Notes on Digital Communications - NPTEL Video Lectures

Surya Teja Paruchuri*,
* Email: surya@terpmail.umd.edu

***Abstract*—This notes wad designed as a brief review material from my understanding of the Digital Communication Lectures.**

I. Introduction

1) The non-deterministic nature or random nature of the channel is due to the presence of the noise. (Types- Natural like lightning, Man-made like switching.)
2) Predominant noise is Thermal Noise. Random, but have statistical information about this noise. This is modeled as Gaussian Noise, thus resulting in AWGN channel.
3) White noise- Means that there is no correlation between different instances of the sample.
4) source coding is possible because samples might not be equiprobable, or are correlated.
5) advantages of digital communications: Cheaper due to digital electronics which are also programmable, can achieve near capacity rates.
6) An important parameter in Channel is Noise Variance.

II. Sampling

1) Band-limited signal- A signal whose
$$H(f) = 0, for|f| > \omega$$
Similarly, a stochastic process is band-limited if the power spectral density is limited to a certain portion of spectrum.
2) recovering from the samples, sampled at greater than Nyquist's sampling rate:
   a) first interpolate and then pass through an low pass filter.
   b) And the best filter would have an impulse response of the form
$$h(t) = \frac{sin(\pi f_s t)}{\pi f_s t}$$

III. Quantizartion, PCM and Delta Modulation

1) For a uniform quantizer, with a very small Δ and that the PDF of the samples is smooth, then
$$f_Q(q) = \frac{1}{\Delta}$$
and thus
$$E[q^2] = \frac{\Delta^2}{12},$$
which is the energy of the quantization noise.
2) Non-uniform Quantization = non-linear transformer (compressor) + uniform quantizer.
3) The inverse for compressor is expander.
4) logarithmic compressors are popular compressors: μ -law compressor (USA, Canada) and A-Law compressor (Europe).
5) Delta Modulation- uses only 1 bit for representation of the change. But the samples should have high correlation, i.e, the signals doesn't change abruptly. And the distortion in very high slope region is called Slopeoverload distortion, and the distortion in almost constant signal region is called granular distortion.

IV. Probability and Random Processes-1

1) another definition: A,B are independent events if $P(A/B) = P(B)$

V. Probability and Random Processes-2

Functions of Random Variables:

1) x is a RV, and let y=g(x).
2) if y= ax+b;then
$$F_Y(y) = F_X(x \leq \frac{y-b}{a})$$
3) Let $\underline{X}$,$\underline{Y}$be the vector of RVS such that $X1 = f_1(Y1, Y2, Y3, ..YN)$,$X2 = f_2(Y1, Y2, Y3, .., YN)$,$X3 = f_3(Y1, Y2, Y3, .., YN)$,and so on, and also $Y1 = g_1(X1, X2, X3, .., XN)$, $Y2 = g_2(X1, X2, X3, .., XN)$, $Y3 = g_3(X1, X2, X3, .., XN)$,and so on. And let $\underline{F}$,$\underline{G}$ represent the vector of f and g functions, then the Jacobian Matrix is as shown below:
$$J = \begin{vmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_2}{\partial y_1} & \cdots & \frac{\partial f_N}{\partial y_1} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_1}{\partial y_N} & \frac{\partial f_2}{\partial y_N} & \cdots & \frac{\partial f_N}{\partial y_N} \end{vmatrix}$$
then
$$P_{\underline{Y}}(Y1, Y2, Y3, ...YN) = P_{\underline{X}}(\underline{f}(\underline{y})).|J|$$
4) Quick Tip: $|A^{-1}| = \frac{1}{|A|}$.
5) nth Moment:
$$E[X^n] = \int_{-\infty}^{+\infty} x^n.p_X(x)dx$$
6) nth Central Moment:
$$E[(X-\overline{X})^n] = \int_{-\infty}^{+\infty} (x-\overline{X})^n.p_X(x)dx$$
7) Mean is the 1st Moment.
8) Variance
$$\sigma_X^2 = E[(X-\overline{X})^2] = E[X^2] - \overline{X}^2$$
9) Joint Moment: $E(X_1^k, X_2^n)$
10) Covariance (X1,X2):
$$= E[(X1-\overline{X1})(X2-\overline{X2})] = E[X1X2] - \overline{X1}.\overline{X2} \quad (1)$$
11) covariance matrix for X1,X2,X3,...XN:
$$\begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,N} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{N,1} & \mu_{N,2} & \cdots & \mu_{N,N} \end{bmatrix}$$
where $\mu_{(i,j)}$ = covariance of Xi,Xj.

Random or Stochastic Process:

1) Random Processes- types: continuous time and discrete time.
2) Random process or stochastic process is function of random variable with respect to time, i.e, the output random variable changes with respect to time.

3) If t1 ¿ t2 ¿ t3 ¿ …¿ tn, where t is the time, and X_{t1},X_{t2}, X_{t3}, dots X_{tn}, then we define the random process using the joint PDF.
4) if

$$p_{Xt1,Xt2,....Xtn}(xt1, xt2, ...xtn) = p_{Xt1+\delta,Xt2+\delta,....Xtn+\delta}(xt1 + \delta, xt2 + \delta, ...xtn + \delta) \quad (2)$$

, then this process is strictly stationary.
5) Correlation (Auto):

$$\Phi(t1, t2) = E(X_{t1}, X_{t2}) = \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} x_{t1}.x_{t2}p(x_{t1}, x_{t2})dt1dt2 \quad (3)$$

6) if the random process is stationary, then $\Phi(t1, t2) = \Phi(t1 + t, t2 + t)$, where t is some random time duration. This is a function of t1-t2, and thus $\Phi(t1 - t2) = \Phi(t)$. And if some X(t)m which is not strictly stationary, but still satisfies this condition, then X(t) is said to be Wide-Sense Stationary.
7) For a complex valued Random Variable:

$$\Phi(t1, t2) = \frac{1}{2}.E[X_{t1}, X^*_{t2}]$$

8) Random processes X(t), Y(t) are
   a) independent if:

$$p_{X,Y}(x_{t1}, x_{t2}, x_{t3}, \dots, x_{tn}, y_{t1'}y_{t2'}, \dots, y_{tn'}) = p_X(x_{t1}, x_{t2}, \dots, x_{tn}).p_Y(y_{t1}, y_{t2}, \dots, y_{tn}) \quad (4)$$

   b) uncorrelated if:

$$E(x_{t1}, y_{t2}) = E[x_{t1}].E[y_{t2}]\forall t1, t2$$

9) Power Spectral Density:

$$\Phi(\tau) \Longleftrightarrow \Phi(f) = \int_{-\infty}^{\infty}\Phi(t).e^{-j2\pi f\tau}d\tau$$

and at $\tau = 0$,

$$\Phi(0) = \int_{-infty}^{+infty}\Phi(f).df = E(|X_t|^2) \geq 0$$

10) Power Spectral Density of the output signal at a LTI filter:

$$\Phi_{Y,Y}(f) = \Phi_{X,X}(f)|H(f)|^2$$

11) In a White Gaussian Noise, the PSD is 1,which means in time domain, it is an impulse function, and thus clearly there is no correlation between two samples.

VI. Channels and their Models-1

We begin by assuming that $x(t)$ undergose a transformation with a function $g(n)$ and then some noise is added to this signal, to represent the complete channel.

1) Binary symmetric channel (BSC): simplest. TX 0 or 1 and Rx 0 or 1.
   This can be represented by the famous Butterfly diagram. Let Probability of error = p. Then $P(0/1) = p = P(1/0)$ and $P(0/0) = 1 - p = P(1/1)$. We cannot compare two BSC with their p values known,as a simple transformation can make a worse channel better than the other.
2) Binary Channel: Similar to BSC, but $P(0/1) = pand$
   $P(1/0) = q$,
   $wherep \neq q$. Examples: Telegraph channels and LAN channels.
3) Binary Erasure Channel: Bits are not received erroneously, but they are received in either correct form or corrupted form i.e erased form. So the output can be one of the three states 0,1,erased(e). Example: +5V is used for 1 and -5V is used for 0; and then if receiver receives a value close to 0V then it says that the channel is erased. Another use: In puncturing the data.
4) Binary error and erasure channel: Bits could be erased or could be received in error i.e (0/1).
5) Discrete Memoryless Channel:Is a generalization of all the above channels. Discrete because the input alphabet is set of N discrete symbols. Similarly output alphabet is also discrete with say M symbols. Thus we will have a **probability transition matrix** :
   $|P(\frac{y_j}{x_i})|, where 1 \leq j \leq M and 1 \leq i \leq N$
   The present symbols don't depend on the past values.
6) If while transmitting (i+1)th symbol, we have information about what symbol was received during ith symbol transmission, then the channel is modeled as having feedback. Shannon has proved that there is no capacity gain in a feedback channel, compared to a channel without feedback channel
7) Other Channels models: Continuous Channels, Waveform Channels, Linear filter channels/ channels with memory, and wireless fading channels.

VII. Channels and their Models -2

1) Discrete time, but continuous values channels: AWGN channel- $y_i = x_i + z_i$, where z is the white noise, gaussian with variance σ^2, i.e the Z's RV at times are independent and identical distributed RVs.
2) Waveform Channels: Continuous time continuous value channels: continuous Channels-
   $y(t) = x(t) + z(t)$, and z is white noise with variance $N_0/2$.
3) AGN channels with memory: $y_i = x_i + z_i$, but here Z is not white. Thus there is a correlation in time, but still gaussian noise
4) Linear filter channels: Now we consider a channel where we also include a $g(n) or h(n)$ transformation function. Thus we have $y_i = x_i * g_i + z_i$, where z is additive white gaussian noise.
5) AWGN channel can be represented in the form of a Linear filter Channel, and vice versa.
6) Wireless Channel/Fading Channels/Multi-path Channel: components
   a) LOS component.
   b) Reflections.
   c) Difraction.
   d) scattering.
7) For a strictly LOS channel, The channel gain $G \propto \frac{1}{(\frac{d}{\lambda})^2}$.
8) In general $G \propto \frac{1}{(\frac{d}{\lambda})^\mu}$, where $2 \leq \mu \leq 4$.
9) delay spread - The time period between the arrival of first multi-path component and the arrival of the last significant multi-path component. Related to coherence bandwidth.

VIII. Information Theory-1

Elements of Information Theory:

1)

$$Probability of event \propto \frac{1}{Codeword length} \propto \frac{1}{Amount of Information acquired} \quad (5)$$

2) Example: Consider an experiment, where 8 horses race every day and probabilities of winning are 1/2,1/4,1/8,1/16,1/64,1/64,1/64,1/64. The winners information has to be transmitted.
3) The amount of information for a RV $F_X(x)$ is $\frac{1}{\log_\mu P_X(x)}$, where μ is the number of levels in the representation, (2 for binary).
4) Entropy: is the average amount of information of the random variable. Thus
$$H(X) = \sum_i p_X(i) log(\frac{1}{p_X(i)})$$
This is also the average code length of the code word.
5) units of information:
$bits \rightarrow \log_2$
$nats \rightarrow \log$
6) Binary entropy function :
$H(p) = -p\log p - (1-p)\log(1-p)$, where p is the probability of occurrence of say 1.
7) For a random variable to have maximum information (entropy), then the pdf of the RV should be uniform.
8) Properties of the Entropy: Entropy is always a function of the probability density values and not the RV itself.
   a) $H(x) \geq 0$
   b) **Source coding Theorum (by Shannon)**: Let average code word be (L_{av})
   H(x) $\leq L_{av} \leq$ H(x) + 1;
   An entropy of H(x) is achieved by coding in blocks of size N (Ntuples)
   N.H(x) $\leq L_n \leq$ H(x) + $\frac{1}{N}$;
9) Joint Entropy:
$$H(x,y) = \sum_x \sum_y p_{X,Y}(x,y) \log \frac{1}{p_{X,Y}(x,y)}$$
10) For independent RV's X,Y,
$$H(x,y) = H(x) + H(y)$$

IX. Information Theory -2

1) We see that
$$H(y/X=x) = -\sum_y p(Y=y/X=x) \log p(Y=y/X=x)$$
2) Conditional Entropy: Average uncertainty left in Y, after knowing X
$$H(Y/X) = \sum_x p_X(x) H(Y/X=x)$$
This is always less than the total uncertainty in Y.
3) Chain Rule:
$$H(X,Y) = H(X) + H(Y/X)$$
4) Mutual Information between X,Y:
$$I(x,y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x).p(y)}$$
The same can also be defined as
$$I(x,y) = H(X) - H(X/Y) = H(Y) - H(Y/X)$$
5) **Channel Coding Theorem:** I(x,y) is a function p(x) and p(x/y). In a communication systems, with x as input and y as output, p(y/x) is given by the channel, i.e, channel commands this distribution, and p(x) is defined by the system designer. If we Tx x with p(x), and receive y for a given channel's p(y/x), then I(x,y) is fixed; and we want to estimate x from the what we receive, which is the mutual information (the shared region in the Venn diagram of the H(x) and H(y))
Formal definition: For any ϵ >0, and R < I(x,y), if we transmit at a rate R, then average probability of error < ϵ.

We try to optimize by choosing a p(x), which maximizes the mutual information.
6) Coding Theorem: c $\cong maxp(x)I(x,y)$. Converse is "If R > C, then reliable communication is not possible".

X. Band Pass Representation -1

1) Why modulation: Antenna Dimensions, Channel characteristics, Multiplexing.
2) Consider a band-pass signals with f_c as the center frequency. Let s(t) be the time-domain representation of the band-pass signal, and s(f) the frequency domain representation (which has both +ve component and -ve component). The positive component of the spectrum is
$$s_+(f) = 2.u(f).s(f)$$
The time domain representation is
$$s_+(t) = \int_\infty^{+\infty} s_+(f) e^{-j2\pi ft} dt$$
$$F^{-1}[2u(t)] * F^{-1}[s(f)]$$
$$F^{-1}[2u(f)] * s(t)$$
The signal $s_+(t)$ is called the pre-envelope of $s(t)$. We know that $F^{-1}[2u(t)] = \delta(t) + \frac{j}{\pi t}$. Thus
$$s_+(t) = s(t) + \frac{j}{\pi .t} * s(t)$$
The Hilbert Transform of s(t) is $\hat{s}(t) = \frac{1}{\pi .t} * s(t)$ where $\frac{1}{\pi t}$ is called the Hilbert Transformer. The Fourier Transform of Hilbert transform is
$$H(f) = \begin{cases} -j; & \text{if} f > 0 \\ 0; & \text{if} f = 0 \\ +j; & \text{if} f < 0 \end{cases}$$
and $|H(f)| = 1$; and
$$\angle H(f) = \begin{cases} \frac{-\pi}{2}; & \text{for } f > 0 \\ \frac{+\pi}{2}; & \text{for} f < 0 \end{cases}$$
3) Thus Hilbert transformer is a 90 degree phase shifter.
4) Thus $s_+(t)$ is a band-pass filter, and as since it has only +ve part of the spectrum, it is complex signal.
5) Let $s_l(t) = s_+(f + f_c)$, which we get by multiplying with the carrier of f_c, i.e by shifting it to 0 center frequency. Expanding this, we get
$$S_l(t) = [s(t) + j\hat{s}(t)]e^{-j2\pi f_c t}$$
$s_l(t)$ = x(t) + jy(t)is a low-pass and complex signal, where
$$x(t) = s(t)cos(2\pi f_c t) + \hat{s}(t)sin(2\pi f_c t)$$
$$y(t) = -s(t)sin(2\pi f_c t) + \hat{s}(t)cos(2\pi f_c t)$$

6) Here $s_l(t)$ is the low-pass equivalent, which is complex signal of a real, band-pass signal s(t).
7) Thus while designing, we design the $s_l(t)$, and then use this to get the $s(t)$.
8) This converts the problem of designing a signal for real, band pass channel to designing a signal for complex, low pass channel.

XI. Bandpass Signal Representation-2

1) The band-pass signal has an envelope.
2) we have

$$s_l(t) = a(t)e^{j\theta(t)}$$

, where

$$a(t) = \sqrt{x^2(t) + y^2(t)}$$

, and

$$\theta(t) = \tan^{-1}(\frac{y(t)}{x(t)})$$

Thus, s(t) = Re$[s_l(t)e^{j2\pi f_c t}]$
Thus

$$s(t) = a(t)cos(2\pi f_c t + \theta(t)$$

and a(t) is the envelope.
3) Energy conversion in Band-pass signal in terms of a low-pass equivalent signal.

$$\begin{aligned} E &= \int_{-\infty}^{+\infty} s^2(t)dt \\ &= \int_{-\infty}^{+\infty} Re[s_l(t)e^{j2\pi f_c t}]dt. \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} |s_l(t)^1|dt + \frac{1}{2}\int_{-\infty}^{+\infty} s_l(t)^2 \cos 4\pi f_c t + 2\theta(t)dt \end{aligned}$$

Now in the second integral as the frequency is twice, we may assume that the envelope doesn't change much, and thus the integral will be negligible. Thus

$$E \approx \frac{1}{2}\int_{-\infty}^{+\infty} |s_l(t)^2|dt$$

4) The impulse response of the base-band channel is :

$$\begin{aligned} H_+(f) &= \begin{cases} H(f); & \text{if} f > 0 \\ 0; & otherwise \end{cases} \\ &= u(f).H(f) \end{aligned}$$

Thus

$$H_l(t) = H_f(f + f_c)$$

So

$$H_l(f - f_c) = \begin{cases} H(f); & f > 0 \\ 0; & \text{otherwise} \end{cases}$$

Thus,

$$H(f) = H_{f-f_c} + H_l^*(-f - f_c)$$

$$\begin{aligned} h(t) &= h_l(t)e^{j2\pi f_c t} + h_l^*(t)e^{-j2\pi f_c t} \\ &= 2Re[h_l(t)e^{2\pi f_c t}] \end{aligned}$$

5) The Hilbert transform of an even signal is odd symmetric, and for an odd symmetric signal, it's Hilbert transform is even symmetric.

XII. Digital Modulation Techniques-1

1) Types of modulation:
   a) Base-band modulation: if the channel is base-band.
   b) Pass-band modulation: if the channel is pass-band.
2) The simplest modulation is mapping the bits to a voltage level and transmitting a known pulse p(t) with this amplitude.
3) base-band modulated signal equation

$$x(t) = \sum_{\text{k=-}\infty}^{\infty} a_k.p(t - kT)$$

, where p(t)is the base-band modulated pulse.
4) Pass-band modulated signal equation

$$x(t) = \sum_{\text{k=-}\infty}^{\infty} a_k.p^{'}(t - kT)$$

, where $p^{'}(t)$ is the Pass-band modulated pulse, which is a base-band pulse p(t), multiplied with a $\sin/\cos$ signal of high frequency. The above two techniques are Pulse Amplitude Modulation (PAM).
5) When PAM pulse, p(t) is rectangular, then demodulation is as simple as a convolution and sampling at mT intervals.

XIII. Digital Modulation Techniques-2

1) Matched Filter concept: Suppose we have a PAM signal

$$x(t) = \sum_{\text{k=-}\infty}^{\infty} a_k p_1(t - kT)$$

where $p_1(t)$ is the pulse (used to convert the digital binary bits to analog domain signal).Given a channel with impulse response $c(t)$ and that adds a noise $n(t)$. Let us assume we have a receive filter Rx Filter which has an impulse response $a(t)$, whose output is $y(t)$. Also let $r(t)$ be the output after channel effect (this is given to Rx Filter). Lastly, let $p_2(t)$ be the output of the channel if the input is $p_1(t)$. (Also for representation purposes let $p_2(t) = p_1(t) * c(t)$, and $p_2(t) = p_2(t) * a(t)$). Then:

$$\begin{aligned} r(t) &= \sum_{\text{k=-}\infty}^{\infty} a_k p_1(t - kT) * c(t) + n(t) \\ &= \sum_{\text{k=-}\infty}^{\infty} a_k p_2(t - kT) + n(t) \end{aligned}$$

$$\begin{aligned} y(t) &= r(t) * a(t) \\ &= \sum_{\text{k=-}\infty}^{\infty} a_k p_2(t - kT) * a(t) + n(t) * a(t) \\ &= \sum_{\text{k=-}\infty}^{\infty} a_k p_3(t - kT) + n(t) * a(t) \end{aligned}$$

Now sampler samples $y(t)$ multiples of T (i.e kT). By applying a matched filter, we use a Rx Filter whose $a(t) = p_2^*(T - t)$. So for case of proving, let's take consider only a single symbol.

$$\begin{aligned} y((k+1)T) &= \int_{\text{-}\infty}^{+\infty} a((k+1)T - \tau)(a_k p_2(\tau - kT) + n(\tau)d\tau \\ &= a_k \int_{\text{-}\infty}^{+\infty} p_2^*(\tau - kT)p_2(\tau - KT) \\ &+ \int_{\text{-}\infty}^{+\infty} p_2^*(\tau - KT)n(\tau)d\tau \\ &= a_k \int_{\text{-}\infty}^{+\infty} |p_2(t)|^2 dt + Noisecomponent_k \\ &= E.a_k + Noisecomponent_k \end{aligned}$$

2) Sampling the output of Rx Filter at kT gives the maximum value of the corresponding symbol's amplitude a_k. Sampling at any other point in between the symbol will give a lower value of a_k, thus reducing the SNR.
3) The SNR is given by the signal part divided by the noise part.

$$y(T) = a_k \int_{-\infty}^{+\infty} p_2(\tau)a(T-\tau)d\tau + \int_{-\infty}^{+\infty} n(\tau)a(T-\tau)d\tau$$
$$= y_x(T) + y_n(T)$$

4)
$$SNR = \frac{|y_x(T)|^2}{E[|y_n(T)|^2]}$$
$$E[|y_n(T)|^2] = N_0/2 \int_{-\infty}^{+\infty} |a(T-t)|^2 dt$$

XIV. Digital Modulation Techniques-3

1) Proving that matched filter gives the maximum SNR: as per a well known Inequality, Cauchy-Schwartz inequality: for $g_1(t), g_2(t)$ we have
$$\int_{-\infty}^{+\infty} |g_1^*(t)g_2(t)dt|^2 \le \int_{-\infty}^{+\infty} |g_1(t)|^2 dt \int_{-\infty}^{+\infty} |g_2(t)|^2 dt$$
let
$$g_1(t) = a^*(\tau - t) \text{ and, } g_2(t) = p_2(t)$$
we get
$$\int_{-\infty}^{+\infty} |p_2(t)a(T-\tau)d\tau|^2 \le \int_{-\infty}^{+\infty} |p_2(\tau)|^2 d\tau \int_{-\infty}^{+\infty} |a(\tau-T)|^2 d\tau$$
$$\frac{\int_{-\infty}^{+\infty} |p_2(t)a(T-\tau)d\tau|^2}{\int_{-\infty}^{+\infty} |a(\tau-T)|^2 d\tau} \le \int_{-\infty}^{+\infty} |p_2(\tau)|^2 d\tau$$
$$\frac{\int_{-\infty}^{+\infty} |p_2(t)a(T-\tau)d\tau|^2}{\frac{N_0}{2}\int_{-\infty}^{+\infty} |a(\tau-T)|^2 d\tau} \le \frac{2}{N_0}\int_{-\infty}^{+\infty} |p_2(\tau)|^2 d\tau$$
$$SNR \le \frac{2}{N_0}\int_{-\infty}^{+\infty} |p_2(\tau)|^2 d\tau$$
$$SNR \le \frac{2}{N_0} E$$
here E is the energy in the $p_2(t)$
2) As per Cauchy-Schwartz inequality, the equation has maximum value when g_1 is an integer multiple of g_2. Thus for SNR to have maximum SNR, we get matched filter $a(t) = p_2^*(T-t)$.
3) Thus matched filter gives the maximum SNR.
4) Inter Symbol Interference (ISI): considering the effect of transmitting pulses continuously, we get the sampled output as

$$\begin{aligned} y(kT) &= a_k * p_3(kT) + n_1(kT) \\ &= a_k p_3(0) + \sum_{m=-\infty;m\neq k}^{+\infty} a_m p_3(kT - mT) + n_1(kT) \\ &= \text{symbol energy} + ISI(sayg(t) + noise \end{aligned} \tag{6}$$

Thus we should choose $p_1(t)$ such that, we get $p_3(f) = 1$, with a phase of 0.

XV. Digital Modulation Techniques - 4

1) Raised cosine pulse: A pulse of choice for the
$$\frac{\sin\frac{\pi t}{T}}{\frac{\pi t}{T}} \frac{\cos\frac{\pi\alpha t}{T}}{1-4\alpha^2t^2T^2}$$
where $\alpha \ge 0$. $\alpha = 0$ is a special case where it's frequency response is a rect() and in time domain it is sinc function. [Ref the video for it's Fourier transform]. $0 \le \alpha \le 1$
2) Spectrum of above raised cosine pulse:
$$G(f) = \begin{cases} T; & |F| \le \frac{1-\alpha}{2T} \\ Tcos^2\frac{\pi T}{2\alpha}[|f| - \frac{1-\alpha}{2T}]; & \frac{1-\alpha}{2T} \le |f| \le \frac{1+\alpha}{2T} \\ 0; & \frac{1+\alpha}{2T} < |f| \end{cases}$$
3) Unit energy signal is $\phi(t) = \sqrt{\frac{2}{E_p}p(t)\cos\omega_c t}$, which can be used to express the amplitudes of all the symbols, where E_p is the energy of the pulse signal.
4) constellation is the representation of the signal in terms of the unit energy signal.
5) Phase Shift Keying (PSK):
$$x_m(t) = p(t)\cos 2\pi f_c t + \frac{2\pi}{M}m$$
$0 \le m \le M-1$ and $0 \le t \le T$.
6) The energy of each signal is half of the energy of the pulse, thus it a constant energy modulation.
7) If don't want to increase the probability of error in PSK, simply increase the size of the circle (in the constellation). This increases the distance between 2 constellations points and thus maintaining the probability of error while increasing the number of bits
8) Unlike PAM, PSK needs two unit energy vectors to represent's all the constellation diagram.
9) The probability of error is proportional to the distance between two constellation points.

XVI. Digital Modulation Techniques -5

1) Proof that distance to a constellation point is the energy of the signal.
$$\begin{aligned} \text{Energy} = &= \sum_{\text{k}=-\infty}^{\infty} x_m(t)dt \\ &= \sum_{\text{k}=-\infty}^{\infty} (A_m^2\phi_1^2(t) + B_m^2\phi_2^2(t))dt \\ &= A_m^2 + B_m^2 \end{aligned}$$
as energy of ϕ_x is 1 and each are orthogonal. This is the distance between origin and constellation point.
2) distance between constellation points m,n for M point constellation:
$$\begin{aligned} d_{mn}^2 =& |x_m - x_n| \\ &= E_p[1 - \cos\frac{2\pi(m-n)}{M}] \\ d_{min} =& \sqrt{E_p(1-\cos\frac{2\pi}{M})} \end{aligned}$$
3) Quadrature Amplitude Signals (QAM): Joint amplitude modulation of 2 carrier signals.
$$x_m(t) = a_m p(t)\cos 2\pi f_c t + b_m p(t)\sin 2\pi f_c t$$
for $0 \le m \le M$ and $0 \le t \le T$

$$x_m(t) = Re[V_m e^{j\theta_m} p(t) e^{j2\pi f_c t}]$$

, where V_m is the voltage and the θ is the angle of the low-pass equivalent signal.

$$V_{m=\sqrt{a_m^2+b_m^2}}$$

and

$$\theta_m = \tan^{-1} \frac{b_m}{a_m}$$

4) Basic Linear Algebra Concepts:
   a) Field
   b) Vector Space
   c) Linearly independent vectors: If a vector cannot be expressed as a linear combination of the rest, for all vectors in the space, then these vectors are linearly independent.
   d) generating set: a set of vectors is generating if any vector belonging to a vector space can be expressed as a linear combination of the generating set's vectors
   e) The generating function set which are linearly independent is called the basis of vector space V.

XVII. Digital Modulation Techniques-6

1) Nyquist's pulse shaping is used (the Fourier Transform of it should be 1) to avoid Inter Symbol Interference (ISI).
2) Basic Linear Algebra Concepts (Contd.):
   a) Span of a vectors: spanv1,v2,. . . ,vn is a set of vectors all possible linear combinations of the vectors v1,v2,...,vn.
   b) spanv1,v2,v3,.....,vn is $\subset$ Vector Space.
   c) Span of generating vectors is the whole Vector Space.
   d) a vector space U $\subset$ V is called subspace iff a linear combination of two linear vectors in U is also a vector in U.
   e) Length of a vector v =x1,x2,x3,...xn is $|v| = \sqrt{x_1^2 + x_2^2 + x_3^2 + ... + x_n^2}$
   f) distance between 2 vectors d(v,u) = $|v - u| = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + (x_3-y_2)^2 + ... + (x_n-y_2)^2}$
   g) length of vector of functions f:

$$|f| = \sqrt{\sum_{\text{k=-}\infty}^{\infty} ||f(x)^2 dx}$$

   , which is the square root energy of the signal.
   h) We are concerned only about signals/functions that have finite energy, which form a subspace.
   i) distance between two functions/finite energy signals:

$$d(f,g) = |f-g| = \sqrt{\sum_{\text{k=-}\infty}^{\infty} |f(x)-g(x)|^2 dx}$$

   j) Inner product: of two vectors v,u over vector space V of field F defined using following properties:

$$< v, v >= |v|^2$$

$$< \alpha v, v >= \alpha < v, v >$$

$$< v, \alpha u >= \alpha^* < v, u >$$

   for $\mathbb{R}^n$, inner product is defined as

$$< v, u >= \sum_{i=1}^{n} x_i y_i$$

   for $\mathbb{C}^n$, inner product is defined as

$$< v, u >= \sum_{i=1}^{n} x_i y_i^*$$

   k) orthogonality of two functions v,u: iff ¡v,u¿ = 0.
   l) if a vector has length "1" then it is called normal vectors.
   m) set of vectors v1,v2,...vn is orthogonal iff each pair of vectors is orthogonal.
   n) a set of vectors is orthonormal iff it is orthogonal and the vectors are normal vectors.
   o) Deriving a set of orthonormal vectors from a given sets of vectors V=v0,v1,v2,....vn to express V in terms of these orthonormal vectors:

$$\Phi_0 = \frac{v_0}{|v_0|}$$
$$\Phi_1 = \frac{v_1 - < v_1, \Phi_1 >}{|v_1 - < v_1, \Phi_1 > |}$$
$$\Phi_2 = \frac{v_2 - < v_2, \Phi_1 > - < v_2, \Phi_2 >}{|v_2 - < v_2, \Phi_1 > - < v_2, \Phi_2 > |}$$
$$\vdots = \vdots$$

XVIII. Digital Modulation Techniques-7

1) Frequency Shift Keying (FSK):

$$x_m(t) = \sqrt{\frac{2E}{T}} \cos 2\pi f_c t + 2\pi \Delta f.m.t$$

2) $< x_m(t), x_n(t) >= Esinc(2T\Delta f(m-n))$ is a sinc function of Δ, where E is the energy.
3) For what all Δ's the inner product is zero, i.e., for say m-n=1, we have

$$E.sinc(2T\Delta f) = 0$$

   only if $\frac{1}{2T}$ is divides Δ f

   FSK signal set is orthogonal if $\frac{1}{2T}$ divides Δ f
4) Properties of Orthogonal FSK:
   a) Constant energy modulation
   b) no of constellation points M = N, dimension of signal space.
   c) Bandwidth increases $\propto$ M
   d) Energy and minimum distance $\sqrt{2E}$can be kept constant while no of constellation points increases.
5) Pulse Position Modulation (PPM): p(t) $\rightarrow$ pulse, then

$$x_m(t) = \sqrt{\frac{E}{E_p} p(t - m\Delta t)}; 0 \leq m \leq T-1$$

6) PPM is constant energy, M=N, BW $\propto$ to M.

XIX. Digital Modulation Techniques -8

1) Biorthogonal signals set: if

$$\Phi_1(t), \Phi_2(t), \Phi_3(t).., \Phi_n(t)$$

   is a set of orthonormal functions. If we then have orthogonal signals

$$\sqrt{E}\Phi_1(t), \sqrt{E}\Phi_2(t), \sqrt{E}\Phi_3(t).., \sqrt{E}\Phi_n(t)$$

   whose energy is E, then biorthogonal set of signals are

$$\sqrt{E}\Phi_1(t), .., \sqrt{E}\Phi_n(t), -\sqrt{E}\Phi_1(t), .., -\sqrt{E}\Phi_n(t)$$

2) Properties of Biorthogonal signals: Contant energy, distance between constellation pairs is not uniform, as M increases M increases.

3) PPM is used for ultrawide band application.
4) <u,v >= (<v,u >)*
5) Main types of demodulation: Correlation Receiver and Matched filter receiver.

XX. Digital Modulation Techniques-9

1) Tranformations on signals that retain the equivalence of modified signal to original signal set.
   a) Translation: reduces the average energy used for transmitted
   b) The signal set translated such that the centroid is translated to origin has the lowest Tx energy requirements.
2) Properties of the translated signal:
   a) Energy of the simplex set is less than that of the orthogonal signal.
   b) Inner product of the simplex signal set m,n is $-\dfrac{1}{M}$

XXI. Digital Modulation Techniques-10

1) Why achieving zero ISI is not possible.
   a) Channel may not be known at the time of design
   b) It may be expensive to implement p(t), a(t) to avoid ISI completely.
2) Eye Diagram: To visualize the ISI on an oscilloscope. The output of Rx filter is observed on an oscilloscope with symbol timing as trigger.[REF the video for the explanation at 12:00 minutes]
3) Immunity to noise (a) in eye diagram is the distance between the two levels.
4) Sampling timing error immunity (b): if b is large, then the effect of sampling at different time is better.
5) As the levels increases, the no of levels in the eye diagram also increase.
6) we can model channel and modulator and de-modulator together as a discrete channel time channel. If we also include the mapper, then it discrete in time and amplitude (Such a model can be used while developing Error correction codes).
7) Mapper: The best way to design a mapper is to minimize the bit errors such that even if a symbol is detecting incorrectly. That is the difference between the bits of neighboring symbols should be minimal. (Gray Codes: A maximum difference of 1 bit between neighbors)
   Gray code Example: 000,001,011,010,110,111,101,100. (can be used for 8-PSK)
   This kind of mapping results in a $P_s \approx P_b$, where P_s, P_b are probability of symbol error, and Probability of bit error respectively.
8) Gray coding for 16-QAM:

0000	0001	0011	0010
0100	0101	0111	0110
1100	1101	1111	1110
1000	1001	1011	1010

9) There is trade-off between Bandwidth, Average Energy, Bit Rate and Probability of Error.

XXII. Probability of Error-1

1) Probability of error depends on the Receiver. (We assume a system with matched filter which was already proved to be optimal). And channel is just having AWGN noise with a variance of $\dfrac{N_0}{2}$

$$\begin{aligned} y_i &= \int_0^T y(t)\Phi_i dt \\ &= \int_0^T x_m(t)\Phi_i dt + \int_0^T n(t)\Phi_i dt ** \\ &= x_{mi} + ni \end{aligned}$$

** assuming $x_m(t)$ was transmitted, and n_i is the output of filter $\Phi_i(T-t)$ when excited by n(t).

n_i is Gaussian because n(t) is also Gaussian. Also n_i has zero mean because n(t) is zero mean. We shall compute variance and co-variance between different pairs.

$$En_i, n_j = E\left\{\int_0^T n(t)\Phi_i(t)dt \int_0^T n(\tau)\Phi_j(\tau)d\tau\right\}$$

as the mean is 0

$$\begin{aligned} &= E\left\{\int_0^T \int_0^T n(t)n(\tau)\Phi_i(t)\Phi_j(\tau)dtd\tau\right\} \\ &= \int_0^T \int_0^T E\left\{n(t)n(\tau)\right\}\Phi_i(t)\Phi_j(\tau)dtd\tau \end{aligned}$$

here the $E\left\{n(t)n(\tau)\right\}$

is autocorrelation at $t-\tau$

$$\begin{aligned} &= \int_0^T \int_0^T \frac{N_o}{2}\delta(t-\tau)E\left\{n(t)n(\tau)\right\} \\ &= \frac{N_o}{2}\int_o^T \Phi_i(t)\left(\int_o^T \Phi_j(t)d\tau\right)dt \\ &= \frac{N_o}{2}\int_o^T \Phi_i(t)\Phi_j(t)dt \\ &= \frac{N_o}{2}\delta_{i,j} \end{aligned}$$

Thus the elements of the noise component vector along n0,n1,....nN-1 along different basis functions are independent and thus if x_m is transmitted, $y_m = x_m + n_0 + n_1 \cdots + n_{N-1}$

2) **Probability of error for binary PAM**: Average Probability of error is P_{ave} in a binary PAM with d as the distance between the two constellation points:

$$\begin{aligned} P_{ave} &= \frac{1}{1}P\left(\frac{y<0}{x=d/2}\right) + \frac{1}{2}P\left(\frac{y>0}{x=-d/2}\right) \\ &= \frac{1}{1}P\left(\frac{n<-d/2}{x=d/2}\right) + \frac{1}{2}P\left(\frac{n>d/2}{x=-d/2}\right) \\ &= P\left(n>d/2\right) \\ &= Q\left(\frac{d}{\sqrt{2N_0}}\right) \end{aligned}$$

as d= $2\sqrt{Energy_{symbol}}$

$$= Q\left(\sqrt{\frac{2E}{N_0}}\right)$$

3) $Q(x) < e^{\frac{-x^2}{2}}$ gives upper bound.
4) for large number of bit transmitted, the BER $\approx$ Probability of error.

XXIII. Probability of Error -2

1) Probability of error: Binary FSK or binary orthogonal signal set

$$P_e = Q\left(\sqrt{\frac{E_{symbol}}{N_0}}\right)$$

2) While calculating the probability of error, we should account only for those noise components that would shift it outside the decision region. Thus if a constellation point is represented in 2-D, and one half of the plane corresponds to a constellation point, We should calculate the probability of error using only those component that would move it outside the decision region.
3) In 16-QAM we can calculate the probability of error in terms of distance and then substitute the energy using the relation between average energy and distance
4) Probability of error for orthogonal signals can be computed as follows:
   a) Since N=M here, the energy is of the signal transmitted is $\sqrt{E}$
   b) The probabilities of error is same for all constellation points. Thus let's say S1 is transmitted, and we received α at Rx, then the average probability of error for a given α is simply the product of probability of noise $>\alpha$
   c) Then this probability should also be integrated for all values of α with the corresponding probability of occurrences of α.

XXIV. Probability of Error -3

1) Probability of Error for Bi-Orthogonal: We detect the symbols by simply considering the absolute values of the vectors components.
2) A reasonably good upper bound on Probability of error for bi-orthogonal modulation is computed using the union bound shown below:

   Suppose E1,E2,...En are events with probabilities then Union bound says that

$$P_r\left(\bigcup_{i=1}^{t} E_i\right) \leq \sum_{i=1}^{t} P_r\left(E_i\right)$$

3) *Modulated with Memory:*
   a) consider some Base-band modulation. (where 1 is represented using some +ve voltage and 0 with 0V). Here the signal values returns to 0, whenever we transmit a 0.
   b) The NRZ scheme is where we change the voltage on transmission of 1 and we don't change the voltage if we need to transmit 0. This a memory modulation.
   c) Example of memory coding: Differential Encoding: $b_k = a_k \oplus b_{k-1}$, where a represents the information bit and b represents the Tx bit. This is an example in base-band memory modulation.
   d) Another example in Pass-band memory modulation: Differential PSK (DPSK). Let's consider DQPSK- where we change the phase when there is change in transmitted symbol.
   e) Demodulation of memory signals: can be done by identifying the phase and then canceling/subtracting the previous phase. Another way is to decode using IQ components and then multiplying the previous symbol's signals. The angle of the product gives the phase difference.

XXV. Equalizers

1) Consider the discrete model of the communication system, equalizer is placed after the sampled signals are available. The equalizer design is based one of the following 3 criteria:
   a) Peak distortion criteria: Minimize the worst case error $|I_k - \hat{I}_k|$ in absence of noise. The equalizer filter's response is
$$C(z) = \frac{1}{x(z)}$$
   ,where c(z) is the equalizer filter, and x(z) is the channel response of the filter. This is also known as zero forcing.

   While this completely eliminates ISI, The drawback of the peak distortion criteria is that it amplifies the noise.
   b) Minimum Square Error criteria (MSE):Unlike peak criterion, here we minimize the expectation of the mean square of the error.
$$\varepsilon_k = I_k - \hat{I}_k$$
$$\text{MSE J:} = Expectation|\varepsilon_k|^2$$
$$= Expectation|I_k - \hat{I}_k|^2$$
$$C(z) = \frac{1}{x(z) + N_0}$$
   where$N_0/2$is the variance of noise.
   c) Decision Feedback Equalizer: To use ISI effect in the already detected symbols for better detection of incoming symbols. This would reduce the order of the matched filter. The same is done in other direction also, i.e, to use the information from incoming stream which could carry information about already detected symbols to improve the decision of the already detected symbols. Components of the are:
      i) Feed forward filter
      ii) Symbol detector
      iii) feedback filter
$$\hat{I}_k = \sum_{j=-k_1}^{0} c_j y_{k-j} + \sum_{j=1}^{k_2} \acute{c}_j \tilde{I}_{k-j}$$
   The filter can be designed based on again Peak distortion and MSE techniques.
2) Adaptive Equalization: adapting the filtering based on channel variations.
3) Sequence Estimation: A different approach - not quite the same as equalizer. Unlike in equalization, where the received signal is modified and then symbol information is detected, Here we estimate the sequence that is mostly likely to generate the received sequence. (similar to Maximum likelihood detection, where for symbol it is as $\hat{x} = argmax_x f(\frac{y}{x})$.
   Vitterbi Algorithm: While the above algorithm says that we have to wait for all the symbols to be received to decode even the first symbol, Vitterbi algorithm does enable us to decode the symbols with some fixed delay. Vitterbi algorithm is can also useful for decoding a family of codes like block codes and convolution codes. Particularly, they are used/famous for decoding convolution codes where a transmitted signal is passed through a filter to encode and to decode, Vitterbi algorithm is used to decode the received sequence.

XXVI. Source coding -1

1) The need for source coding:
   a) To remove the redundant information.
   b) To reduce the source's data output rate to match to the channel's capacity.
   c) Non-uniform distribution of the source.

2) sampling at or higher than Nyquist's criteria - lossless source coding (Winzip, Compress etc,), where as reducing the information to fit onto the channel is lossy source coding such as quantization, vector quantization, subband coding (JPEG, MPEG).
3) Lossless Source coding for Discrete Memory-Less Sources:
   a) A discrete memory less source generates symbols from the set $X_i \in \{X_0, X_1, \ldots, X_{M-1}\}$in time, each generated symbol is a RV and each symbol in the set has a known probability of occurrence $PX_i = x_i$.
   b) To make a sampled signal memoryless, we need to sample the signal at a rate lower than Nyquist's rate, else sampling it higher rate will result in signal samples with memory. So we should sample at exactly Nyquist's rate to also be able to reconstruct the signal back.
   c) Desired properties of source codes:
      i) non-singular codes: $x_i \neq x_j \Rightarrow C(x_i) \neq C(x_j)$. All the symbols are encoded by different code words (doesn't mean that these are always decodable). Singular codes are not useful.
      ii) uniquely decodable:

$$x_1x_2x_3 \ldots x_m \neq x_1' x_2' \ldots x_m'$$
$$\Rightarrow C(x_1)C(x_2) \ldots C(x_m) \neq C(x_1')C(x_2') \ldots C(x_m')$$

      Ex: $\{10, 00, 11, 110\}$
      iii) Prefix or Prefix-free codes or instantaneous codes: We cannot decode a uniquely decodable code instantaneously. We have to wait for future bits. This is resolved by prefix codes - any codeword should not be a prefix of any other codeword. Ex: $\{0, 10, 110, 1110\}$
4) **Huffmann Codes:**
   a) Arrange the symbols in decreasing probability.
   b) Assign 0 and 1 to the last two symbols in the list. (If any of the last probabilities is a sum of two or more symbols then the bit assigned should be added to all the symbols individually).
   c) Then combine the probabilities of the last two symbols and rearrange the probabilities in the decreasing order.
   d) Repeat until the sum probability is 1.
5) Properties of Huffmann Code: Is a prefix code, optimum code - no code better than huffmann code for any RV.
6) A quick way to identify huffmann codes: There will be atleast 2 codes of equal length which only differ in the last two bits.

XXVII. Source coding -2

1)

Theorem 1. *Kraft Inequality: A prefix code with codeword lengths $l_1, l_2, l_3 \ldots l_M$ exists if and only if*

$$\sum_i 2^{-l_i} \leq 1$$

Proof. §1. Necessity: The above condition is necessary for any code to be prefix-free.
Consider Codeword $C(x_i)$ of length $= l_i$. Let the max codeword length be l_{max}. Then codeword $C(x_i)$ disqualifies $2^{l_{max}-l_i}$ codes on the l_{max} level of binary expansion. As the descendants are disjoint, all the disqualified codes at l_{max} level are

$$\sum_i 2^{l_{max}-l_i} \leq 2^{l_{max}}$$

Thus, by both sides by $2^{l_{max}}$ we get,

$$\sum_i 2^{-l_i} \leq 1$$

§2.Sufficiency: The above condition is sufficient for existence of the prefix-free codes.
Given the lengths $l_1, l_2, l_3 \ldots l_M$ in increasing order which satisfy kraft's inequality. For a codeword length, we pick a codeword in binary expansion at the level equal to the codeword length. As we pick a codeword, it disqualifies some descendants and thus some at the l_{max} level. So if for a given codeword length l_i if we cannot find any codeword at l_i level, then all the l_{max} codes are also disqualified by definition. Thus the sum of additional terms of 2^{l_i+} would exceed 1, which is against the given statement. Thus if kraft's inequality is satisfied, then a prefix code should be constructed for the given lengths.

□

2) Kraft's inequality also holds for uniquely decodable codes. ().
3) Equality means the codes exhaust all the nodes at the l_{max} level in binary expansion.
4) Optimal Codes: a code for RV X if there is no code for the same RV with smaller average length.

Theorem 2. *Let L be the average length of any of the possible optimal code. Then*

$$H(x) \leq L \leq H(x) + 1$$

Proof. §1. For any uniquely decodable code, the expected length $L(C, X) \geq H(X)$

Let $z = \sum_i 2^{-l_i}$and $q_i = 2^{-l_i}/z$.

Known inequality: $\sum_i p_i log(\frac{1}{q_i}) \geq \sum_i p_i log(\frac{1}{p_i})$

And the above inequality becomes an equality if$p_i = q_i$ (7)

Where q_i is like a probability mass function. (We take the known equality without proof).

$$\begin{aligned} L(C, X) &= \sum_i p_i l_i \\ &= \sum_i p_i \left(\log \frac{1}{q_i} - \log z \right) \\ &\geq \sum_i p_i \left(\log \frac{1}{p_i} - \log z \right) \\ &\geq H(x) \end{aligned}$$

Thus L is greater than H(x) for decodable code. We can achieve maximum compression when the inequality becomes an equality, which can be achieved only if

$$log(z) = 0 \Rightarrow \sum_i 2^{-l_i} = 1$$

i.e equality in Kraft inequality and (8)

$$p_i = q_i = 2^{l_i} \text{(since z=1)} \Rightarrow l_i = \log \frac{1}{p_i}$$

§2. There exists code with $L(C,X) \leq H(x)+1$.
For all i define

$$l_i = \left\lceil \log_2 \frac{1}{p_i} \right\rceil$$

Kraft inequality is satisfied by the above defined $l_i s$:

$$= \sum_i 2^{-l_i}$$
$$= \sum_i 2^{\log_2 \frac{1}{p_i}}$$
$$= \sum_i p_i = 1$$

There is a prefix code with these $l_i s$.

$$\begin{aligned} L(C,X) &= \sum_i p_i \left\lceil \log \frac{1}{p_i} \right\rceil \\ &\leq \sum_i p_i \left(\log \frac{1}{p_i} + 1 \right) \\ &= H(X) + 1 \end{aligned} \tag{9}$$

□

XXVIII. Source Coding -3

1) **Block-wise source coding:** Better compression can be achieved by coding many symbols together as a block. Instead of coding symbol by symbol, we take sequence of n symbols. Thus we have 2^n possible sequences each with a probability. Now code these sequences.

   a) consider a sequence of n symbols $\{X_1, X_1, X_2 \ldots, X_n\}$, then the entropy of the sequence is $nH(x)$ as the RVs are IID.
   b) Let the optimal code for this sequence is $nH(x) \leq L_{sqe} \leq nH(x)+1$
   c) the average code length $L = \frac{L_{seq}}{n}$, thus

$$H(x) \leq L \leq H(x) = \frac{1}{n}$$

2) The no of values for which probabilities in a block codes have to be computed are m^n, where there are m possible symbols and the block size of the block codes is n
3) as $n \to \infty$, the number of probabilities also tend to ∞
4) Block codes cannot be applied to scenarios with unknown statistics.
5) Shannon-Fano-Elias Code (Principle of arithmetic coding) : Here we use the cummulative probability distribution function of the symbols. Then for each symbol$s_i$we code the midpoint between symbol s_i and s_{i-1} using $l_{s_i} = \left\lceil \log \frac{1}{p(x)} \right\rceil + 1$ bits (Thus we truncate bits after binary point to a maximum of l_{s_i} bits).

   NOTE: The above coded number will be the number just below the midpoint in the $2^{l(x)}$ levels between s_i and s_{i-1} symbols

   Now step size in $l(x)$ -bit quantization is [0,λ] is

$$\begin{aligned} 2^{-l(x)} &= 2^{-\left\lceil \log \frac{1}{p(x)} \right\rceil - 1} \\ &\leq 2^{-\log \frac{1}{p(x)} - 1} \\ &= \frac{p(x)}{2} \end{aligned} \tag{10}$$

   Thus quantization error is less than half-of the probability at x, p(x) or $p(s_i)$. Thus we can say that the coded word for symbol s_i will never cross beyond $p(s_{i-1})$.
6) Average code length for Shannon-Fano-Elias Code is H(x) + 2 (sub-optimal code by itself, but doing a block coding of SFE code gives average code of H(x) as n $\to \infty$. That is it is asymptotically optimal).

XXIX. Source coding -4

1) Arithmetic coding : Basically SFE code for a large block length and iteratively coding. Does this sequentially, and thus reduces the delay in coding. For i tuples (arranged in lexicographical order):

$$\begin{aligned} F(x) &= \sum_{y^{(i)} \leq x^{(i)}} P(y^{(i)}) \\ &= \sum_{y^{(i-1)} \leq x^{(i-1)}} P(y^{(i-1)}) + \sum_{y_i \leq x_i} P(x^{(i-1)}.x_i) \\ &= F(x^{i-1} - 1) + \sum_{y_i \leq x_i} P(x^{(i-1)}) P\left(x_i | \boldsymbol{x}^{(i-1)}\right) \\ &= F(\boldsymbol{x}^{(i-1)} - 1) + P(\boldsymbol{x}^{(i-)}) F(x_i | \boldsymbol{x}^{(i-1)}) \end{aligned} \tag{11}$$

   Thus all we need to compute at i^ht script is the cumulative probability term (the last term in the above equation), which helps in coding iteratively

   The code string actually coded will represent a sub-interval (a,b) in $(F(\boldsymbol{x}^{(i)} - 1), F(\boldsymbol{x}^{(i)}))$. Thus we use 'a' as the code word. And computing the $P\left(x_i | \boldsymbol{x}^{(i-1)}\right)$ is the key, which becomes simple for special cases:

   a) I.I.D source $P\left(x_i | \boldsymbol{x}^{(i-1)}\right) = P(x_i)$
   b) Markov sources $P\left(x_i | \boldsymbol{x}^{(i-1)}\right) = P(x_i | x_{i-1})$
   c) For sources with unknown statistics, we use models such as :

      i) Laplace Model:
$$P\left(X_i = a | \boldsymbol{x}^{(i-1)}\right) = \frac{F_a + 1}{\sum_b (F_b + 1)}$$
      ii) Drichlet model:
$$P\left(X_i = a | \boldsymbol{x}^{(i-1)}\right) = \frac{F_a + a}{\sum_b (F_b + a)}$$

2) Lempel-Zei Code: (- doesn't explicitly need probabilities to construct the code.) Here as bits are received, we place them in the buffer and compare if the string in the buffer is present in the list. If the string in buffer is present in the list, we shall wait for the next bit to arrive and use the updated bit sequence to look for a match in the list. We do this till the last arrive bit makes a new sequence not present in the list. In such case, we will use the index of the string minus the last most bit in the buffer and encode that index and the last bit together against the complete sequence. This bit is transmitted and also stored

in the list. (Note: Here the meaning of the last bit in buffer or list is the most recently received bit in stream).

XXX. Channel Coding

1) Channel coding: To introduce redundancy to be able to correct errors and distortion introduced by the channel.
2) Purpose of Channel coding:
   a) TO reduce the probability of error for finite block of symbols.
   b) To approach channel capacity, we need codes of large length.
3) The ability to correct errors introduced by channel reduces the probability of error.
4) For a BSC channel with p=0.25, then the capacity of the channel is 0.19, thus for 100 bits, only 19 bits can carry information. Thus we should construct code words of length 100 bits which can carry a maximum of 19 bits. Even then, recovering all the information is not feasible because of errors introduced by channel. Thus there is still a probability of error. Now Shannon-Schwartz equation says that as the block size increases from 100 to 1000, this probability of error decreases.
5) Code: set of all code words; code word: an element transmitted
6) code-rate: no of information bits/ no of bits transmitted.
7) *Repetition codes:* Repeating a bit for $n-1$ times. Code:
$$\left\{ 00000\ldots 0, 11111\ldots 1 \right\}$$
Code rate = $\dfrac{1}{n}$. Number of errors that can be corrected:
$$\left\lfloor \frac{n}{2} \right\rfloor$$
8) *Parity Check codes:* n-1 information bits and a parity bit. Code:
$$\left\{ x = (x_0, x_1, \ldots, x_{n-1}) | \textstyle\sum_{i=0}^{n-1} x_i = 0 \right\}$$
where each $x_i \in \{0,1\}$ and in binary $x_0 + x_1 \rightarrow XOR$ and $x_0 + x_1 \rightarrow AND$. Here we can detect upto 1 error and can not correct any errors. Rate of code
$$= \frac{n-1}{n}$$
9) Hamming distance $d_H(\underline{x}, \underline{y})$: for n-tuple $\underline{x}, \underline{y}$ where $\underline{x} = (x_0, x_1, \ldots, x_{n-1})$ and $\underline{y} = (y_0, y_1, \ldots, y_{n-1})$ is number of components that are different in $\underline{x}, \underline{y}$
10)
$$d_{min}(C) = mind_H(\underline{x}, \underline{y})$$
where $\underline{x} \neq \underline{y}$ and C is the code.
11) For repetition codes, $d_{min} = n$, $rate = \frac{1}{n}$and errors that can be corrected $t = \left\lfloor \frac{n}{2} \right\rfloor$
12) For parity $d_{min} = 2$, $rate = \frac{n-1}{n}$
13) Generally, we can correct all the errors in a codeword from code C, if the number of errors $t \leq \left\lfloor \dfrac{d_{min}-1}{2} \right\rfloor$
14) Consider a field $\left\{ \substack{\mathbb{F}\rightarrow \\ 0,1} \right\}$ with + and * as the operators.
15) Hamming codes: (7,4) Hamming code- consider all the non-zero vectors of length 3 written in columns as below:
$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$
$$C = \{\underline{x} = (x_0, x_1, \ldots, x_6) | H\underline{x}^T = \underline{0}\}$$
We know that we can correct max 1 bit error using hamming codes. Thus d_{min} = 3. Here the dimension of C = 4 (all the vectors whose product results in 0; thus no of solutions = matrix dimension - rank; thus C = 4). So we have 2^4 codes, and rate is $\frac{4}{7}$
16) Some other popular codes are Reed-Solomon codes, BCH codes, Convolution codes, LDPC codes

XXXI. Introduction to OFDM

1) DTFT of $x[k]$ is
$$X(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} x_k e^{-j\omega k}$$
2) DTFT is complex in general, even for real signals
3)
$$\delta[n] \overset{\text{DTFT}}{\Longrightarrow} 1$$
with a phase of zero.
4) Inverse DTFT :
$$x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega$$
[The definition of DTFT may not converge].
5) DFT: M-length DFT of x[n] is defined as the samples of the DTFT:
$$\begin{aligned} X_k &= \frac{1}{\sqrt{M}} X(e^{\frac{j2\pi k}{M}}) \\ &= \frac{1}{\sqrt{M}} \sum_{i=-\infty}^{+\infty} x_i e^{\frac{-j2\pi k}{M}} \end{aligned} \tag{12}$$
6) Inverse DFT:
$$x[n] = \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} e^{\frac{j2\pi k}{M}}$$
only if the original time domain signal had less than M components before sampling the DTFT.
7) The DFT matrix is given by
$$D_{m,n} = \left(e^{\dfrac{-j2\pi mn}{M}} \right)$$
where $m, n \rightarrow o, 1, 2, \ldots, x_{M-1}$
8) Cyclic convolution: of x and y is defined as z, where
$$z_k = \sum_{i=0}^{M-1} x_i y_{k-i}$$
, where k-i is taken modulo M
9) Cyclic shift: If y is the cyclic shift of x i.e, $y_i = x_{i-1}$((i-1) is taken modulo M) for all i, then
$$Y_k = e^{-j2\pi k/M} X_k$$
10) $length(x|n| * y|n|) = length(x|n|) + length(y|n|) - 1$
11) Unitary Matrix U: For a complex U, if U^*- conjugate transpose of U is also its inverse
$$U^* U = I$$
12) another advantage of cyclic prefix: Cyclic convolution can be used.
13) Power allocation: Using Water pouring method - The noise variance in the i-th sub-channel is given $dfrac\sigma^2 |C_1|^2$, where σ^2 is the received noise variance and C_i is the ith coefficient of the channel impulse response, then power if inverse of $dfrac\sigma^2 |C_1|^2$.

XXXII. Conclusion

1) Synchronization issues (Carrier, phase and symbol)
2) All the content in this material is about point-to-point communication.
3) Interference from other users.