert transform of a signal g(t) is defined as

$$\hat{g}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{g(\tau)}{t-\tau}d\tau$$

- HT is a linear transformation. Its inverse is given by

$$g(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\hat{g}(\tau)}{t-\tau}d\tau$$

- HT of $\hat{g}(t)$ is -g(t)
- In the frequency domain, we have

$$\hat{G}(f) = -j\,\mathrm{sgn}(f)G(f),$$

$$\mathrm{sgn}(f) = \begin{cases} +1 & f > 0 \\ 0 & f = 0 \\ -1 & f < 0 \end{cases}$$

HT introduces a phase shift of -90 degrees for all positive frequencies of the input signal, and +90 degrees for all negative frequencies. The amplitudes remain intact.

73

Pre-Envelope

- Define the pre-envelope of a signal u(t) as the complex-valued function

$$u_+(t) = u(t) + j\hat{u}(t)$$ $\hat{u}(t)$ is the HT of u(t).

- Its Fourier transform is

$$U_+(f) = U(f) + \mathrm{sgn}(f)U(f) = \begin{cases} 2U(f) & f > 0 \\ U(0) & f = 0 \\ 0 & f < 0 \end{cases}$$

- Pre-envelope removes the negative frequency components
- Similarly define the pre-envelope for negative frequencies:

$$u_-(t) = u(t) - j\hat{u}(t)$$

$$U_-(f) = U(f) - \mathrm{sgn}(f)U(f) = \begin{cases} 0 & f > 0 \\ U(0) & f = 0 \\ 2U(f) & f < 0 \end{cases}$$

74

From complex baseband to real passpand signals

Consider arbitrary complex-valued baseband signal $\tilde{u}(t)$, whose spectrum is limited to [-W,+W]. Define

$$u(t) = \mathrm{Re}\{\tilde{u}(t) e^{j2\pi f_c t}\}$$

Show that u(t) is a real-valued passband signal concentrated around $\pm f_c$.

- Let $u_+(t) = \tilde{u}(t) e^{j2\pi f_c t}$. Then $U_+(f) = \tilde{U}(f - f_c)$ in [f_c+W, f_c-W]

$$u(t) = \mathrm{Re}\{u_+(t)\} = \frac{1}{2}\left(u_+(t) + u_+^*(t)\right) \leftrightarrow U(f) = \frac{1}{2}\left(U_+(f) + \underbrace{U_+^*(-f)}_{\text{in } [-f_c-W,\, -f_c+W]}\right)$$ conjugate symmetry

- u(t) is a real valued passband signal

From real passpand to complex baseband signals

Given any real passband signal u(t), let's find the baseband complex envelope.

- Consider $u_+(t) = u(t) + j\hat{u}(t)$ for which we have

$$U_+(f) = \begin{cases} 2U(f) & f > 0 \\ U(0) & f = 0 \\ 0 & f < 0 \end{cases}$$

Since u(t) is real, it has conjugate symmetry in the frequency domain. Therefore,

$$U(f) = \frac{1}{2}\left(U_+(f) + U_+^*(-f)\right)$$

$$U(f) = \begin{cases} \frac{1}{2} U_+(f) & f > 0 \\ \frac{1}{2} U_+^*(-f) & f \le 0 \end{cases}$$

- In the time domain

$$u(t) = \frac{1}{2}\left(u_+(t) + u_+^*(t)\right) = \mathrm{Re}\{u_+(t)\}$$

- Define complex envelope

$$\tilde{u}(t) = u_+(t) e^{-j2\pi f_c t} = u_I(t) + j u_Q(t)$$

$u_+(t)$ is concentrated around f_c ; $\tilde{u}(t)$ is obtained by shifting the spectrum of $u_+(t)$ to left by f_c ; hence, it is a baseband signal.

Imperial College London

Lecture 4: Noise

Professor Geoffrey Li
Department of Electrical and Electronic Engineering
Imperial College London

77

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

78

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

79

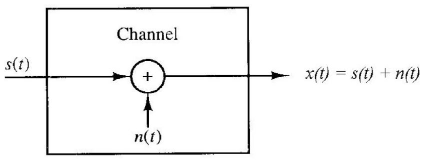
Noise

- Noise is unwanted 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 zero-mean **Gaussian distributions** (following from the central limit theorem!).

80

White Noise

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



  - Its power spectral density (PSD) is constant over **all frequencies**,

Two-sided PSD applies to general complex signals (two-sided bandwidth)

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

  - Factor ½ indicates 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 is only defined for stationary noise.

- An infinite bandwidth is a purely theoretic assumption:
  - Valid as long as noise PSD is flat over the bandwidth of interest.

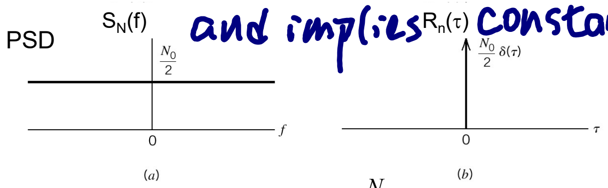
81

Noise is a random process:
- properties in ensembles, e.g. Gaussian
- properties in time, e.g. white.

White and Gaussian Noise

- White noise

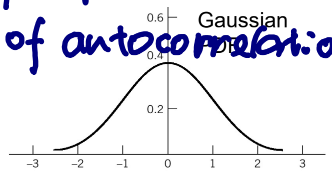
"white" means uncorrelated in time, and implies constant PSD.



  - Autocorrelation function of $n(t)$: $R_n(\tau) = \frac{N_0}{2}\delta(\tau)$
  - Samples at different time instants are uncorrelated.

PDF: sample probability distribution at specific time.
PSD: power distribution in frequency (FT of autocorrelation)

- Gaussian noise: the distribution at any time instant is Gaussian
  - Gaussian noise can be colored



- White noise ≠ Gaussian noise
  - White noise can be non-Gaussian
- In communications, we typically assume additive white Gaussian noise (AWGN)

82

Examples:
- white but non-Gaussian: any statistically uncorrelated sequence
- Gaussian but non-white: non-zero correlation in time.

Ideal Low-pass White Noise

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

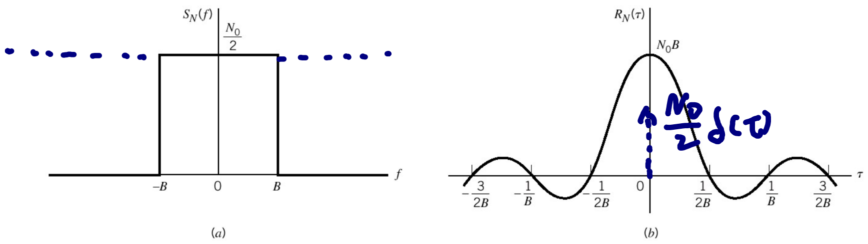
$$S_N(f) = \begin{cases} \frac{N_0}{2}, & |f| \le B \\ 0, & \text{otherwise} \end{cases} \qquad \text{Power } P_N = N_0 B$$

- By Einstein-Wiener-Khinchine relation, autocorrelation function

$$R_n(\tau) = E[n(t)n(t+\tau)] = N_0 B \,\text{sinc}(2\pi B\tau)$$

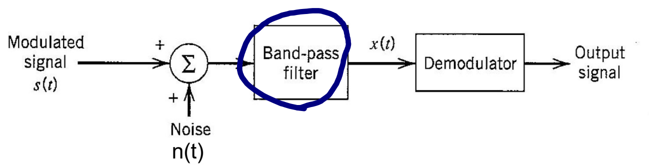
- Samples at Nyquist frequency $2B$ are uncorrelated instead of $\delta(t)$.

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

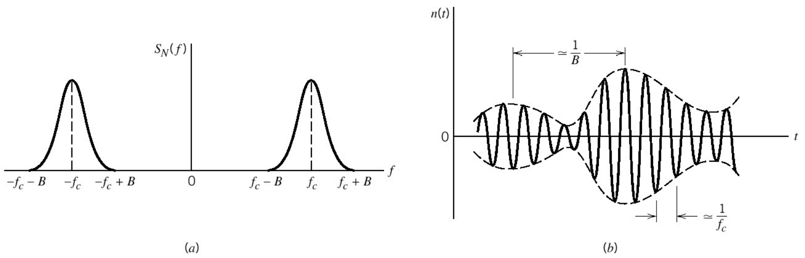


Band-pass Noise

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



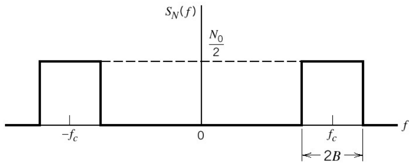
- Any noise that enters the receiver will therefore be band-pass in nature: its spectral magnitude is non-zero only for some band concentrated around the carrier frequency f_c (sometimes called **narrowband noise**).



Example

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

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



Power $P_N = 2N_0B$

- Autocorrelation function

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

  - follows from 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:

$$R_n(\tau) = 0, \text{ for } \tau = k/(2B),\ k = 1, 2, \ldots$$

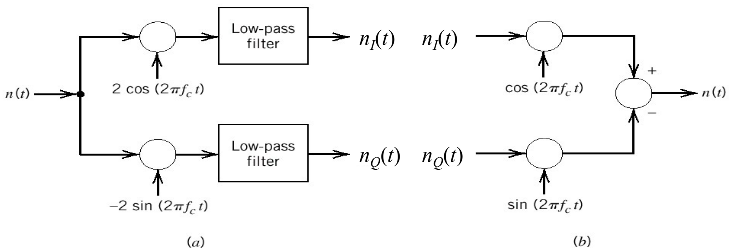
85

Band-pass Noise

- n(t) in canonical form:

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

- $n_I(t)$ and $n_Q(t)$ are fully representative of the band-pass noise.
  - Given band-pass noise, one may extract in-phase and quadrature components (using LPF of bandwidth B).
  - Given the two components, one may generate band-pass noise. This is useful in computer simulation.



86

Properties of Baseband Noise

- If noise $n(t)$ has zero mean, then so do $n_I(t)$ and $n_Q(t)$.
- If noise $n(t)$ is Gaussian, then so are $n_I(t)$ and $n_Q(t)$.
- $n_I(t)$ and $n_Q(t)$ have the same variance (i.e., same power) as $n(t)$
- Both in-phase and quadrature components have the same PSD:

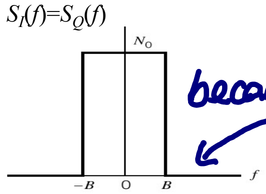
$$S_{N_I}(f) = S_{N_Q}(f) = \begin{cases} S_N(f-f_c)+S_N(f+f_c), & |f|\le B \\ 0, & \text{otherwise} \end{cases}$$

Noise Power

- For **ideally filtered narrowband noise**, the PSDs of $n_I(t)$ and $n_Q(t)$ are therefore given by

$$\boxed{S_I(f) = S_Q(f) = \begin{cases} N_0, & |f|\le B \\ 0, & \text{otherwise} \end{cases}}$$

$S_I(f)$=$S_Q(f)$



because 2B by LPF.

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

$$P_{N_I} = P_{N_Q} = 2N_0B$$

Phasor Representation

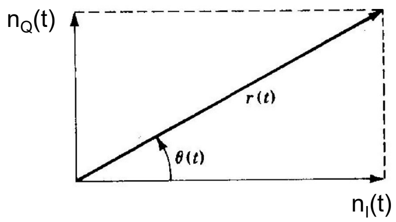
- We may write band-pass noise in the alternative form:

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

– $r(t) = \sqrt{n_I(t)^2 + n_Q(t)^2}$: the envelop of the noise

$f(r) = \frac{r}{\sigma^2} e^{\frac{r^2}{2\sigma^2}}$

– $\varphi(t) = \tan^{-1}\left(\frac{n_Q(t)}{n_I(t)}\right)$: the phase of the noise



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

89

Distribution of Envelop and Phase

It can be shown that if $n_I(t)$ and $n_Q(t)$ are independent Gaussian-distributed with zero mean and the same variance, what are the distributions of the magnitude $r(t)$ and the phase $\phi(t)$?

Denote: $n_I(t) \to x;\ n_Q(t) \to y;\ r(t) \to r;\ \phi(t) \to \theta,$

then $r = \sqrt{x^2+y^2}, (r \geq 0);\ \ \theta = \mathrm{Tan}^{-1}\frac{y}{x}, \theta \in [-\pi, \pi]$, or $x = r\cos(\theta), y = r\sin(\theta)$

$$J = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin(\theta) \\ \sin\theta & r\cos(\theta) \end{pmatrix},\ \det(J) = |J| = r$$

$$f_{n_I(t)}(x) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, f_{n_Q(t)}(y) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{y^2}{2\sigma^2}\right\},$$

$$f_{n_I(t)n_Q(t)}(x,y) = f_{n_I(t)}(x) f_{Q(t)}(x) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}$$

$$f_{r(t)\phi(t)}(r,\theta) = f_{n_I(t)n_Q(t)}(x,y)|\det(J)| = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} r = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{r^2}{2\sigma^2}\right\} r$$

$$f_{r(t)}(r) = \int_{-\pi}^{\pi} f_{r(t)\phi(t)}(r,\theta)d\theta = \frac{1}{\sigma^2}\exp\left\{-\frac{r^2}{2\sigma^2}\right\} r = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}},\ \ f_{\phi(t)}(\theta) = \int_0^{+\infty} f_{r(t)\phi(t)}(r,\theta)dr = \frac{1}{2\pi}$$

$f_{r(t)}(r) f_{\phi(t)}(\theta) = f_{r(t)\phi(t)}(r,\theta)$, then $r(t), \phi(t)$ are independent

90

Distribution of Envelop and Phase

- It can be shown that if $n_I(t)$ and $n_Q(t)$ are independent Gaussian-distributed with zero mean and the same variance, then the magnitude $r(t)$ has a **Rayleigh** distribution, the phase $\phi(t)$ is **uniformly** distributed, and they are independent.

- What if a sinusoid $A\cos(2\pi f_c t)$ is mixed with noise? What is the distribution of the magnitude? How about the phase?

Mixed signal:

A WGN.

$$A\cos(2\pi\omega_c t) + n_I(t)\cos(2\pi\omega_c t) - n_Q(t)\sin(2\pi\omega_c t)$$
$$= (A + n_I(t))\cos(2\pi\omega_c t) - n_Q(t)\sin(2\pi\omega_c t)$$

Magnitude:

$$\mathrm{r}(t) = \sqrt{\left(A + n_I(t)\right)^2 + n_Q^2(t)}$$

Distribution of Envelop and Phase

If a sinusoid $A\cos(2\pi f_c t)$ is mixed with noise?

Denote: $n_I(t) \to x;\ n_Q(t) \to y;\ r(t) \to r; \phi(t) \to \theta,$

then $r = \sqrt{(A+x)^2 + y^2}, (r \geq 0);\ \ \theta = \mathrm{Tan}^{-1}\frac{y}{A+x}, \theta \in [-\pi, \pi],$

or $A + x = r\cos(\theta) \to\ x = r\cos(\theta) - A, y = r\sin(\theta)$, $\det(J) = |J| = r$

$$f_{r(t)\phi(t)}(r,\theta) = f_{n_I(t)n_Q(t)}(x,y)|\det(J)| = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} r$$

$$= \frac{r}{2\pi\sigma^2}\exp\left\{-\frac{(r\cos\theta - A)^2 + (r\sin\theta)^2}{2\sigma^2}\right\} = \frac{r}{2\pi\sigma^2}\exp\left\{-\frac{A^2 + r^2 - 2Ar\cos\theta}{2\sigma^2}\right\}$$

$$f_{r(t)}(r) = \int_{-\pi}^{\pi} f_{r(t)\phi(t)}(r,\theta)d\theta = \frac{r}{\sigma^2}\exp\left\{-\frac{A^2+r^2}{2\sigma^2}\right\}\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\left\{\frac{Ar\cos\theta}{\sigma^2}\right\}d\theta$$

$$= \frac{r}{\sigma^2}\exp\left\{-\frac{A^2+r^2}{2\sigma^2}\right\} I_0\left(\frac{Ar}{\sigma^2}\right)$$ (Rice distribution)

$I_0\left(\frac{Ar}{\sigma^2}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\left\{\frac{Ar\cos\theta}{\sigma^2}\right\}d\theta$: the modified Bessel function of the first kind of order zero

$$r=\sqrt{(A+x)^2+(B+y)^2},\quad \theta=\tan^{-1}\frac{B+y}{A+x}\qquad \begin{cases} x = r\cos\theta - A \\ y = r\sin\theta - B \end{cases}$$

$$f_{r(t)\theta(t)}(r,\theta) = f_{n_I(t)n_Q(t)}(x,y)\,|\det(J)|$$

$$= \frac{r}{2\pi\sigma^2}\exp\left\{-\frac{(r\cos\theta-A)^2+(r\sin\theta-B)^2}{2\sigma^2}\right\}$$

$$= \frac{r}{2\pi\sigma^2}\exp\left\{-\frac{r^2+A^2+B^2-2Ar\cos\theta-2Br\sin\theta}{2\sigma^2}\right\}$$

Distribution of Envelop and Phase

- It can be shown that if $n_I(t)$ and $n_Q(t)$ are independent Gaussian distributed with zero mean and the same variance, then the magnitude $r(t)$ has a **Rayleigh** distribution, the phase $\phi(t)$ is **uniformly** distributed, and they are independent.
- What if a sinusoid $A\cos(2\pi f_c t)$ is mixed with noise?

  Mixed signal:

  $$A\cos(2\pi\omega_c t) + n_I(t)\cos(2\pi\omega_c t) - n_Q(t)\sin(2\pi\omega_c t)$$
  $$= (A + n_I(t))\cos(2\pi\omega_c t) - n_Q(t)\sin(2\pi\omega_c t)$$

  Magnitude:

  $$r(t) = \sqrt{\left(A+n_I(t)\right)^2 + n_Q^2(t)}$$

  the magnitude will have a **Rice** distribution!

- What is the pdf of $r(t) = \sqrt{\left(A_I+n_I(t)\right)^2 + \left(A_Q+n_Q(t)\right)^2}$ if A_I, A_Q are constant.

  → also follows Rice distribution

One point bonus for the first 5 students who submit the correct answer!

93

$$f_{r(t)}(r) = \frac{r}{2\pi\sigma^2}\exp\left\{-\frac{r^2+A^2+B^2}{2\sigma^2}\right\}\int_{-\pi}^{\pi}\exp\left\{\frac{r(A\cos\theta+B\sin\theta)}{\sigma^2}\right\}d\theta$$

$$\phi=\tan^{-1}\frac{B}{A}\qquad \frac{r}{\sigma^2}\exp\left\{-\frac{r^2+A^2+B^2}{2\sigma^2}\right\}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\left\{\frac{r\sqrt{A^2+B^2}\cos(\theta-\phi)}{\sigma^2}\right\}d\theta\right\}$$

$$= \frac{r}{\sigma^2}\exp\left\{-\frac{r^2+A^2+B^2}{2\sigma^2}\right\} I_0\left(\frac{r\sqrt{A^2+B^2}}{\sigma^2}\right)$$

Summary

- White noise: PSD is constant over an infinite bandwidth.
- Gaussian noise: PDF is Gaussian.
- Bandpass noise
  - In-phase and quadrature components $n_I(t)$ and $n_Q(t)$ are low-pass random processes.
  - $n_I(t)$ and $n_Q(t)$ have the same PSD.
  - $n_I(t)$ and $n_Q(t)$ 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.

94

Imperial College London

Lecture 5: Noise Performance of DSB

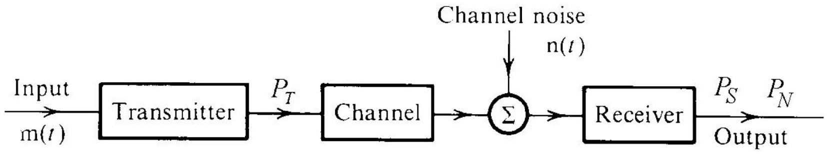
Professor Geoffrey Li
Department of Electrical and Electronic Engineering
Imperial College London

Outline

- SNR of baseband analogue transmission
- DSB-SC – a 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

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

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)
  - Managing the wide range of power levels in communication systems
  - Example:
    - ratio of 2 → 3 dB; 4 → 6 dB; 10 → 10dB

Transmitted Power

- P_T: transmitted power
- Limited by: equipment capability, battery life, cost, government restrictions, interference with other channels,…
- 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 and compare the *SNR* values we obtain.

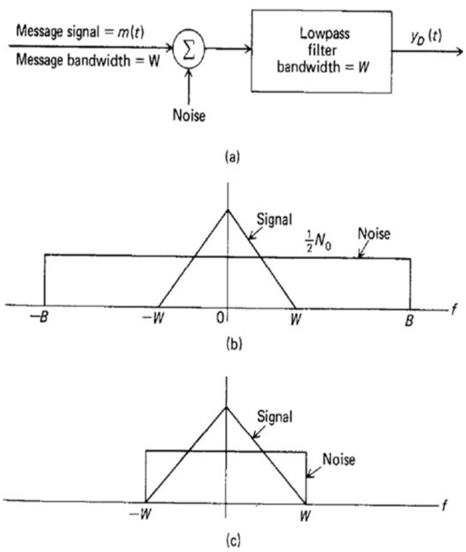
99

A Baseband Communication System

- It does not use modulation
- It is suitable for transmission over wires
- Transmit power is identical to message power:

  $P_T = P$
- If the unit channel gain or no propagation loss, then

  $P_S = P_T = P$
- Results can be extended to band-pass systems



100

Baseband SNR

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

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

- Improve SNR by:
  - increasing the transmitted power ($P_T \uparrow$),
  - restricting the message bandwidth ($W \downarrow$),
  - making the channel/receiver less noisy ($N_0 \downarrow$).

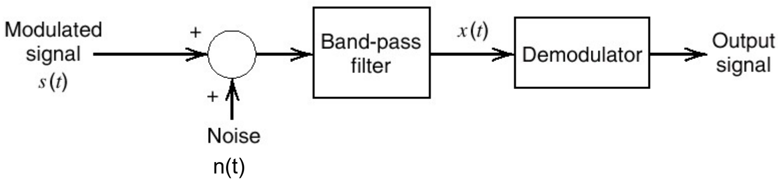
101

Double Sideband-Suppressed Carrier (DSB-SC) Modulation

- General form of a DSB-SC signal (suppressed carrier AM):

$$s(t) = m(t) A \cos(2\pi f_c t)$$

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

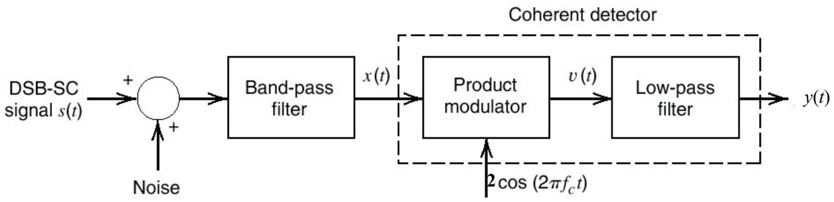


102

DSB-SC Receiver with Noise

- Synchronous detection= Product detection = Coherent detection

  (detection and demodulation are used interchangeably)



- Received (and filtered) noisy signal is

noise in sin branch is ignored

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

LPF

$$[Am(t)+n_I(t)](1+\cancel{\cos(4\pi f_c t)}) - n_Q(t)\,\cancel{\sin(4\pi f_c t)}$$

Synchronous Detection for DSB-SC

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

  $v(t) = 2\cos(2\pi f_c t)\, x(t)$

$2\cos^2\theta = 1 + \cos(2\theta)$

$2\sin\theta\cos\theta = \sin(2\theta)$

$$= 2\cos(2\pi f_c t)\left[Am(t)\cos(2\pi f_c t) + n_I(t)\cos(2\pi f_c t) - n_Q(\mathrm{t})\sin(2\pi f_c t)\right]$$

$$= Am(t) + n_I(t) + n_I(t)\cos(4\pi f_c t) - n_Q(t)\sin(4\pi f_c t)$$

- Use a LPF to keep $y(t) = Am(t) + n_I(t)$ } Quadrature noise component completely cleaned!

- 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 $n_I(t)$:

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

Comparison

- *SNR* at the receiver output:

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

- **Transmit power:**

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

Tx power : $\frac{A^2 P}{2}$

Rx power = $A^2 P$

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

- Comparison with

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

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

105

Imperial College London

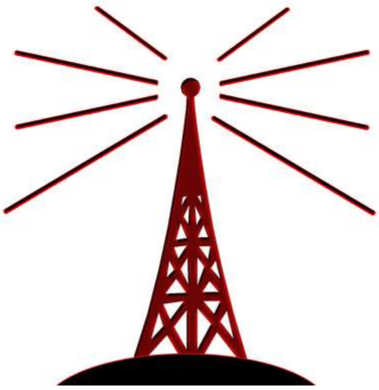
Lecture 6: Noise Performance of SSB and Conventional AM

Professor Geoffrey Li
Department of Electrical and Electronic Engineering
Imperial College London

106

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

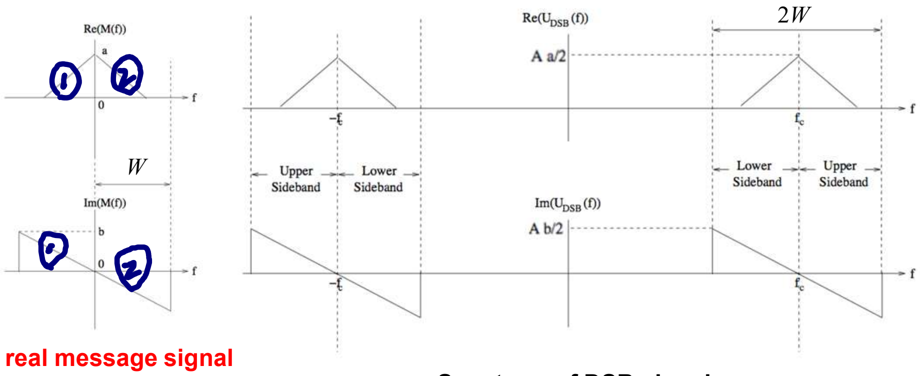


107

real signal: spectrum conjugate symmetry

DSB → SSB

only need 1 or 2

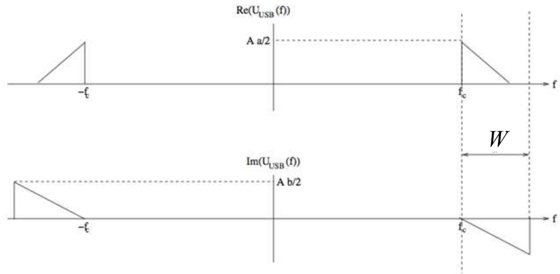


real message signal
complex conjugate symmetric spectrum

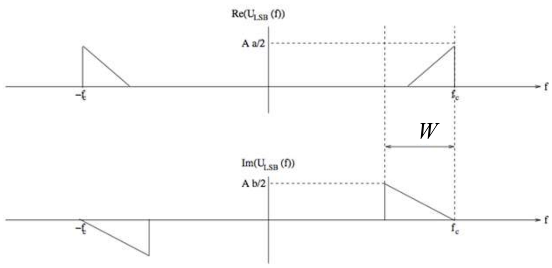
Spectrum of DSB signal

One side band is enough to reconstruct the message!

Single sideband (SSB)



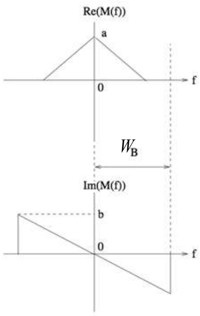
(a) Upper Sideband Signaling

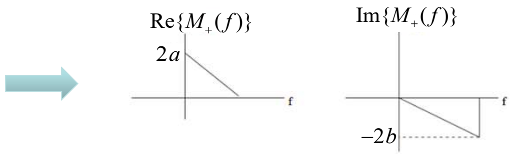


(b) Lower Sideband Signaling

Message can be recovered by moving SSB components left and right by f_c, and low pass filtering (just like DSB).

SSB Modulated Signal





$$m(t) \Leftrightarrow M(f)$$

$$\hat{m}(t) \Leftrightarrow \hat{M}(f) = -j\,\mathrm{sgn}(f)M(f)$$

$\hat{m}(t)$ Hilbert Transform of $m(t)$

$$M_+(f) = \begin{cases} 2M(f), & f > 0 \\ M(0), & f = 0 \\ 0, & f < 0 \end{cases}$$

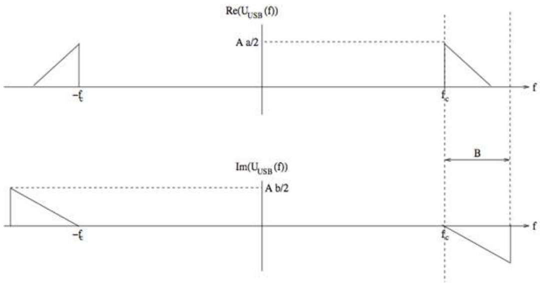
$$= M(f) + j[-j\,\mathrm{sgn}(f)M(f)]$$

$$= M(f) + j\hat{M}(f)$$

Analytic signal $m_+(t)$ removes negative frequency components.

$$m_+(t) = m(t) + j\hat{m}(t)$$

$$m_-(t) = m(t) - j\hat{m}(t)$$



(a) Upper Sideband Signaling

$$s(t) = \frac{A}{4} m_+(t) e^{j2\pi f_c t} + \frac{A}{4} m_-(t) e^{-j2\pi f_c t}$$

SSB signal

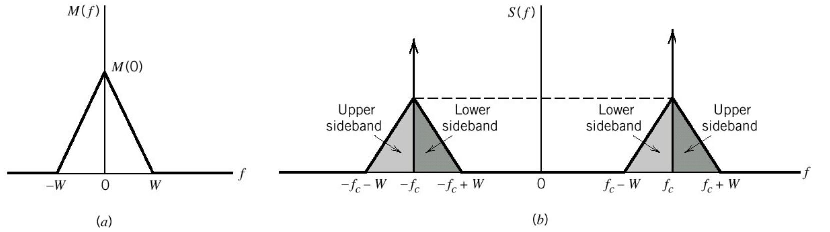
SSB modulation: $\begin{cases} u_I(t) \leftarrow m(t) \\ u_Q(t) \leftarrow \hat{m}(t) \end{cases}$

SSB Modulation

- I component: the message, Q component: its Hilbert transform

$$
\begin{aligned}
s(t) &= \frac{A}{4} m_+(t) e^{j2\pi f_c t} + \frac{A}{4} m_-(t) e^{-j2\pi f_c t} \\
&= \frac{A}{4}\left(m(t) + j\hat{m}(t)\right) e^{j2\pi f_c t} + \frac{A}{4}(m(t) - j\hat{m}(t)) e^{-j2\pi f_c t} \\
&= \frac{A}{2} m(t)\cos\left(2\pi f_c t\right) - \frac{A}{2}\hat{m}(t)\sin\left(2\pi f_c t\right)
\end{aligned}
$$

- $\hat{m}(t)$ and $m(t)$ have the same power P. $\frac{A^2P}{8}$ $\frac{A^2P}{8}$
- 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 $n_I(t)$ the lower-sideband in-phase noise (different from the double-sideband noise in DSB).
- Using coherent detection:

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

- After low-pass filtering,

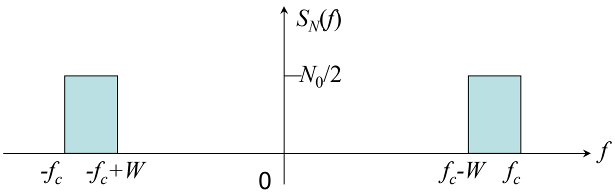
$$
y(t) = \frac{A}{2} m(t) + n_I(t)
$$

112

Noise Power

- Noise power of $n_l(t)$ = noise power of band-pass noise

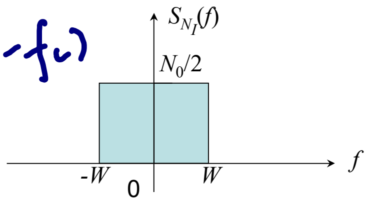
  = N_0W (halved compared to DSB)



Lower-sideband noise

Challenges:
① extract f_c
② PAPR

$S_{N_l}(f) = S_N(f+f_c) + S_N(f-f_c)$



Baseband noise

113

Output SNR

- Signal power $A^2P/4$

- SNR at output

$$\boxed{SNR_{SSB} = \frac{A^2P}{4N_0W}}$$

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

$$SNR_{baseband} = \frac{A^2P}{4N_0W}$$

- Conclusion: SSB achieves the same SNR performance as DSB-SC (and the baseband model),

but, only requires half the bandwidth!

114

Standard AM, DSB, and SSB

- **Standard AM**:

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

Usually $A \geq \mathrm{m_p} = \max|m(t)|$

Modulation Index $\mu = \frac{m_p}{A} \leq 1$

- **DSB**:

$$s_{DSB}(t) = m(t)\mathrm{A}\cos(2\pi f_c t)$$

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

115

Standard AM: Synchronous Detection

- Pre-detection signal:

$$\begin{aligned} x(t) &= [A + m(t)]\cos(2\pi f_c t) + n(t) \\ &= [A + m(t)]\cos(2\pi f_c t) + n_I(t)\cos(2\pi f_c t) - n_Q(t)\sin(2\pi f_c t) \\ &= [A + m(t) + n_I(t)]\cos(2\pi f_c t) - n_Q(t)\sin(2\pi f_c t) \end{aligned}$$

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

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

- LPF

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

116

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

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

117

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:

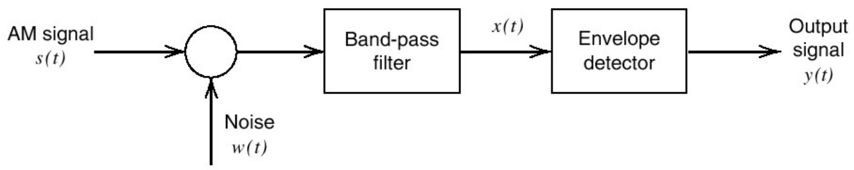
$$\frac{P}{A^2 + P} < 1$$

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

118

Non-coherent Receiver

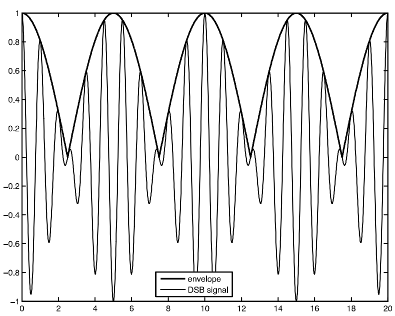
- The envelope (or magnitude of complex envelope) does not depend on carrier phase
- Suppose we can extract the envelope of a passband signal
  - Does not require carrier sync
- Can we recover the message?

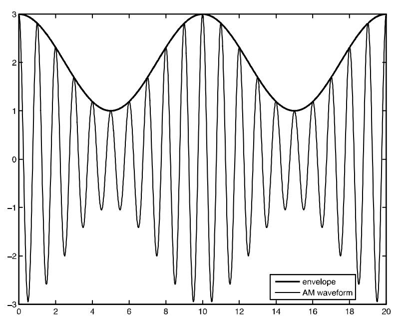


Model of AM envelope detector

What does the envelope tell us?

Example: sinusoidal message waveform

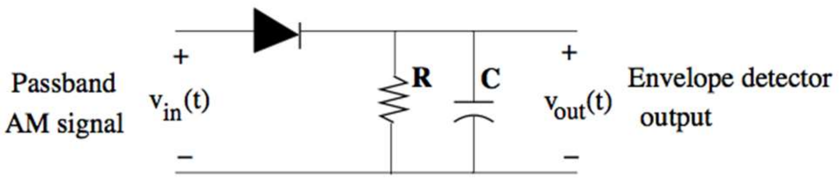




DSB modulated signal
Envelope = message magnitude
➔ Envelope detection loses message sign

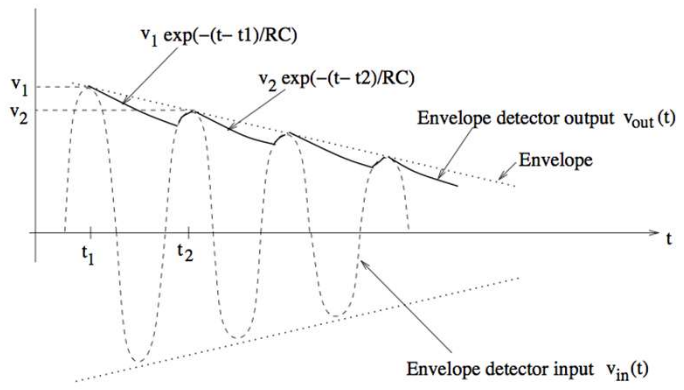
DSB signal + strong carrier component
Envelope = message + DC
➔Envelope detector + DC block recovers message info

How do we do envelope detection?



Positive carrier cycle ➔ capacitor charges up (reaches value of envelope)
Negative carrier cycle ➔ capacitor discharges with RC time constant

Envelope detector operation

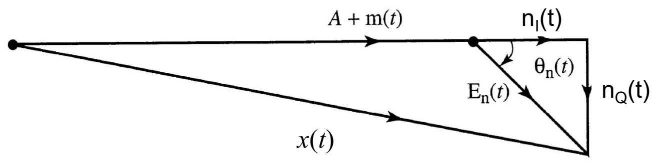


- Positive carrier cycle ➔ capacitor charges up (reaches value of envelope)
- Negative carrier cycle ➔ capacitor discharges with RC time constant
- Should not discharge too fast during negative cycle
- Should react fast enough to follow variations in envelope (which depend on message bandwidth B)

$$\frac{1}{f_c} \ll RC \ll \frac{1}{B}$$

Envelope Detection for Standard AM

- Phasor diagram of the signals present at an AM receiver



- Envel

$$y(t) = \text{envelope of } x(t)$$

- Equation is too complicated $\sqrt{[A + m(t) + n_I(t)]^2 + n_Q^2(t)}$
- Consider limiting cases to put noise in an additive form

Small Noise Case

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

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

- Then

$$n_Q(t) << [A + m(t) + n_I(t)]$$

- Then

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

- Thus

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

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

- And in terms of baseband SNR:

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

- **Valid for small noise only!**

Large Noise Case: Threshold Effect

- Information is lost!
- **Threshold effect**: below some carrier-to-noise ratio level (very low *A*), performance of envelope detector deteriorates very rapidly (not the case in coherent detection)

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

Lecture 7: Frequency Modulation (FM)

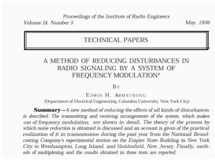
Professor Geoffrey Li
Department of Electrical and Electronic Engineering
Imperial College London

Outline

- Recap of FM
- FM system model in noise
- PSD of noise
- Derivation of FM output SNR
- References
  - Notes of Communication Systems, Chap. 3.4.1-3.4.2.
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 6

Invention of FM

- Edwin Howard Armstrong invented wideband FM
- He patented the regenerative circuit in 1914 (later challenged in court), the superheterodyne receiver in 1918 and the super-regenerative circuit in 1922.
- Presented "**A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation**", (first paper to describer FM radio) before the New York section of the Institute of Radio Engineers (now IEEE) on 6 Nov. 1935 (published in 1936).



Proceedings of the Institute of Radio Engineers

Volume 24, Number 5 — May, 1936

TECHNICAL PAPERS

A METHOD OF REDUCING DISTURBANCES IN RADIO SIGNALING BY A SYSTEM OF FREQUENCY MODULATION*

By

EDWIN H. ARMSTRONG

(Department of Electrical Engineering, Columbia University, New York City)

Summary—*A new method of reducing the effects of all kinds of disturbances is described. The transmitting and receiving arrangements of the system, which makes use of frequency modulation, are shown in detail. The theory of the process by which noise reduction is obtained is discussed and an account is given of the practical realization of it in transmissions during the past year from the National Broadcasting Company's experimental station on the Empire State Building in New York City to Westhampton, Long Island, and Haddonfield, New Jersey. Finally, methods of multiplexing and the results obtained in these tests are reported.*

"Several engineers said after the demonstration that they consider Dr. Armstrong's invention one of the most important radio developments since the first earphone crystal sets were introduced."

- Pushed by RCA, FM band shifted to 88–108 MHz in 1945
- RCA offered Armstrong $1,000,000 for a non-exclusive, royalty-free license to use his FM patents. He refused to protect other licensed companies, which would have to pay 2% royalties
- Committed suicide in 1954

129

FM vs. AM

- Fundamental difference between AM and FM:
- AM: information contained in the signal **amplitude**

  ⇒ Additive noise: corrupts the modulated signal directly
- FM: information contained in the signal **frequency**

  ⇒ effect of noise is determined by the extent it changes the frequency of the modulated signal
- Consequently, FM signal is affected less by noise than AM

130

Frequency and Phase

- A carrier waveform:

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

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

- When

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

we may say that

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

131

Frequency and Phase

- A carrier waveform

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

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

- When

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

we may say that

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

- Generalization: **instantaneous frequency** $f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt}$ $\theta_i(t) = 2\pi\int_0^t f_i(\tau)\,d\tau + \theta_i(0)$

$$\theta_i(t) = 2\pi\int_0^t f_i(\tau)\,d\tau + \theta_i(0)$$

132

FM: $s(t) = A\cos[2\pi f_c t + 2\pi k_f \int_0^{\tau} m(\tau) d\tau]$

PM: $s(t) = A\cos[2\pi f_c t + k_p m(t)]$

FM

- Instantaneous frequency is varied linearly with message:

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

  - k_f is the **frequency sensitivity** of the modulator.

- Hence (assuming $\theta_i(0)=0$):

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

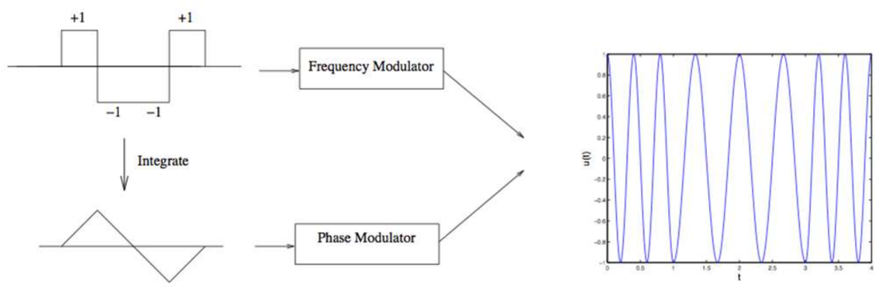
- Modulated signal: $s(t) = A\cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau) d\tau\right]$

- **Phase modulation** (PM): $s(t) = A\cos\left[2\pi f_c t + k_p m(t)\right]$

- Note:

133

PM – FM Equivalence



(a) Messages used for angle modulation (b) Angle modulated signal

- FM signal is equivalent to a PM signal where the modulating signal is $\int_0^t m(\tau) d\tau$
- Hence, properties of PM and FM modulation are similar
- **We focus on FM**

134

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

FM Basics

- $m_p = \max|m(t)|$: peak message amplitude

$$f_c - k_f m_p \le f_i(t) \le f_c + k_f m_p$$

- **frequency deviation:** deviation of f_i(t) from carrier frequency:

$$\Delta f = k_f m_p$$

$\Delta f = k_f m_p$

- **deviation ratio/modulation index:**

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

W: message bandwidth

$\beta = \frac{\Delta f}{W}$

Example: Sinusoidal message $m(t) = A_m \cos 2\pi f_m t$

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

$$\theta(t) = 2\pi k_f \int_0^t A_m \cos(2\pi f_m \tau)\, d\tau = \frac{A_m k_f}{f_m}\sin(2\pi f_m t)$$

$$\beta = \frac{A_m k_f}{f_m}$$

$$\theta(t) = \beta \sin 2\pi f_m t$$

135

Bandwidth of FM

- **deviation ratio/modulation index:**

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

  - *W*: message bandwidth
  - Small β: **narrow-band FM**
  - Large β: **wide-band FM**

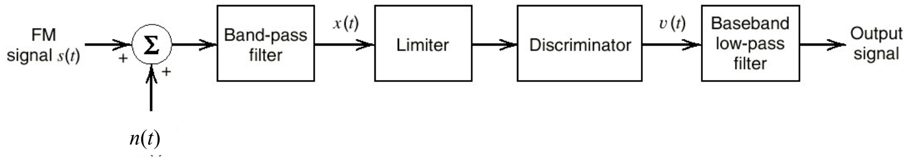
- **Carson's rule of thumb**: Transmission bandwidth of FM

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

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

136

FM Receiver



- **Bandpass filter**: removes signals outside bandwidth of $fc \pm B_T/2$
  - ⇒ 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 whose instantaneous amplitude is proportional to instantaneous frequency
  - ⇒ it recovers the message signal
- **Final baseband low-pass filter**: has a bandwidth of W
  - ⇒ removes out-of-band noise

137

Linear Argument at High SNR

- FM is nonlinear (modulation & demodulation): superposition does not hold
- We will show that for *high SNR*, noise and message signals are approximately independent of each other:

  Output ≈ Message + Noise (i.e., no other nonlinear terms)

138

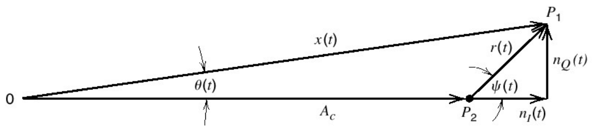
$$x(t) = A\cos(2\pi f_c t + \phi(t)) + r(t)\cos(2\pi f_c t + \phi(t) + \psi(t) - \phi(t))$$
$$= A\cos(2\pi f_c t + \phi(t)) + r(t)\left(\cos(2\pi f_c t + \phi(t))\cos(\psi(t) - \phi(t)) - \sin(2\pi f_c t + \phi(t))\sin(\psi(t) - \phi(t))\right)$$

Phase Noise in High SNR

$$x(t) = A\cos\left[2\pi f_c t + \phi(t)\right] + \{n_I(t)\cos(2\pi f_c t) + n_Q(t)\sin(2\pi f_c t)\}$$
$$= A\cos\left[2\pi f_c t + \phi(t)\right] + r(t)\cos\left[2\pi f_c t + \psi(t)\right]$$

$$\phi(t) = 2\pi k_f \int_0^t m(\tau)d\tau$$

Phasor diagram of the FM carrier and noise signals



- Instantaneous phase of the resultant phasor:

$$\theta(t) = \phi(t) + \tan^{-1}\left[\frac{r(t)\sin(\psi(t) - \phi(t))}{A + r(t)\cos(\psi(t) - \phi(t))}\right]$$

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

$$\theta(t) \cong \phi(t) + \frac{r(t)}{A}\sin(\psi(t) - \phi(t))$$

139

Phase Noise in High SNR

The discriminator output is

$$v(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} \cong k_f m(t) + n_d(t)$$ ➡ Noise is additive!

where the noise terms is

$$n_d(t) = \frac{1}{2\pi A}\frac{d}{dt}\left[r(t)\sin(\psi(t) - \phi(t))\right]$$

If Ψ(t) is uniformly distributed over [0,2π], then Ψ(t)-Φ(t) is also uniform over [0,2π]!!! ➡ $n_d(t)$ is independent of message!

Prove it, one point bonus!

$$n_d(t) \cong \frac{1}{2\pi A}\frac{d}{dt}\left[r(t)\sin\psi(t)\right] = \frac{1}{2\pi A}\frac{d}{dt}n_Q(t)$$

Additive noise $n_d(t)$ at the discriminator output is determined by the carrier amplitude A and the quadrature component $n_Q(t)$ of the narrowband noise.

140

Discriminator Output

Output Signal Power: $P_S = k_f^2 P$

P: average power of the message signal

Power of Noise?

- Noise term:

$$n_d(t) = \frac{1}{2\pi A}\frac{dn_Q(t)}{dt}$$

if $x(t) \leftrightarrow X(f)$

- From Fourier theory: $\frac{dx(t)}{dt} \leftrightarrow j2\pi f X(f)$

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

Noise PSD

- It follows 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

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

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

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

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

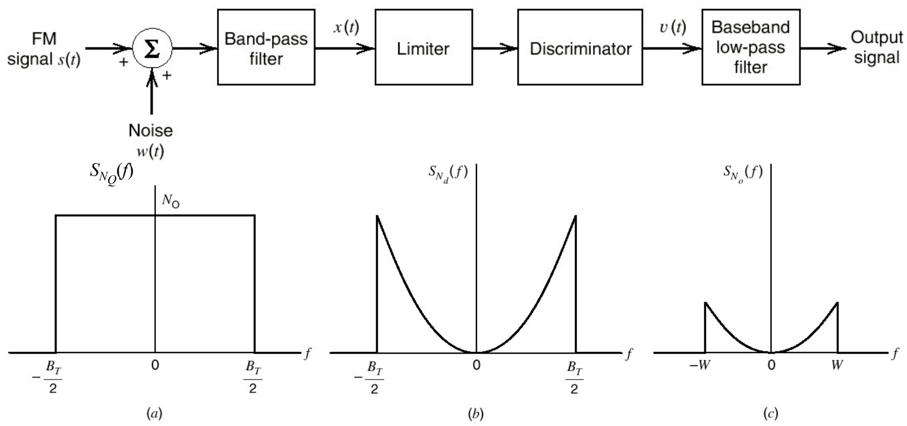
B_T: transmission bandwidth of the FM signal

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

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

For wideband FM, W is typically smaller than $B_T/2$

Noise PSD



(a) Power spectral density of quadrature component $n_Q(t)$ 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.

143

Noise Power

- Average noise power at the receiver output:

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

- Thus,

$$P_N = \int_{-W}^{W} \frac{f^2}{A^2} N_0 df = \frac{2N_0 W^3}{3A^2}$$

- Average noise power at the output of an FM receiver

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

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

144

Output SNR

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

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

- 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 carrier power is large compared to noise power

Bandwidth-SNR Tradeoff

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

In wideband FM, transmission bandwidth B_T is proportional to Δf.

At high SNR, increase in B_T ($\beta = \Delta f / W$) provides quadratic increase in output SNR.

Threshold effect

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

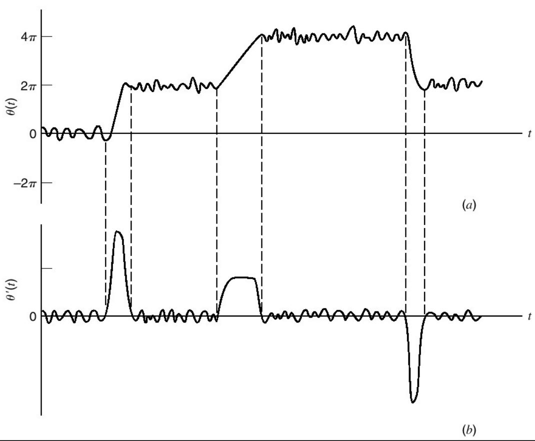
Carrier-to-noise ratio: $\rho = \dfrac{A^2}{2N_0 B_T}, \quad B_T = 2W(\beta+1)$

- Below the threshold (ρ<10) FM receiver breaks (i.e., significantly deteriorates).
- Analyzed by S. O. Rice (very complicated!), the noise in FM receiver is called "click noise" or "Rice noise"

147

Qualitative Discussion

- As noise changes randomly, 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

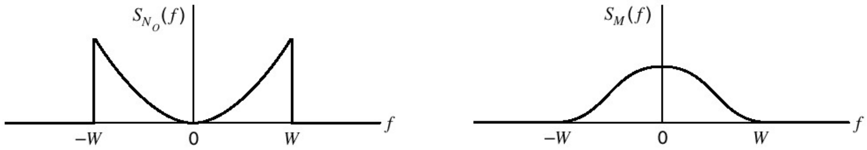


Number of clicks increases as ρ decreases

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

148

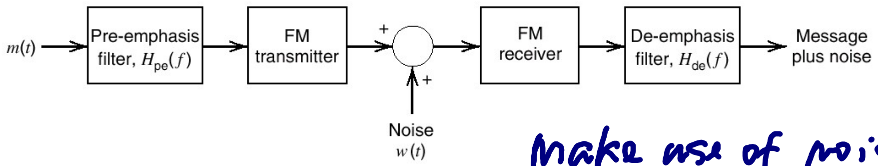
Improve Output SNR



- PSD of noise at detector output $\propto$ square of frequency.
- PSD of a message typically decays towards the ends of its band
- 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

149

Pre-emphasis and De-emphasis



Make use of noise PSD.

- $H_{pe}(f)$: artificially emphasizes high frequency components of the message prior to modulation (before noise is introduced).
- $H_{de}(f)$: de-emphasizes high frequency components at the receiver, and restore the original PSD of the message.
- In theory, $H_{pe}(f) \propto f$, $H_{de}(f) \propto 1/f$
- This can improve output SNR by around 13 dB.

150

Improvement Factor

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

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

- PSD 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

$$I = \frac{P_N \text{ without pre/de-emphasis}}{P_N \text{ with pre/de-emphasis}} = \frac{\frac{2N_0W^3}{3A^2}}{\int_{-W}^{W} \frac{f^2}{A^2} |H_{de}(f)|^2 N_0 df} = \frac{2W^3}{3\int_{-W}^{W} f^2 |H_{de}(f)|^2 df}$$

Example

- (a) Pre-emphasis filter

$$H_{pe}(f) \cong 1 + j\frac{f}{f_0}$$

$$f_0 = 1/(2\pi rC),$$

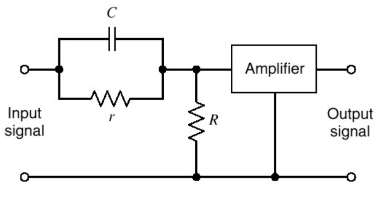
- (b) De-emphasis filter

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

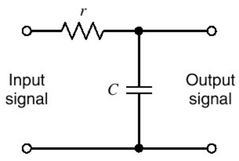
- Improvement

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

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



(a)



(b)

Comparison of Analog Systems

- Assumptions:
  - single-tone modulation, $m(t) = A_m \cos(2\pi f_m t)$
  - message bandwidth $W = f_m$
  - for AM system, modulation index, $\mu = \frac{m_p}{A} = 1$, $m_p = \max|m(t)|$
  - For FM system, modulation index, $\beta = \frac{\Delta f}{W} = 5$, $\Delta f = k_f m_p$ (used in commercial FM transmission, with $\Delta f = 75$ kHz, and $W = 15$ kHz)
- SNR expressions for various modulation schemes:

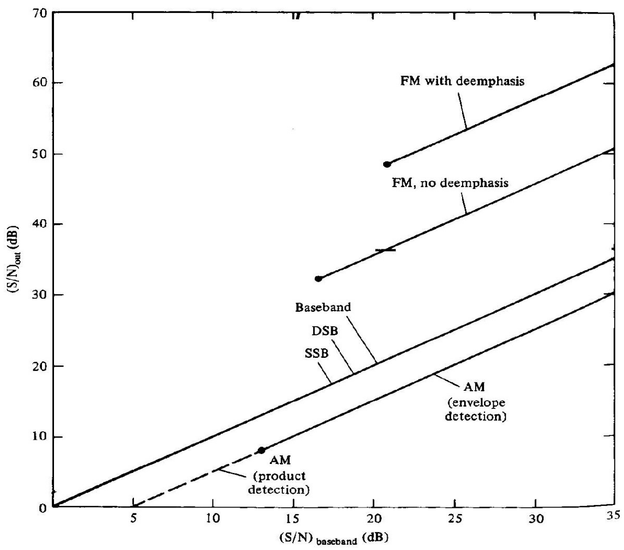
$$SNR_{DSB-SC} = SNR_{baseband} = SNR_{SSB}$$

$$SNR_{AM} = \frac{\mu^2}{2+\mu^2} SNR_{baseband} = \frac{1}{3} SNR_{baseband}$$

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

without pre/de-emphasis

Performance of Analog Systems



Conclusions

- (Full) AM: SNR performance is 4.8 dB worse than a baseband system, transmission bandwidth is $B_T = 2W$
- DSB: SNR performance is identical to a baseband system, transmission bandwidth is $B_T = 2W$
- SSB: SNR performance is again identical, transmission bandwidth is only $B_T = W$
- FM: SNR performance is 15.7 dB better than a baseband system, transmission bandwidth is $B_T = 2(\beta + 1)W = 12W$

  (with pre- and de-emphasis SNR performance is increased by about 13 dB with the same transmission bandwidth).

155

Imperial College London

Lecture 8: Digital Representation of Signals

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

Imperial College London

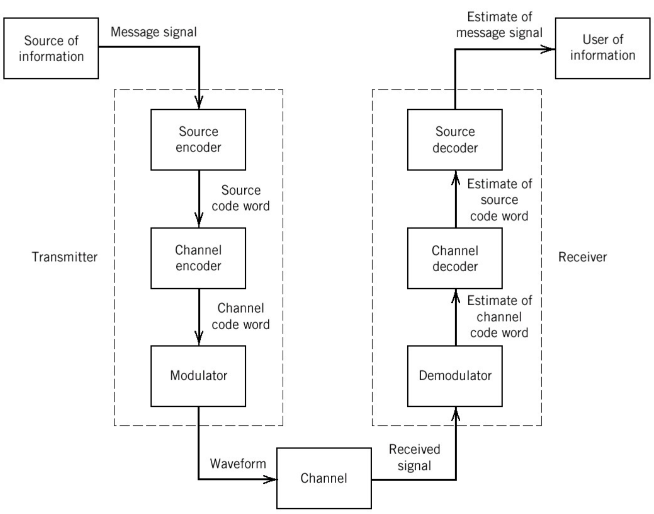
156

Outline

- Introduction to digital communication
- Quantization (A/D) and noise
- PCM
- Companding
- Line codes
- References
  - Notes of Communication Systems, Chap. 4.1-4.3
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 7

157

Block Diagram of Digital Communication



158

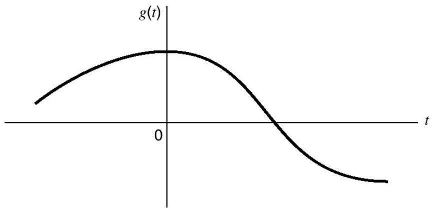
Why Digital?

- 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.

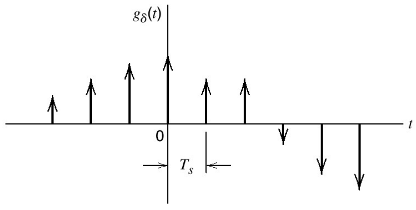
159

Sampling

- How densely should we sample an analog signal so that we can reproduce its form accurately?
- A signal whose spectrum 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



(a)



(b)

160

Quantization

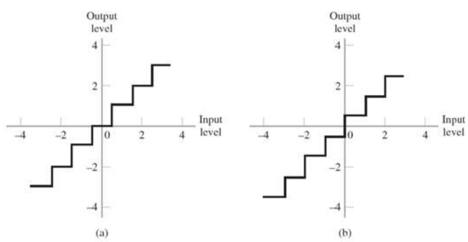
- Quantization is the process of transforming the sample amplitude into a discrete amplitude taken from a finite set of possible amplitudes.
- The more levels, the better approximation.
- No need to transmit exact values: human sense (eye, ear) can only detect finite differences
- Consider memoryless and instantaneous quantization: quantization at time t is independent of other samples
- Quantizers can be of a uniform or nonuniform type.

(a) (b)

Quantizer

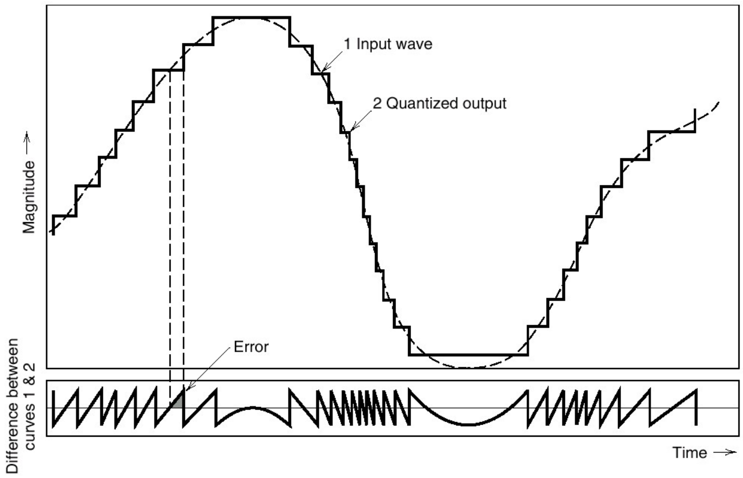
(a) (b)

- Amplitudes m_1, …, m_L are decision levels
- Decision region $\mathcal{I}_k=\{m_k<m\leq m_{k+1}\}$, k=1 ,…, L
- At quantizer output, decision region k is represented by an amplitude v_k
- v_k, k=1, …, L, are called the representation or reconstruction levels
- Mapping v = g(m) is the quantizer characteristic



Quantization Noise

- Error between the input and the output signals



163

Variance of Quantization Noise

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

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

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

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

- **Quantization noise variance**

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

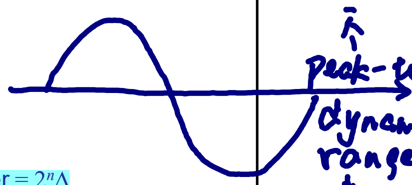
$$= \frac{1}{\Delta}\frac{q^3}{3}\bigg|_{-\Delta/2}^{\Delta/2} = \frac{1}{\Delta}\left[\frac{\Delta^3}{24} - \frac{(-\Delta)^3}{24}\right] = \frac{\Delta^2}{12}$$

164

SNR

- Assume that encoded symbol has n bits
  - 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 message signal
- Assume that message signal fully loads the quantizer:

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



- SNR at the quantizer output:

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

165

SNR

$$\Delta = \frac{2m_p}{2^n} \Rightarrow \Delta^2 = \frac{4m_p^2}{2^{2n}}$$

(assume fully loaded)

$$SNR_o = \frac{12P}{\frac{4m_p^2}{2^{2n}}} = \frac{3P}{m_p^2}2^{2n}$$

•In dB:

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

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

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

•**Each extra bit in the encoder adds 6 dB to the output SNR**

166

Example

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

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

$$SNR_o = \frac{3P}{m_p^2} 2^{2n}$$

- Maximum signal value: $m_p = A_m$
- We have

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

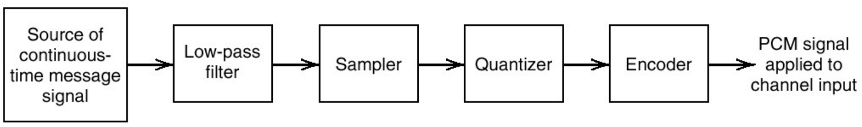
- In dB

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

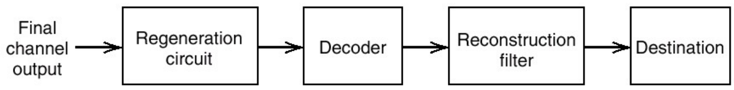
- Audio CDs: $n = 16 \Rightarrow SNR > 90$ dB

PCM = sample + quantize + encode

Pulse-Coded Modulation (PCM)



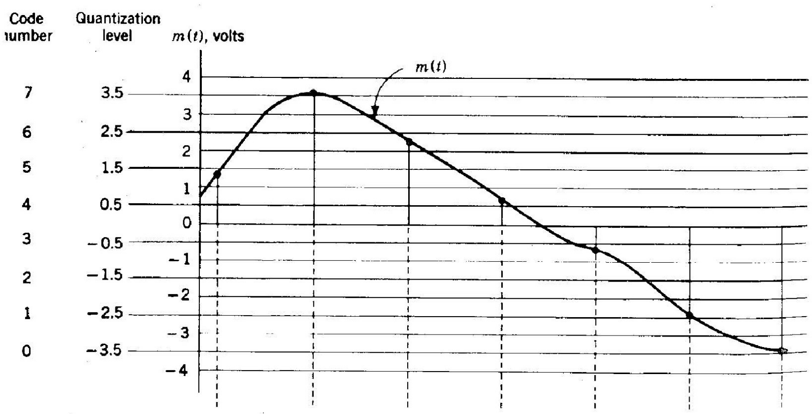
(a) Transmitter



(c) Receiver

- Sample the message signal above Nyquist rate
  - Low-pass filter is applied to prevent aliasing
- Quantize each sample
- Encode discrete amplitudes into a binary codeword
- PCM: not modulation in usual sense; type of Analog-to-Digital Converter.

The PCM Process



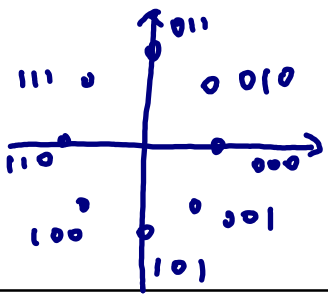
Sample value	1.3	3.6	2.3	0.7	−0.7	−2.4	−3.4
Nearest quantization level	1.5	3.5	2.5	0.5	−0.5	−2.5	−3.5
Code number	5	7	6	4	3	1	0
Binary representation	101	111	110	100	011	001	000

169

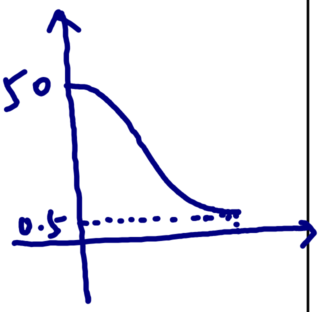
Grey Mapping

- 4-level coding:
  - Numerical Mapping: {0, 1, 2, 3} → {00, 01, 10, 11}
  - Gray Mapping: {0, 1, 2, 3} → {01, 00, 10, 11}
- 8-lvel Gray mapping:
  - {0,1,2, 3, 4, 5, 6, 7} → {011,010,000,001,101,100,110,111}
- General to any number of levels, 2^n

Gray mapping:
adjacent letters only
differ by 1 position



170



$$e_1 = \frac{\frac{\Delta}{2}}{50}$$

$$e_2 = \frac{\frac{\Delta}{2}}{0.5}$$

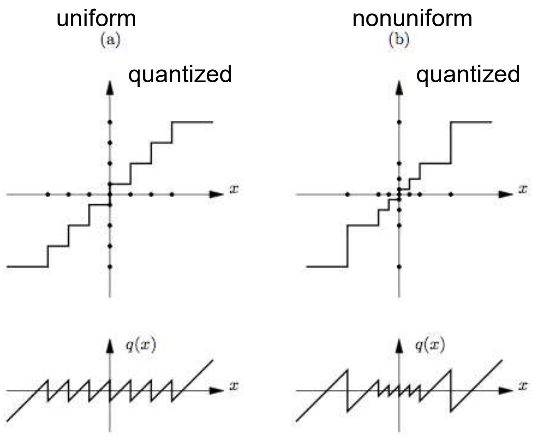
Problem With Uniform Quantization

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

171

Solution: Nonuniform Quantization

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



172

Companding = Compressing + Expanding

- A practical (and equivalent) solution to nonuniform quantization: uniform
  - Compress the signal ← pre-processing for less quantization loss
  - Quantize it (using a uniform quantizer)
  - Transmit it
  - Expand it ← recover by inverse
- Companding corresponds to pre-emphasis/ de-emphasis, as in FM
- Pre-distorting a message signal in order to achieve better performance in the presence of noise, and then remove the distortion at the receiver
- Exact SNR gain obtained with companding depends on the exact form of the compression used
- With proper companding, output SNR can be made insensitive to peak to average power ratio
- Ideally, compression and expansion are exactly inverse of each other

173

Compander Standards: μ-Law vs A-Law

μ-Law (North America and Japan, typical μ = 255)

$$y = F(x) = sgn(x)\frac{\ln(1+\mu|x|)}{\ln(1+\mu)}, |x| < 1$$

$$F^{-1}(y) = sgn(y)(1/\mu)((1+\mu)^{|y|} - 1)$$

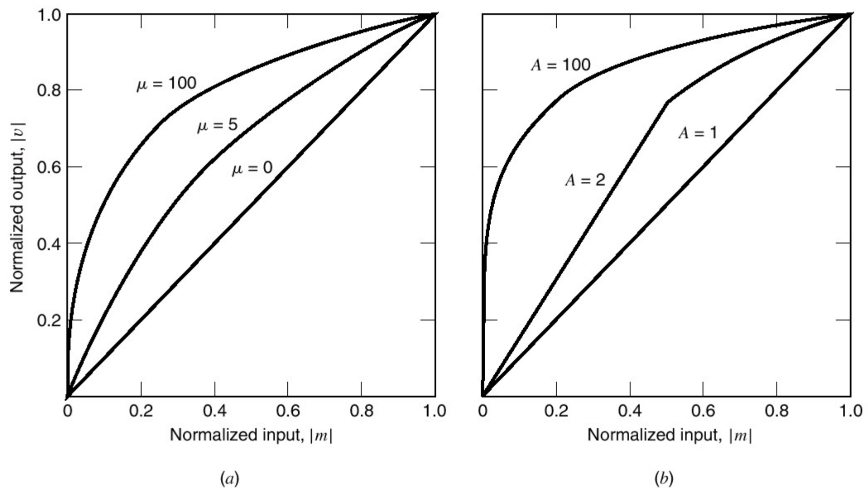
A-Law (in most countries of the world; typical A = 87.6)

$$y = F(x) = \begin{cases} sgn(x)\dfrac{A|x|}{1+lnA}, & |x| < \dfrac{1}{A} \\ sgn(x)\dfrac{1+\ln(A|x|)}{1+lnA}, & \dfrac{1}{A} \le |x| \le 1 \end{cases}$$

$$x = F^{-1}(y) = \begin{cases} sgn(y)\dfrac{|y|(1+lnA)}{A}, & |y| < \dfrac{1}{1+lnA} \\ sgn(y)\dfrac{\exp(|y|(1+lnA)-1)}{A}, & \dfrac{1}{1+lnA} \le |y| \le 1 \end{cases}$$

174

Compander Standards: 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: μ = 255, A = 87.6.

Applications of PCM & Variants

- Speech:
  - PCM: 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.

Line Coding

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

177

Desired Properties of a Line Code

- Transmission bandwidth: as small as possible
- Power efficiency: as small power as possible for given bandwidth and target error probability
- Error detection and correction capability
- Favorable power spectral density: PSD should match the channel characteristics
- Adequate timing content: Should allow extracting timing information
- Transparency: Data should be transmitted reliably regardless of its binary pattern

178

Line Codes

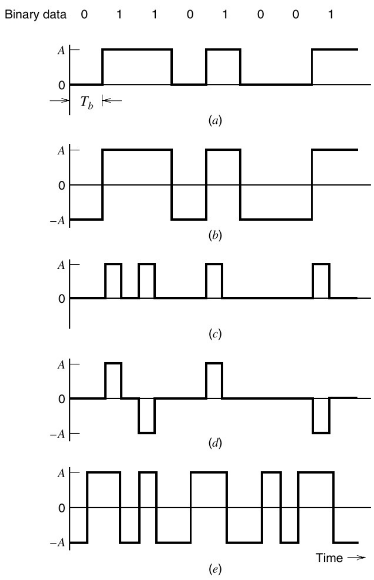
Unipolar nonreturn-to-zero (NRZ) signaling (on-off signaling)

Polar NRZ signaling

Unipolar Return-to-zero (RZ) signaling

Bipolar RZ signaling

Manchester code



Amplitude Spectrum of Line Codes

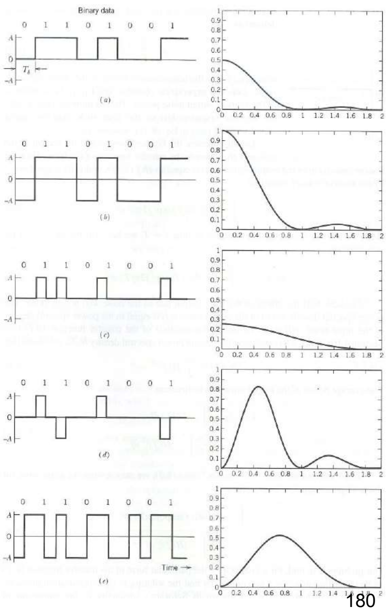
Unipolar nonreturn-to-zero (NRZ) signaling (on-off signaling)

Polar NRZ signaling

Unipolar Return-to-zero (RZ) signaling

Bipolar RZ signaling

Manchester code



Summary

- Digitization of signals requires
  - Sampling: a signal of bandwidth W is sampled at the Nyquist frequency 2W
  - Quantization: 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
  - "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

181

Imperial College London

Lecture 9: Baseband Digital Transmission

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

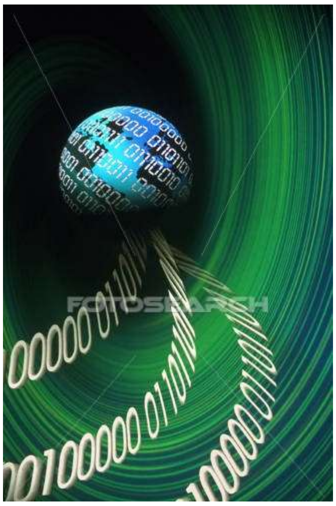
Imperial College London

182

Outline

- 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



183

Analog and Digital Communications

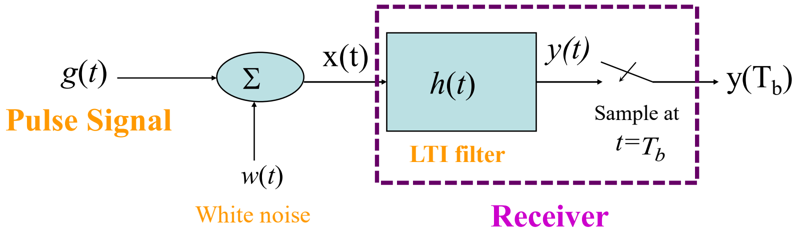
- Difference between analog and digital communication systems:
  – Analog communication systems: reproduce transmitted waveform accurately

    ⇒ *Usually* use signal-to-noise ratio to assess the quality of the system
  – Digital communication systems: recover the transmitted symbol correctly

    ⇒ *Usually* use the probability of error at the receiver to assess the quality of the system

184

Matched Filter



$$x(t) = g(t) + w(t) \qquad 0 \le t \le T_b$$

- w(t): white noise with zero mean and PSD $\frac{N_o}{2}$
- Receiver wants to detect the pulse in presence of additive noise
  - Receiver knows what pulse shape it is looking for
- Goal: Design a receive (linear) filter that minimizes the effect of noise
  - Optimize the design of the filter

Matched Filter

- Filter output will be:

$$y(t) = g(t) * h(t) + w(t) * h(t)$$
$$= g_o(t) + n(t)$$

$$g_o(t) = g(t) * h(t) = \int_{-\infty}^{+\infty} g(\tau) h(t-\tau) d\tau$$

- We want instantaneous power of signal component $g_o(t)$ at time $t=T_b$ as large as possible compared to noise component n(t)

- Maximize peak pulse signal-to-noise ratio (η)

$$\eta = \frac{|g_o(T_b)|^2}{E[n^2(T_b)]} = \frac{\text{instantaneous power}}{\text{average power}}$$

Matched Filter Derivation

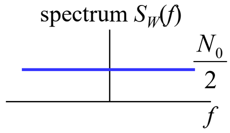
- Noise

$$n(t) = w(t) * h(t)$$

$$S_N(f) = \underbrace{S_W(f)}_{\text{AWGN}}\ \underbrace{S_H(f)}_{\text{Matched filter}} = \frac{N_0}{2} |H(f)|^2$$

$$E[n^2(t)] = \int_{-\infty}^{\infty} S_N(f)\ df = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2\ df$$

Noise power spectrum $S_W(f)$



- Signal

$$g_o(t) = g(t) * h(t) \qquad G_0(f) = H(f) G(f)$$

$$g_o(t) = \int_{-\infty}^{\infty} H(f)\ G(f)\ e^{j 2 \pi f t}\ df$$

$$|g_o(T_b)|^2 = \left| \int_{-\infty}^{\infty} H(f)\ G(f)\ e^{j 2 \pi f T_b}\ df \right|^2$$

187

Matched Filter Derivation

- Find $h(t)$ that maximizes pulse peak SNR η

$$\eta = \frac{\left| \int_{-\infty}^{\infty} H(f) G(f)\ e^{j 2 \pi f T_b}\ df \right|^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2\ df}$$

- Schwartz's inequality: Consider two energy signals $\phi_1(x)$ and $\phi_2(x)$ We have

$$\left| \int_{-\infty}^{\infty} \phi_1(x) \phi_2^*(x) dx \right|^2 \le \int_{-\infty}^{\infty} |\phi_1(x)|^2 dx \int_{-\infty}^{\infty} |\phi_2(x)|^2 dx$$

equality holds if and only if $\phi_1(x) = k\phi_2(x)$ for an arbitrary constant k.

188

Matched Filter Derivation

Let $\varphi_1(f) = H(f)$ and $\varphi_2(f) = G^*(f)e^{-j2\pi f T_b}$

$$\left| \int_{-\infty}^{\infty} H(f)\, G(f)\, e^{j2\pi f T_b}\, df \right|^2 \le \int_{-\infty}^{\infty} |H(f)|^2\, df \int_{-\infty}^{\infty} |G(f)|^2\, df$$

$$\eta = \frac{\left| \int_{-\infty}^{\infty} H(f)\, G(f)\, e^{j2\pi f T_b}\, df \right|^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2\, df} \le \frac{2}{N_0} \int_{-\infty}^{\infty} |G(f)|^2\, df$$

$$\eta_{\max} = \frac{2}{N_0} \int_{-\infty}^{\infty} |G(f)|^2\, df \quad \text{occurs}$$

if $H_{opt}(f) = kG^*(f)e^{-j2\pi f T_b}$

Hence, $h_{opt}(t) = kg^*(T_b - t) = kg(T_b - t)$

189

Properties of Matched Filters

- Impulse response is

$$h_{\text{opt}}(t) = k\, g(T_b - t)$$

T_b **:** symbol period,

$g(t)$**:** transmitter pulse shape, k: gain

– scaled, time-reversed and shifted version of $g(t)$

– duration and shape determined by pulse shape $g(t)$

- Maximizes peak pulse SNR

$$\eta_{\max} = \frac{2}{N_0} \int_{-\infty}^{\infty} |G(f)|^2\, df = \frac{2}{N_0} \int_{-\infty}^{\infty} |g(t)|^2\, dt = \frac{2E}{N_0} = \text{SNR}$$

– does not depend on pulse shape $g(t)$

– proportional to signal energy (energy per bit) E

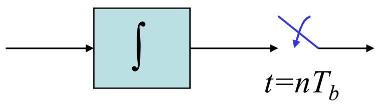
– inversely proportional to noise power spectral density

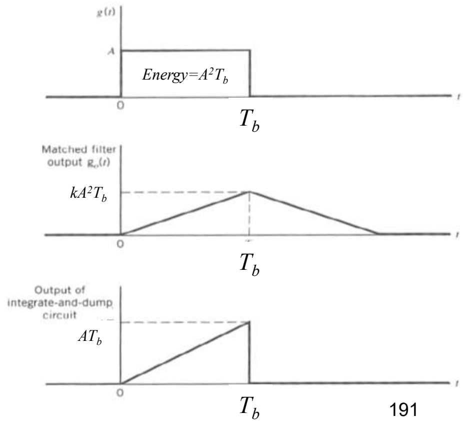
190

Matched Filter for Rectangular Pulse

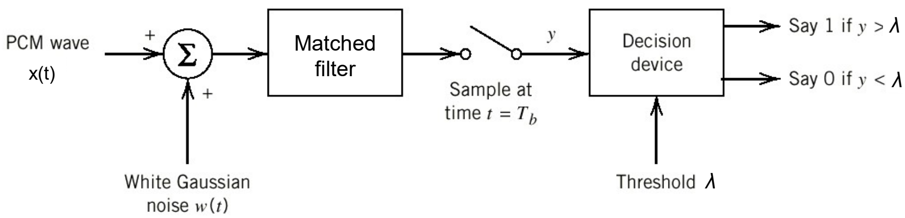
- Matched filter for rectangular pulse shape

  -Impulse response of matched filter is a rectangular pulse of same duration
- Convolve input with rectangular pulse of duration T_b sec and sample result at T_b sec is equivalent to
  - Integrate for T_b sec
  - Sample at symbol period T_b sec
  - Reset integration for next time period
- Integrate and dump circuit





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 with double-sided noise PSD of $N_0/2$
  - Matched filter is rectangular (set $kAT_b = 1$ for simplicity)
- Effect of additive noise on digital transmission: symbol 1 may be mistaken for 0, and vice versa ⇒ **bit errors**
- What is the probability of a bit error?

Distribution of Noise

- After the matched filter, the pre-detection signal is

$$Y = y(T_b) = \frac{1}{T_b}\int_0^{T_b} x(t)dt = s + \underbrace{\frac{1}{T_b}\int_0^{T_b} w(t)dt}_{\text{noise, N}}$$

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

$$\sigma^2 = E[N^2] = \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} E[w(t)w(u)]\,dt\,du$$

$$= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} R_W(t,u)\,dt\,du = \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} \frac{N_o}{2}\delta(t-u)\,dt\,du = \frac{N_o}{2T_b}$$

- *N* is a Gaussian random variable with pdf :

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

193

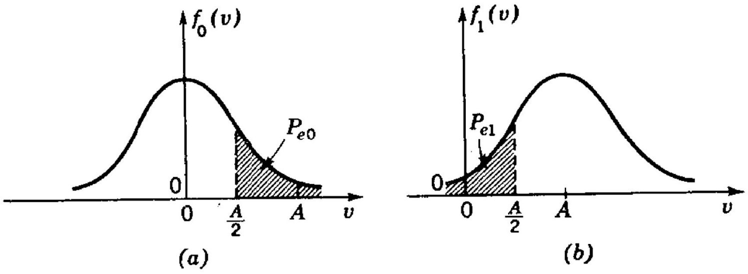
Decision

- If a symbol 0 were transmitted: $Y = N$
  - *Y* will have a pdf of $N(0, \sigma^2)$
- If a symbol 1 were transmitted: $Y = A + N$
  - *Y* will have a PDF of $N(A, \sigma^2)$
- Use λ as the decision threshold:
  - if $Y < \lambda$, choose symbol 0
  - if $Y > \lambda$, choose symbol 1

194

Errors

- Two cases of decision error:
  - (i) symbol 0 was transmitted, but symbol 1 is decided
    - with probability P_{e0}
  - (ii) symbol 1 was transmitted, but symbol 0 is decided
    - with probability P_{e1}



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

195

Case I

- Probability of (i) = Probability (error | symbol 0 was transmitted) $\times$ Probability (symbol 0 is transmitted)

$$P_i = P_{e0} \times p_0$$

where:

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

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

196

Case II

- Probability of (ii) = Probability (error | symbol 1 was transmitted) × Probability (symbol 1 is transmitted)

$$P_{ii} = P_{e1} \times p_1$$

where:

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

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

197

Total Error Probability

- Total error probability:

$$\begin{aligned} P_e(\lambda) &= P_i + P_{ii} \\ &= p_1 P_{e1} + (1-p_1) P_{e0} \\ &= p_1 \int_{-\infty}^{\lambda} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(n-A)^2}{2\sigma^2}\right) dn + (1-p_1) \int_{\lambda}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n^2}{2\sigma^2}\right) dn \end{aligned}$$

- Optimal threshold (ignore the derivation):

$$\lambda = -\frac{\sigma^2}{A} \ln \frac{p_1}{1-p_1} + \frac{A}{2}$$

- al probability: ($p_1 = p_0 = 1 - p_1$) ⇒ $\lambda = A/2$ and $P_{e0} = P_{e1}$
- Transmit symbols with unequal probability: if $p_0 > p_1$, then $\lambda < A/2$, and $P_{e0} < P_{e1}$

198

Calculation of P_e for $p_1 = p_0 = .5$

- 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:

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

199

Probability of Error

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

Energy of a pulse is $E = A^2 T_b$

We transmit a pulse only half of the time (on average) -> Average energy per bit (E_b):

$$E_b = \frac{A^2 T_b}{2}$$

Noise variance: $\sigma^2 = \dfrac{N_o}{2T_b}$

Probability of error in terms of energy per bit and noise PSD:

$$P_e = Q\left(\sqrt{\frac{E_b}{N_o}}\right)$$

200

Example

- **Example 1:** $A/\sigma = 7.4 \Rightarrow P_e = 10^{-4}$

⇒ For a transmission rate of 10^5 bits/sec, there will be an error every 0.1 seconds on the average

- **Example 2:** $A/\sigma = 11.2 \Rightarrow P_e = 10^{-8}$

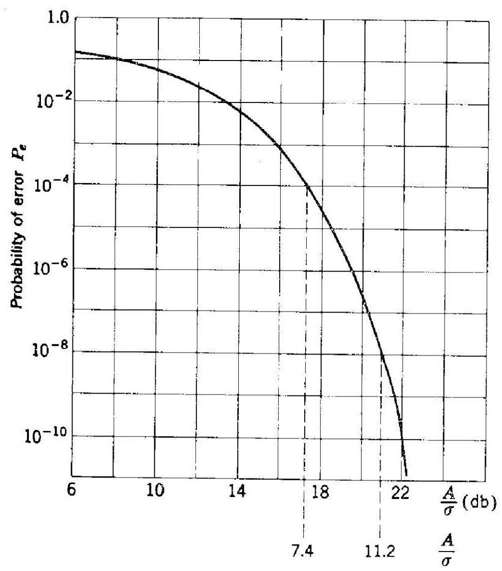
⇒ For a transmission rate is 10^5 bits/sec, there will be an error every 17 mins on the average

- 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).

201

Probability of Bit Error



202

Q-function

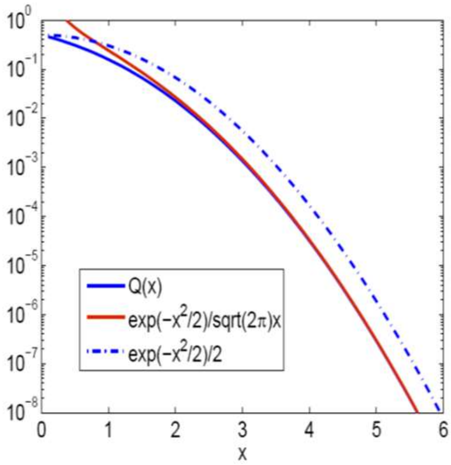
- Upper bounds and good approximations:

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

  - which becomes tighter for large *x*, and

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

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



Imperial College London

Lecture 10: Digital Modulation

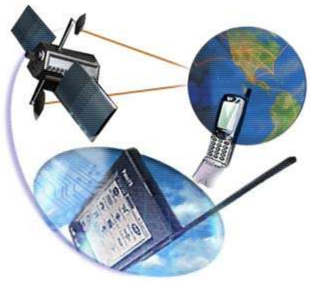
Professor Geoffrey Li

Department of Electrical and Electronic Engineering

Imperial College London

Outline

- ASK and its bit-error rate (BER)
- FSK and its BER
- PSK and its BER
- References
  - Notes of Communication Systems, Chap. 4.5
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 9



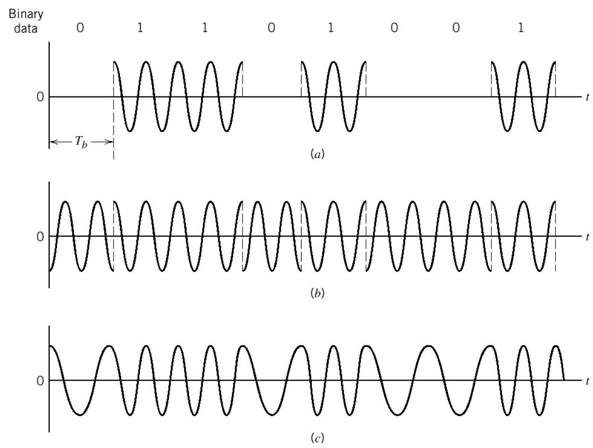
205

Basic Forms: ASK, PSK, and FSK

(a) Amplitude-shift keying (ASK).

(b) Phase-shift keying (PSK).

(c) Frequency-shift keying (FSK).



206

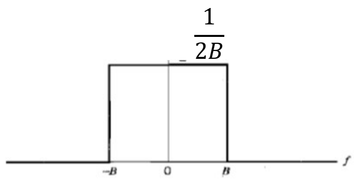
Bandwidth, Symbol Duration and Rate

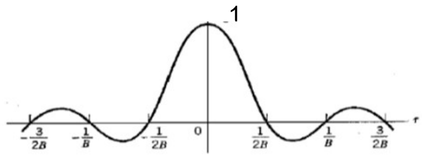
Consider a baseband waveform with bandwidth, B,

$$P(f) = \begin{cases} \frac{1}{2B}, & |f| \le B, \\ 0 & otherwise. \end{cases}$$

Its corresponding time-domain waveform can be expressed as

$$p(t) = \int_{-B}^{B} \frac{1}{2B} e^{-j2\pi f t} dt = \frac{\sin(2\pi Bt)}{2\pi Bt} = \mathrm{sinc}(2\pi Bt)$$





Symbol duration, $T = \frac{1}{2B}$, or symbol rate, $R = \frac{1}{T} = 2B$

A modulated waveform with symbol duration, T, bandwidth will be $B = \frac{1}{2T} = \frac{R}{2}$

(W is used instead of B in the subsequent discussion)

207

Evolution: Digital Passband Modulation → PSK

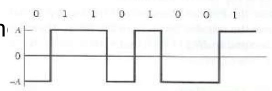
- In baseband, linearly modulated waveform is given by

$$u(t) = \sum_n b[n] p(t - nT)$$

$\{b[n]\}$: the sequence of symbols to be transmitted

$p(t)$: the modulating pulse.

- If symbols take values from {-1, +1} and m pulse is a rectangular time-limited pulse.



$$s(t) = u(t)\cos(2\pi f_c t)$$

- In passband, we can simply transmit $s(t) = \cos(2\pi f_c t + \pi)$ $\quad s(t) = \cos(2\pi f_c t)$

- For the nth symbol interval, $nT \le t \le (n+1)T$, we have if $b[n]=+1$ and if $b[n]=-1$

208

- Binary antipodal modulation switches the phase of the carrier between 0 and π

PSK Signal Constellations

- BPSK modulates only the I-component. We can modulate both I and Q components,

$$s(t) = u_I(t)\cos(2\pi f_c t) - u_Q(t)\sin(2\pi f_c t)$$

where

$$u_I(t) = \sum_n b_I[n]p(t - nT) \qquad u_Q(t) = \sum_n b_Q[n]p(t - nT)$$

- Assume again $b_I[n]$ and $b_Q[n]$ take values from $\{-1, +1\}$, and rectangular time-limited pulse. For the nth symbol interval, $nT \le t \le (n+1)T$

$s(t) = \cos(2\pi f_c t) - \sin(2\pi f_c t) = \sqrt{2}\cos(2\pi f_c t + \pi/4)$ if $b_I[n]=+1$, $b_Q[n]=+1$

$s(t) = \cos(2\pi f_c t) + \sin(2\pi f_c t) = \sqrt{2}\cos(2\pi f_c t - \pi/4)$ if $b_I[n]=+1$, $b_Q[n]=-1$

$s(t) = -\cos(2\pi f_c t) - \sin(2\pi f_c t) = \sqrt{2}\cos(2\pi f_c t + 3\pi/4)$ if $b_I[n]=-1$, $b_Q[n]=+1$

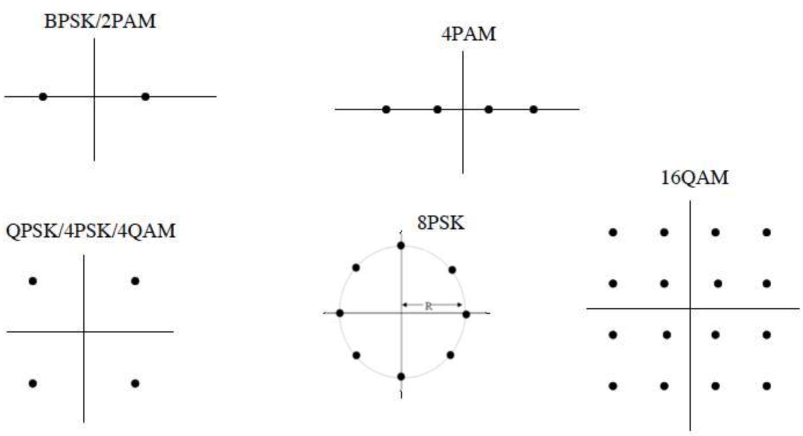
$s(t) = -\cos(2\pi f_c t) + \sin(2\pi f_c t) = \sqrt{2}\cos(2\pi f_c t - 3\pi/4)$ if $b_I[n]=-1$, $b_Q[n]=-1$

- Modulation switches the phase among $\{\pm\pi/4, \pm 3\pi/4\}$, hence it is called quadrature phase-shift keying (QPSK)

209

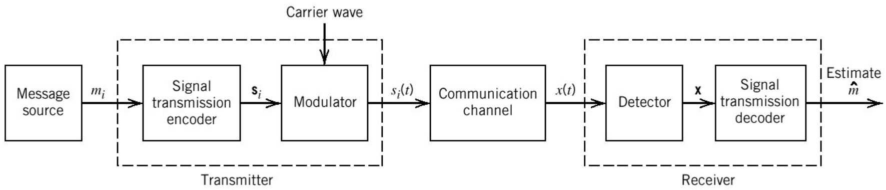
Quadrature Amplitude Modulation (QAM)

$$s(t) = u_I(t)\cos(2\pi f_c t) - u_Q(t)\sin(2\pi f_c t)$$



210

Demodulation



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

211

ASK

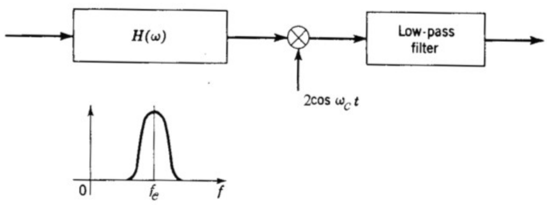
- Amplitude shift keying (ASK) = on-off keying (OOK)

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

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 $[f_c - W, f_c + W]$, $(2W = \frac{1}{T_b})$

212

Coherent Demodulation

- Pre-detection signal:

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

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

$$\begin{aligned} y(t) &= [A(t)+n_I(t)]2\cos^2(2\pi f_c t)-n_Q(t)2\sin(2\pi f_c t)\cos(2\pi f_c t) \\ &= [A(t)+n_I(t)](1+\cos(4\pi f_c t))-n_Q(t)\sin(4\pi f_c t) \end{aligned}$$

- After low-pass filtering:

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

Bit-Error Rate (BER) of ASK

- **Reminder:** The PSD of in-phase noise component, $n_I(t)$, is N_o, double the PSD of the original band-pass noise $n(t)$.
  - BER analysis is similar to that for baseband transmission (with adjusted signal and noise powers)
- Assume equi-probable transmission of 0s and 1s
- Then the decision threshold must be $A/2$ and the probability of error is given by:

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

- Transmission energy for a pulse is

$$E = A^2 T_b / 2$$

- Average energy per bit:

$$E_b = A^2 T_b / 4$$

only half chance to transmit the pulse (when b[n]=1)

- Noise variance: $\sigma^2 = N_o \times 2W = N_o \times \frac{1}{T_b} = N_o/T_b$

$$P_{e,ASK} = Q\left(\sqrt{\frac{E_b}{N_o}}\right)$$

Probability of error in terms of energy per bit and noise PSD

ASK Example

ASK Modulation System with Coherent Demodulation:

- Carrier Amplitude, $A = 0.7$ v
- Standard Deviation of White Gaussian Noise, $\sigma = 0.125$ v
- Symbols "0" and "1" with equal probability

What is BER?

$$P_e = Q\left(\frac{A}{2\sigma}\right) = Q\left(\frac{0.7}{2\times 0.125}\right) = Q(2.8) = 2.8\times 10^{-3}$$

$$Q(x) \le \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}x}$$

How to convert $\frac{A}{\sigma}$ and $\frac{E_b}{N_o}$ into dB?

$20\log_{10}\frac{A}{\sigma} = 20log_{10}\frac{0.7}{0.125} = 14.96$ dB, $10\log_{10}\frac{E_b}{N_o}$

PSK

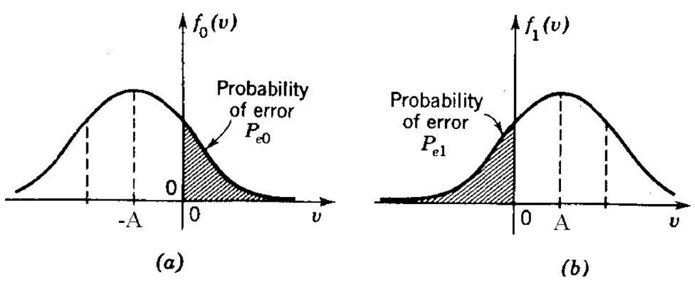
- Phase shift keying (PSK)

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

- Use coherent detection to eventually get detection signal:

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

- Pdfs for PSK for equi-probable 0s and 1s in noise (use threshold 0 for detection):
  - (a): symbol 0 transmitted
  - (b): symbol 1 transmitted



BER of PSK

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

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

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

- Average energy per bit: $E_b = A^2 T_b / 2$
- Noise variance: $\sigma^2 = N_o / T_b$
- Probability of error in terms of energy per bit and noise PSD: $P_{e,PSK} = Q\left(\sqrt{\frac{2E_b}{N_o}}\right)$

217

PSK Example

PSK Modulation System with Coherent Demodulation:

- Carrier Amplitude, $A = 0.7$ v
- Standard Deviation of White Gaussian Noise, $\sigma = 0.125$ v
- Symbols “0” and “1” with equal probability

What is BER?

$$P_e = Q\left(\frac{A}{\sigma}\right) = Q\left(\frac{0.7}{0.125}\right) = Q(5.6) = 1.1 \times 10^{-8}$$

$$Q(x) \leq \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}x} = 1.1 \times 10^{-8}$$

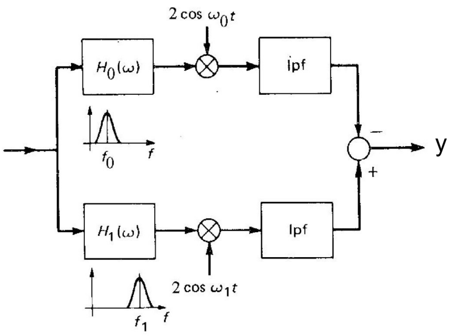
218

Frequency Shift Keying (FSK)

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

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

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



FSK Output

- Each branch = an ASK detector

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

- $n_0(t)$: noise output of top branch
- $n_1(t)$: noise output of bottom branch
- Each noise term has identical statistics to $n_l(t)$
- Output if transmitting “1”

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

- Output if transmitting “0”

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

The Sum of Two RVs

- Noise is sum or difference of two independent zero-mean rvs:
  - 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/uncorrelated 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

$$\sigma_Y^2 = E[X_1^2] + E[X_2^2] = \sigma_1^2 + \sigma_2^2$$

221

BER of FSK

- Set detection threshold to 0
- Difference from PSK: 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 variances add, so it is doubled
  - $\hat{\sigma}^2 = 2\sigma^2 = \frac{2N_o}{T_b}$ $(\sigma^2 = N_o/T_b)$, $\hat{\sigma} = \sqrt{2}\,\sigma$

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

- Average energy per bit: $E_b = A^2T_b/2$
- Probability of error in terms of energy per bit and noise PSD:

$$P_{e,FSK} = Q\left(\sqrt{\frac{E_b}{N_o}}\right)$$

222

Minimum-Shift Keying (MSK)

More on FSK:

- symbol 0 → frequency f_0, symbol 1 → frequency f_1,
- the unmodulated carrier frequency: $f_c = \frac{1}{2}(f_0 + f_1)$,
- frequency separation: $\Delta f = |f_1 - f_0|$,
- bandwidth: $W \approx \Delta f$,
- the symbol period: T ($f_c T \gg 1$ in practice, $\cos(2\pi f_1 t)$ and $\sin(2\pi f_0 t)$ orthogonal within the symbol period).

Δf and/or T should be large enough to make $\cos(2\pi f_1 t)$ and $\sin(2\pi f_0 t)$ are orthogonal !

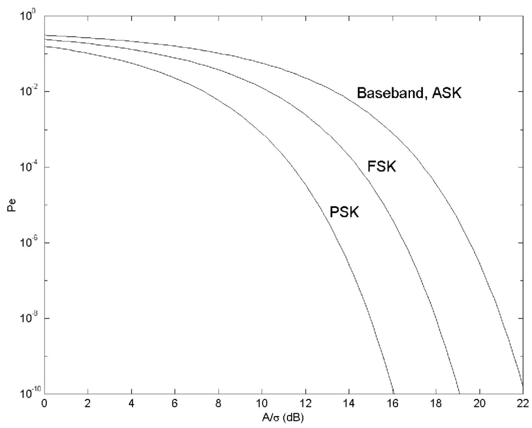
Minimum-shift keying (MSK): using the minimum separation $\Delta f = \frac{1}{2T}$.

Why $F = 1/2T$ is the minimum separation?

$$\frac{1}{T}\int_o^T \cos(2\pi f_1 t)\cos(2\pi f_0 t)\,dt = \frac{1}{2T}\int_o^T (\cos(2\pi(f_1+f_0)t) + \cos(2\pi(f_1 - f_0)t))\,dt$$

$$= \frac{1}{2T}\int_o^T (\cos(2\pi 2 f_c)t) + \cos(2\pi\Delta f t))\,dt = \frac{1}{2}\frac{\sin(4\pi f_c T)}{4\pi f_c T} + \frac{1}{2}\frac{\sin(2\pi\,\Delta f T)}{2\pi\Delta f T}$$

Comparison of Three Schemes



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

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

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

ASK: $\frac{E_b}{N_o} = \frac{A^2}{4\sigma^2}$; FSK: $\frac{E_b}{N_o} = \frac{A^2}{2\sigma^2}$; PSK: $\frac{E_b}{N_o} = \frac{A^2}{2\sigma^2}$

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, it was used in (Advanced Mobile Phone System) 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.
  - 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

Lecture 11: Noncoherent Demodulation

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

Imperial College London

Outline

- Noncoherent demodulation of ASK
- Noncoherent demodulation of FSK
- Differential demodulation of DPSK
- References
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 9

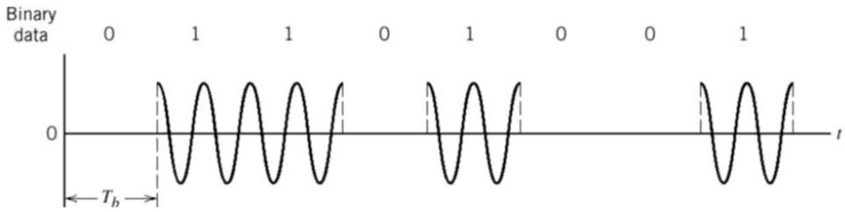
227

Coherent vs. Noncoherent Demodulation

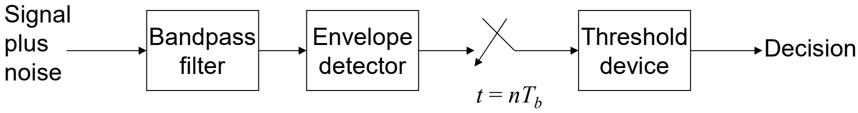
- Coherent demodulation assumes perfect synchronization.
  - Needs a phase locked loop.
- Accurate phase synchronization may be difficult in a dynamic channel.
  - Phase synchronization error is due to varying propagation delays, frequency drift, instability of the local oscillator, effects of strong 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.
- Phase Φ is assumed to be uniformly distributed on $[0, 2\pi]$.
- Circuitry/receiver is simpler, but analysis is more difficult!

228

ASK Waveform



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

- Envelope

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\sin(2\pi f_c t)$$

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

Distribution of the Envelope

- Symbol 0 sent → envelope, $r(t) = \sqrt{n_I^2(t) + n_Q^2(t)}$,

  with Rayleigh distribution

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

- Symbol 1 sent →the envelope, $r(t) = \sqrt{(A + n_I(t))^2 + n_Q^2(t)}$ with Rician distribution

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

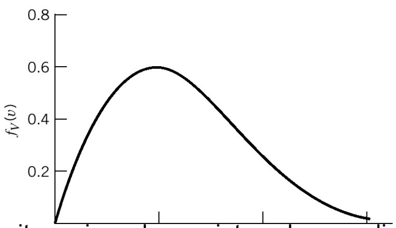
- 1st case dominates the error probability when $\frac{A}{\sigma} \gg 1$.

231

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:

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

232

Rician Distribution

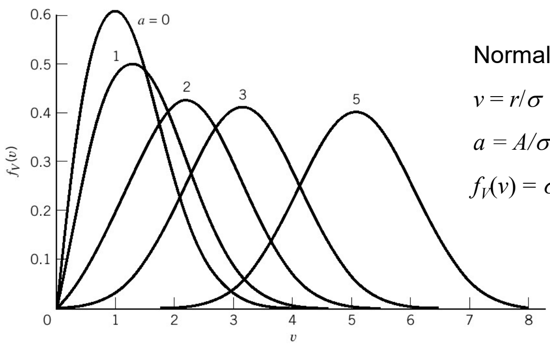
- If X has nonzero mean *A*, R has Rician distribution:

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

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

$v = r/\sigma$

$a = A/\sigma$

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

Error Probability

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

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

- Final expression

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

- Coherent demodulation

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

- Noncoherent demodulation results in some performance degradation. Yet, for large SNR, performances are similar.

ASK Example

ASK Modulation System with Noncoherent Demodulation:

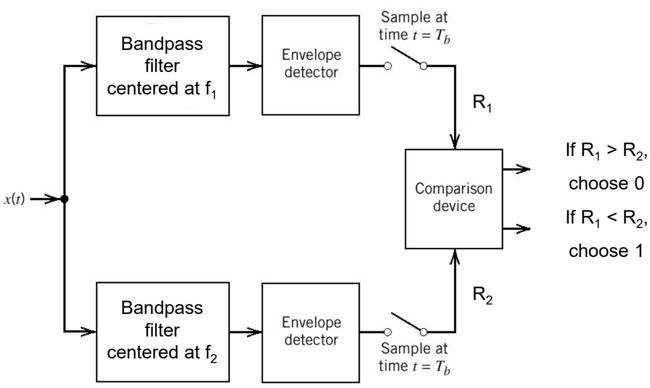
- Carrier Amplitude, $A = 0.7$ v
- Standard Deviation of White Gaussian Noise, $\sigma = 0.125$ v
- Symbols “0” and “1” with equal probability

What is BER?

$$P_e = \frac{1}{2}\exp(-\frac{A}{8\sigma}) = \frac{1}{2}\exp(-\frac{0.7^2}{8\times 0.125^2}) = 9.9\times 10^{-3}$$

For ASK with coherent detection, $P_e = 2.8 \times 10^{-3}$

Noncoherent Demodulation of FSK



Distribution of Envelope

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

$$y_1(t) = n_1(t)$$
$$y_2(t) = n_2(t) + A\cos(2\pi f_2 t)$$

- Again, first branch has Rayleigh distribution

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

- while second has Rice distribution

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

- Note that envelopes R_1 and R_2 are independent.

237

BER of FSK

- Error occurs if Rice < Rayleigh

$$\begin{aligned}
P_e &= P(R_2 < R_1) \\
&= \int_0^\infty \int_{r_2}^\infty \frac{r_2}{\sigma^2} e^{-(r_2^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar_2}{\sigma^2}\right) \frac{r_1}{\sigma^2} e^{-r_1^2/(2\sigma^2)} dr_1 dr_2 \\
&= \int_0^\infty \frac{r_2}{\sigma^2} e^{-(r_2^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar_2}{\sigma^2}\right) \int_{r_2}^\infty \frac{r_1}{\sigma^2} e^{-r_1^2/(2\sigma^2)} dr_1 dr_2 \\
&= \int_0^\infty \frac{r_2}{\sigma^2} e^{-(2r_2^2+A^2)/(2\sigma^2)} I_0\left(\frac{Ar_2}{\sigma^2}\right) dr_2 \\
&= \frac{1}{2} e^{-A^2/(4\sigma^2)} \int_0^\infty \frac{x}{\sigma^2} e^{-(x^2+\alpha^2)/(2\sigma^2)} I_0\left(\frac{\alpha x}{\sigma^2}\right) dx \qquad x = \sqrt{2}r_2,\ \alpha = A/\sqrt{2}
\end{aligned}$$

- Observe that the integrand is a Rician density

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

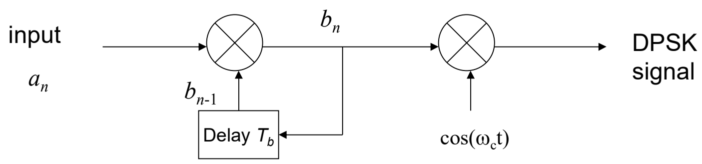
- Coherent demodulation

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

238

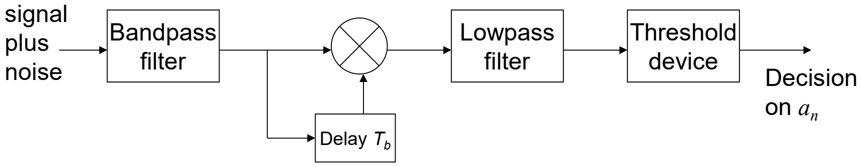
DPSK: Differential PSK

- 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, $b_n \in \pm 1$.



239

Differential Demodulation



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

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

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

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

240

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

PSK Example

DPSK Modulation System with Differential Demodulation:

- Carrier Amplitude, $A = 0.7$ v
- Standard Deviation of White Gaussian Noise, $\sigma = 0.125$ v
- Symbols “0” and “1” with equal probability

What is BER?

$$P_e = \frac{1}{2} exp\left(-\frac{A^2}{2\sigma^2}\right) = \frac{1}{2} exp\left(-\frac{0.7^2}{2 \times 0.125^2}\right) = 7.7 \times 10^{-8}$$

For PSK with coherent demodulation, $P_e = 1.1 \times 10^{-8}$.

Summary and Comparison

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

ASK: $\frac{E_b}{N_o}=\frac{A^2}{4\sigma^2}$; FSK: $\frac{E_b}{N_o}=\frac{A^2}{2\sigma^2}$; PSK: $\frac{E_b}{N_o}=\frac{A^2}{2\sigma^2}$

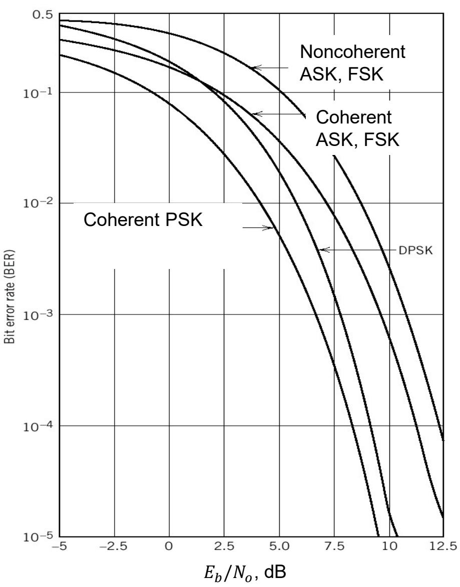
243

Summary and Comparison

ASK: $\frac{E_b}{N_o}=\frac{A^2}{4\sigma^2}$

FSK: $\frac{E_b}{N_o}=\frac{A^2}{2\sigma^2}$

PSK: $\frac{E_b}{N_o}=\frac{A^2}{2\sigma^2}$



244

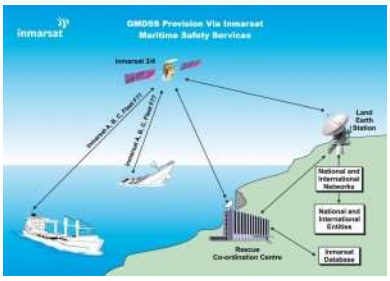
Conclusions

- 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.

245

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



246

Imperial College London

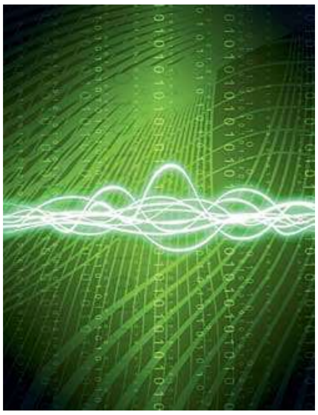
Lecture 12: Entropy and Data Compression

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

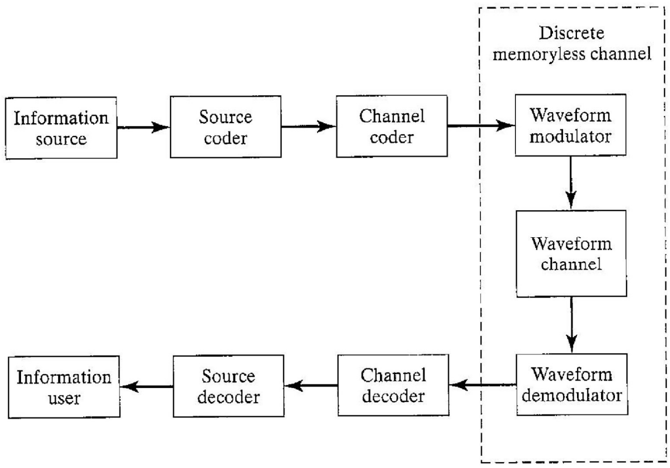
Imperial College London

Outline



- What is information theory?
- Entropy
  - Definition of information
  - Entropy of a source
- Source coding (compression)
  - Source coding theorem
  - Huffman coding
- References
  - Notes of Communication Systems, Chap. 5.1-5.4
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 10

Model of a Digital Communication System



249

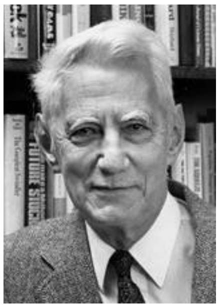
What is Information?

- Information: any new knowledge about something
  - How to store information efficiently?
  - How to transmit information over noisy channels?
- Information is everywhere
  - Collected by sensory system, transmitted via nervous system, processed in brain, ...
  - Stored in DNA, in hard-drives, in books, ...
  - Transmitted over the phone line, over the air, over generations, ...
- Information, knowledge, data, meaning, news, notice, instruction, ...
- Can we quantify information?

250

What is Information Theory

- C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, 1948.



- *“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.”*

  Claude E. Shannon

- Two fundamental questions in information theory:
  - What is the ultimate limit on data compression?
  - What is the ultimate transmission rate of reliable communication over noisy channels?

251

Measuring Information

- The game of “20 questions”

- R. V. L. Hartley, “Transmission of Information,” Bell Systems Technical Journal, July 1928.

- Consider an unknown s of which we only know that it belongs to set S with N elements. Learning s amounts to $\log_2 N$ bits of information, or

  Let $s \in S = \{s_1, s_2, \ldots, s_N\}$, then we get $\log_2 N$ bit of information (presumption: s takes $s_1, s_2, \ldots$ with equal probability).
  For example, $N = 4, \log_2 4 = 2\ bits$

252

What about non-uniform distribution?

For example: Almost all students at Imperial are smart.

- ❑ event that an imperial student is smart
  - ➔ not so informative
- ❑ event that an imperial student is not smart
  - ➔ very informative

- 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. (it seems like a continuous function of the probability distribution !)

253

Definition

- Amount of information in a symbol *s* with probability p:

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

- Properties:
  - $p = 1 \Rightarrow I(s) = 0$: a deterministic symbol contains no information
  - $0 \le p \le 1 \Rightarrow 0 \le I(p) \le \infty$: information measure is monotonic and non-negative
  - $p = p_1 \times p_2 \Rightarrow I(s) = I(p_1) + I(p_2)$: information from statistically independent events is additive.

- Logarithm base 2 is commonly used, resulting in **bits**
- **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.

254

Discrete Memoryless Source(DMS)

- 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 and identically distributed (i.i.d.)

- Assume that each symbol has probability

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

Source Entropy

- If symbol s_k has occurred, this corresponds to

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

bits of information.

- 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:** average amount of information per source symbol:

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

- Units: bits/symbol.

Meaning of Entropy

- What does entropy tell us about the source?
- It is the amount of uncertainty before we receive it
- It tells us how many bits of information per symbol we get on the average by learning the source realization.
- Relation with thermodynamic entropy
  - In **thermodynamics**: entropy measures disorder and randomness;
  - In **information theory**: entropy measures uncertainty.

257

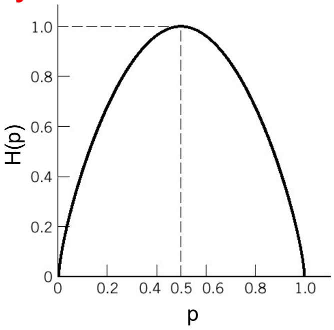
Example: Entropy of a Binary Source

- 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 $p = 1/2$.



$$\frac{\partial H(S)}{\partial p} = 0 \Rightarrow -(1-p)\frac{-1}{1-p} - (-1)\log_2(1-p) - p\frac{1}{p} - \log_2 p = 0 \Rightarrow p = \frac{1}{2}$$

- In general, entropy is maximized when the source is uniformly distributed over its alphabet.

For an N-symbol source, S, with probabilities $p_1, p_2, \ldots, p_N$ ($p_1 + \cdots + p_N = 1$), $\max H(S) = \log_2 N$ when $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$.

258

Example: 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.

Source Coding Theory

- What is the minimum number of bits 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.

Information Content of English

Encoding English Words Letter-by-Letter

- In English, on average there are 4.5 letters per word
- With space (ignoring punctuation and capitalization) we need 5.5 characters per word
- We need 5 bits to encode each letter (26 letters)
- We need 27.5 bits per word

Encoding English Words Word-by-Word

- Assume 171,476 English words (from Google)
- Equivalent to 18 bits per word ($2^{18} = 262{,}144$)

Encoding English Semantically

Average Codeword Length

- l_n: number of bits used to code the k-th symbol
- N: total number of symbols
- p_n: probability of occurrence of symbol n
- Define the **average codeword length**:

$$\bar{L} = \sum_{n=1}^{N} p_n l_n$$

- 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 long codewords.

Frequencies of Words in English

1	the	26	they	51	when	76	come
2	be	27	we	52	make	77	its
3	to	28	say	53	can	78	over
4	of	29	her	54	like	79	think
5	and	30	she	55	time	80	also
6	a	31	or	56	no	81	back
7	in	32	an	57	just	82	after
8	that	33	will	58	him	83	use
9	have	34	my	59	know	84	two
10	I	35	one	60	take	85	how
11	it	36	all	61	people	86	our
12	for	37	would	62	into	87	work
13	not	38	there	63	year	88	first
14	on	39	their	64	your	89	well
15	with	40	what	65	good	90	way
16	he	41	so	66	some	91	even
17	as	42	up	67	could	92	new
18	you	43	out	68	them	93	want
19	do	44	if	69	see	94	because
20	at	45	about	70	other	95	any
21	this	46	who	71	than	96	these
22	but	47	get	72	then	97	give
23	his	48	which	73	now	98	day
24	by	49	go	74	look	99	most
25	from	50	me	75	only	100	us

Considering most frequent 8727 words ($\log_2 8727 = 14.4$ bits),
the entropy of English word is found to be only 9.14 bits/word.

Can we reach this or beyond?

263

Codeword Length

- In a system with 2 symbols that are equally likely:
  - Probability of each symbol to occur: p = ½, $H(p) = 1$ bit
  - Best one can do: encode each with 1 bit only, 0 or 1, $\bar{L} = 1 = H(p)$ bit
- In a system with 2 symbols that are unequally likely:
  - $H(p) < 1$ bit
  - Encode each with 1 bit only, 0 or 1, $\bar{L} = 1 > H(p)$ bit
- A system with N (=2^k for some integer k) symbols that are equally likely:
  - Probability of each symbol to occur: $p = 1/N$
  - One needs $\bar{L} = \log_2 N = \log_2(1/p) = -\log_2 p$ bits to represent the symbols. For example, n=4, L=2 bits
- What is the minimum average codeword length for a particular source?

264

Fixed Length Coding

The same codeword length of different codewords

Example: 4-symbol source with $P\{a_1\} = \frac{1}{2}, P\{a_2\} = \frac{1}{4}, P\{a_3\} = \frac{1}{8}, P\{a_4\} = \frac{1}{8}$.

Encoding table I: $a_1 \to 00, a_2 \to 01, a_3 \to 10, a_4 \to 11$.

(different symbols corresponding to different codewords.)

For example, using table I

$$a_1 a_3 a_4 a_3 \to 00{,}10{,}11{,}10 \to 00101110$$
$$10110100 \to 10{,}11{,}01{,}00 \to a_3 a_4 a_2 a_1$$

Fixed length coding is always uniquely decodable as long as you assign different symbols to different codewords!

Source entropy: $H(S) = \frac{1}{2}\log_2\frac{1}{1/2} + \frac{1}{4}\log_2\frac{1}{1/4} + \frac{1}{8}\log_2\frac{1}{1/8} + \frac{1}{8}\log_2\frac{1}{1/8} = 1.75$ bits

265

Variable Length Coding

Variable length code: codewords may have different codewords

Example: 4-symbol source with $P\{a_1\} = \frac{1}{2}, P\{a_2\} = \frac{1}{4}, P\{a_3\} = \frac{1}{8}, P\{a_4\} = \frac{1}{8}$.

Encoding Table II: $a_1 \to 0, a_2 \to 10, a_3 \to 110, a_4 \to 111$

Encoding using Table II: $a_1 a_3 a_4 a_1 \to 01101110$

Decoding using Table II: $01101110 \to 0{,}110{,}111{,}0 \to a_1 a_3 a_4 a_1$

Encoding Table III: $a_1 \to 0, a_2 \to 11, a_3 \to 110, a_4 \to 111$

Encoding using Table III: $a_1 a_3 a_4 a_1 \to 01101110$

Decoding using Table III: $01101110 \to 0{,}110{,}111{,}0 \to a_1 a_3 a_4 a_1$;

$01101110 \to 0{,}11{,}0{,}111{,}0 \to a_1 a_2 a_1 a_4 a_1$

Not uniquely decodable!!!

Some variable length codes are not uniquely decodable

266

Average Code Length and Coding Theorem

Average Length: For system/source with N symbols, $\{a_n\}_{n=1}^{N}$, the average length of a coding will be

$$\bar{L} = \sum_{n=1}^{N} l_n p_n$$

where l_n is the length of the codeword corresponding to a_n and p_n is the probability of a_n.

Coding Theorem

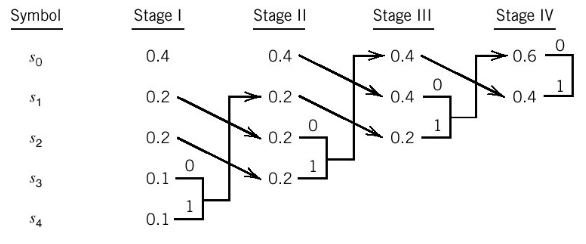
Given a discrete memoryless source of entropy $H(S)$, average codeword length for any uniquely decodable source coding scheme, $\bar{L}$, is bounded by $H(S)$, that is,

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

Huffman Coding

- **Huffman Coding (among other algorithms): uniquely decodable with average coding length $\bar{L}$ satisfying $H(S) \leq \bar{L} < H(S) + 1$.**
- It yields the shortest average codeword length.
- **Basic idea: choose codeword lengths so that more-probable sequences have shorter codewords**
- Huffman Code construction
  - Sort source symbols in order of decreasing probability.
  - Take 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
- 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

- Average codeword length:

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

- Huffman code is not unique (you can reorder equal probabilities)
- Huffman code is the most efficient prefix code
- More than the entropy $H(S)$ = 2.12 bits per symbol.
  ⇒ Room for further improvement.

269

Compound Symbol using Huffman Coding

- Two symbol source: two symbols s_1, s_2
  - probabilities $\Pr\{s_1\} = p_1, \Pr\{s_2\} = p_2$
  - $H_1(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2$
  - Average length of Huffman code: $\mathrm{H}_1(S) \le \overline{L_1} < H_1(S) + 1$
- Compound-symbol source by combining every two symbols:
  - Four compound symbols $s_1s_1, s_1s_2, s_2s_1, s_2s_2$
  - Probabilities
    $$\Pr\{s_1s_1\} = p_1^2, \Pr\{s_1s_2\} = \Pr\{s_2s_1\} = p_1p_2, \Pr\{s_2s_2\} = p_2^2$$
  - Compound-symbol source entropy
    $$H_2(S) = 2H_1(S)$$
  - Average length of Huffman code per the compound-symbol: $\overline{L_2}$
    $$H_2(S) \le \overline{L_2} < H_2(S) + 1$$
  - Average length per symbol: $\overline{L_2}/2$
    $$2H_1(S) \le \overline{L_2} < 2H_1(S) + 1 \rightarrow H_1(S) \le \overline{L_2}/2 < H_1(S) + 1/2$$

270

Compound Symbol using Huffman Coding

- Compound-symbol source by combining K symbols:
  - Probability: $\Pr\{s_{n_1}s_{n_2}\dots s_{n_K}\} = p_{n_1}p_{n_2}\dots p_{n_K} = \prod_{k=1}^{K} p_{n_k}$
  - Compound-symbol source entropy
    $$H_K(S) = KH_1(S)$$
  - Average length of Huffman code per compound symbol: $\overline{L_K}$
    $$H_K(S) \le \overline{L_K} < H_K(S) + 1$$
  - Average length per symbol: $\overline{L_K}/K$
    $$H_1(S) \le \frac{\overline{L_K}}{K} < H_1(S) + \frac{1}{K}$$

- When $K \to \infty$,

$$H_1(S) \le \lim_{K\to\infty} \frac{\overline{L_K}}{K} \le H_1(S) + \lim_{K\to\infty} \frac{1}{K}$$

Average length per symbol: $\lim_{K\to\infty} \frac{\overline{L_K}}{K} = H_1(S)$

271

Application: File Compression

- A drawback of Huffman coding is that it requires knowledge of a probabilistic model, which is not always available a priori.
- 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.

272

Summary

- 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**: For a discrete memoryless source, minimum average codeword length for any source coding scheme is *H*(*S*).

- Huffman coding: An efficient source coding algorithm.

273

Imperial College London

Lecture 13: Channel Capacity and Coding

Professor Geoffrey Li

Department of Electrical and Electronic Engineering

Imperial College London

274

Outline

- Channel capacity
  - Discrete memoryless channels
  - Gaussian channel
  - Channel coding theorem

- Error Correction Coding

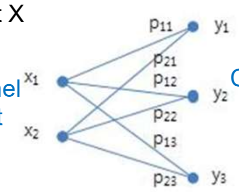
- References
  - Notes of Communication Systems, Chap. 5.5-5.6
  - Haykin & Moher, Communication Systems, 5th ed., Chap. 10

275

Discrete Memoryless Channel

- Channel output Y is a noisy version of channel input X

Channel input



Channel output

- Input alphabet: $X = \{x_0, x_1, \ldots, x_{J-1}\}$
- Output alphabet: $Y = \{y_0, y_1, \ldots, y_{K-1}\}$
- Transition probabilities: $p(y_k \mid x_j) = P(Y = y_k \mid X = x_j)$
- Assume that the input is selected based on $p(x_j) = P(X = x_j) \quad for \quad all \quad j$
- Joint probability distribution is given by $p(x_j, y_k) = p(y_k \mid x_j) p(x_j)$
- Marginal distribution of the channel output is found as

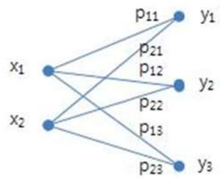
$$\begin{aligned} p(y_k) &= P(Y = y_k) \\ &= \sum_{j=0}^{J-1} p(y_k \mid x_j) p(x_j) \end{aligned}$$

276

Conditional Entropy

- Conditional entropy

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



- This is a random variable that takes values on

$$H(X|Y=y_0), H(X|Y=y_1), \ldots, H(X|Y=y_{K-1})$$

$$p(y_0), p(y_1), \ldots, p(y_{K-1})$$

with probabilities . The mean entropy is given by

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

277

Mutual Information

- Difference $H(X)-H(X|Y)$ is the uncertainty resolved by observing channel output. Define the mutual information as

$$I(X;Y)=H(X)-H(X|Y)$$

- We have

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

- Mutual information is
  - Non-negative: $I(X;Y)\geq 0$
  - Symmetric: $I(X;Y)=I(Y;X)$

- For a given channel, $p(y_k|x_j)$, for $x_1,\ldots,x_J$ and $y_1,\ldots,y_K$, $I(X,Y)$ depends on $p(x_j)$, for $j=1,\ldots,J$.

278

Channel Coding Theorem

- Capacity of a discrete memoryless channel is the maximum mutual information between the input and output, where the maximization is over all possible input probability distributions.

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

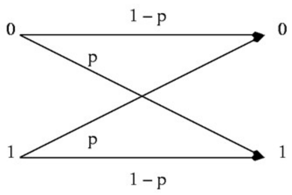
- How to calculate?
  - usually very complicated if analytically, except some symmetrical cases.
  - easily to calculate numerically.

- **If the transmission rate $R \leq C$, then there exists a coding scheme such that R bits per channel use can be transmitted over the channel with an arbitrarily small probability of error.**

- **Conversely, if $R>C$, error probability is always bounded above zero when the transmission rate is above the capacity.**

- How to code? We only know its existence but don't know how.

Binary Symmetric Channel (BSC)



•We should optimize mutual information over binary input distributions:

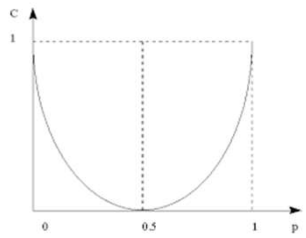
$$C = \max_{p(x_j)} I(X;Y) = 1 - h(p)$$

•We have $I(X;Y) = H(Y) - H(Y|X)$

$$H(Y|X) = p_0 H(Y|X=0) + p_1 H(Y|X=1)$$

$$= p_0 h(p) + p_1 h(p) = h(p)$$

$$h(p) = -p\log_2 p - (1-p)\log_2(1-p)$$



Where .

We then have $p_0 = p_1 = 1/2$

$$I(X;Y) = H(Y) - h(p) \leq 1 - h(p)$$

It is easy to see that this can be achieved by setting:

Additive White Gaussian Noise (AWGN) channel

- Capacity of an additive white Gaussian noise (AWGN) channel:

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

  - B: bandwidth of the channel
  - P: average signal power at the receiver
  - N_0: single-sided PSD of noise

- How can we achieve this rate?
  - Design powerful error correcting codes to correct as many errors as possible.
  - Use good modulation schemes that do not lose information in the detection process.
  - It took around 6o years to reach it, No simple way!

281

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)

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

282

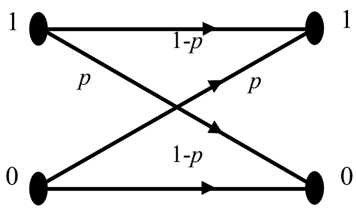
Noise and Errors

- Noise can corrupt the information we wish to transmit.
- Corruption of a signal should be avoided if possible.
- Different systems will generally require different levels of protection against errors.
- Consequently, a number of different channel coding techniques have been developed to detect and correct different types and number of errors.

283

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 (BSC)**.



- Error probabilities are symmetric, errors are stationary and statistically independent.

284

Simple Error Checks

- If the error probability is small and information is fairly fault tolerant, it is possible to use simple methods to detect errors.

- **Repetition** – Repeating each bit in the message
  - If two symbols in an adjacent pair are different, it is likely that an error has occurred.
  - However, this is not very efficient (bit rate is halved).
  - One repetition provides a means for error detection, but not for error correction. More repetitions are needed 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.
  - A single parity bit only allows error detection, not error correction.
  - More efficient than simple repetition.

285

Block Codes

- An important class of codes that can detect and correct some errors are block codes
  - Encode a series of symbols from the source, a 'block', into a longer string: codeword or code block
  - Error detection: if the received coded block is not a valid codeword
  - Error correction: "decode" and associate a corrupted block to a valid coded block by its proximity (as measured by the "Hamming distance")

286

Boolean Algebra

- Two numbers: 0, 1
- Addition, +: $0+0=1+1=0;\ 0+1=1+0=1$
- Multiplication, ×: $0\times 0=0\times 1=1\times 0=0;\ 1\times 1=1$

  Calculation order: same as regular number calculation (multiplication first, from left to right), for example,

  $$1\times 1+1\times 0+0=(1\times 1)+(1\times 0)+0=1+0+0=1$$

- Algebraic in Modulo 2:

  Vector: $a=\begin{pmatrix}1\\0\\1\end{pmatrix}$,

  Matrix: $\mathrm{B}=\begin{pmatrix}1 & 1 & 0\\1 & 0 & 1\end{pmatrix}$,

  $$Ba=\begin{pmatrix}1 & 1 & 0\\1 & 0 & 1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix}=\begin{pmatrix}1\times 1+1\times 0+0\times 1\\1\times 1+0\times 0+1\times 1\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$$

287

Linear Block Codes

- An (n, k) binary linear block code takes a block of k bits of source data and encodes them using n bits.
  - 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**.
  - **Linearity**: the Boolean sum of any two codewords must be another codeword,

    e.g., if $a=(1\quad 0\quad 0)$ and $b=(1\quad 0\quad 1)$ are codewords,

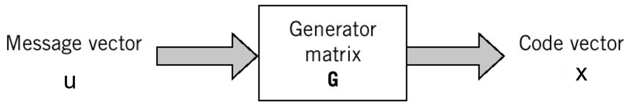
    then $\mathrm{c}=a+b=(1+1\quad 0+0\quad 0+1)=(0\quad 0\quad 1)$ is, too.
  - The set of codewords forms a vector space, within which mathematical operations can be defined and performed.

288

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),

$$x = u \cdot G$$



$$u = (1 \quad 0),\ G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},\ x = uG = (1 \quad 0)\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = (1 \quad 0 \quad 1)$$

- Linearity: Summation of two codewords is another codeword.
  If $x_1 = u_1 G, x_2 = u_2 G$,
  then $x_1 + x_2 = u_1 G + u_2 G = (u_1 + u_2)G$ is another codeword!

289

Hamming Distance

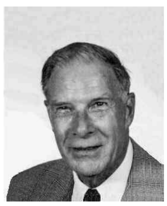
- **Hamming Weight**
  - Hamming weight of a binary vector $\boldsymbol{a}$ (written as $w_H(\boldsymbol{a})$), is the number of non-zero elements it contains.
  - Hence,
    001110011 has a Hamming weight of 5.
    000000000 has a Hamming weight of 0.
- **Hamming Distance**
  - Hamming Distance between 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

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

Richard Hamming (1915 - 1998) established code theory and method when he worked at AT&T Bell Labs, New Jersey, USA



290

Example: (7,4) Hamming Code

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

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

i. Data bits: u=(1010) → code word c=uG=(1010101)
ii. Data bits: u=(1101) → code word c=uG=(1101001)
iii. Data bits: u=(0010) → code word c=uG=(0010110)
iv. Calculate the Hamming distances between each pair of codewords 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]

Example: (7,4) Hamming Code

$$d(\boldsymbol{a},\boldsymbol{b}) = d(0,\boldsymbol{a}+\boldsymbol{b}) = w(\boldsymbol{a}+\boldsymbol{b})$$

Data bits: 1010 → codeword = 1010101
Data bits: 1101 → codeword = 1101001
Distances: 3 4

Data bits: 1010 → codeword = 1010101
Data bits: 0010 → codeword = 0010110
Distances: 1 3

Data bits: 1101 → codeword = 1101001
Data bits: 0010 → codeword = 0010110
Distances: 4 7

(7,4) Hamming Code

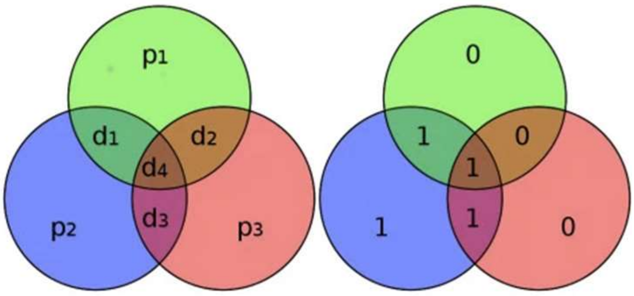
Data bits: $u=(d_1, d_2, d_3, d_4)$

Generate matrix:

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

Codeword: $c=uG=(d_1, d_2, d_3, d_4, d_2+d_3+d_4, d_1+d_3+d_4, d_1+d_2+d_4)$

Parity bits: $p_1=d_1+d_2+d_4$, $p_2=d_1+d_3+d_4$, $p_3=d_2+d_3+d_4$



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. The minimum Hamming distance of a code is the same as minimum weight of non-codewords.
- 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.

Error Detection

- The number of errors that can be detected is then (d_{min}–1), since d_{min} errors may turn an input code word into a different valid code word. Less than d_{min} errors will turn an input code word into a vector that is not a valid code word.
- Number t of errors that can be corrected is $t = \left\lfloor \frac{d_{min}-1}{2} \right\rfloor$, 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 be 'nearer' to the input code word.
- (7,4) Hamming code has d_{min} = 3. It can detect one or two bit errors, and correct any single bit error.

295

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: LDPC, fountain codes, network coding, polar codes in 5G

296