DISCRETE MEMORYLESS CHANNEL

- **Transmission rate over a noisy channel**

Repetition code

Transmission rate

- **Capacity of DMC**

Capacity of a noisy channel Examples

➢ All these transition probabilities from xi to yj are gathered in a transition matrix.

➢ The (i ; j) entry of the matrix is P(Y = yj /jX = xi), which is called forward transition probability.

➢ In DMC the output of the channel depends only on the input of the channel at the same instant and not on the input before or after.

➢ The input of a DMC is a RV (random variable) X who selects its value from adiscrete limited set X.

➢ The cardinality of X is the number of the point in the used constellation.

➢ In an ideal channel, the output is equal to the input.

➢ In a non-ideal channel, the output can be different from the input with a given probability.

- **Transmission rate:**

➢ H(X) is the amount of information per symbol at the input of the channel.

➢ H(Y) is the amount of information per symbol at the output of the channel.

➢ H(XjY) is the amount of uncertainty remaining on X knowing Y .

➢ The information transmission is given by:I (X; Y) = H(X) − H(XjY) bits/channel use

➢ For an ideal channel X = Y , there is no uncertainty over X when we observe Y . So all the information is transmitted for each channel use: I (X;Y) = H(X)

➢ If the channel is too noisy, X and Y are independent. So the uncertainty over X remains the same knowing or not Y , i.e. no information passes through the channel: I (X; Y) = 0.

- **Hard and soft decision:**

➢ Normally the size of constellation at the input and at the output are the same, i.e., jXj = jYj

- In this case the receiver employs hard-decision decoding.
- It means that the decoder makes a decision about the transmitted symbol.

- It is possible also that jXj 6= jY j.
- In this case the receiver employs a soft-decision.

Channel models and channel capacity:

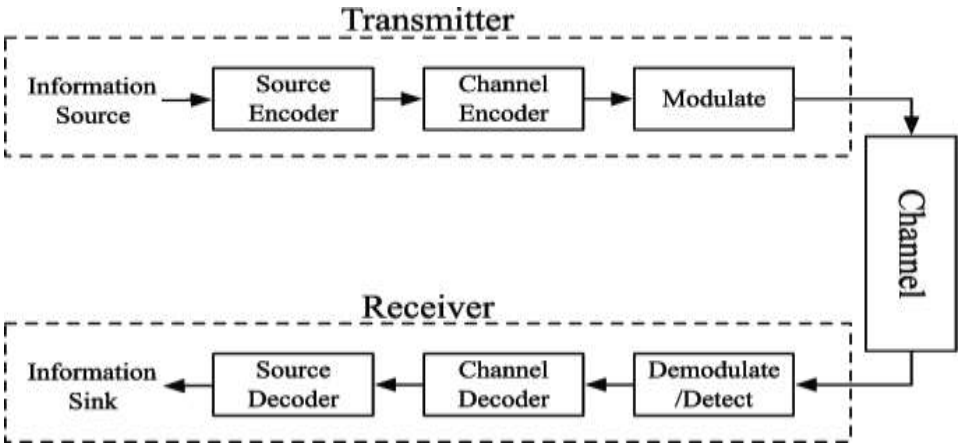

Fig1.2 Block Diagram of Digital Communication System

(Source: https://www.researchgate.net/figure/Block-diagram-of-a-typical-communication-system-doi101371-journalpone0082935g001_fig15_259457178)

1. The encoding process is a process that takes a k information bits at a time and maps each k-bit sequence into a unique n-bit sequence. Such an n-bit sequence is called a code word.

2. The code rate is defined as k/n.

3. If the transmitted symbols are M-ary (for example, M levels), and at the receiver the output of the detector, which follows the demodulator, has an estimate of the transmitted data symbol with

(a). M levels, the same as that of the transmitted symbols, then we say the detector has made a hard decision;

(b). Q levels, Q being greater than M, then we say the detector has made a soft decision.

Channels models:

1. Binary symmetric channel (BSC):

If (a) the channel is an additive noise channel, and (b) the modulator and demodulator/detector are included as parts of the channel. Furthermore, if the modulator employs binary waveforms, and the detector makes hard decision, then the channel has a discrete-time binary input sequence and a discrete-time binary output sequence.

Note that if the channel noise and other interferences cause statistically independent errors in the transmitted binary sequence with average probability p, the channel is called a BSC. Besides, since each output bit from the channel depends only upon the corresponding input bit, the channel is also memoryless.

2. Discrete memory less channels (DMC):

A channel is the same as above, but with q-ary symbols at the output of the channel encoder, and Q-ary symbols at the output of the detector, where $Q \geq q$. If the channel and the modulator are memory less, then it can be described by a set of qQ conditional probabilities

$$P\ (Y = y_i | X = x_j) \equiv P\ (y_i | x_j),\ i = 0,1,...,Q-1;\ j = 0,1,...,\ q-1$$

Such a channel is called discrete memory channel (DSC).

If the input to a DMC is a sequence of n symbols u_1 , u_2 ,..., u_n selected from the alphabet X and the corresponding output is the sequence v_1 , v_2 ,..., v_n of symbols from the alphabet Y, the joint conditional probability is

$$P(Y_1 = v_1, Y_2 = v_2, ..., Y_n = v_n \,|\, X_1 = u_1, X_2 = u_2, ..., X_n = u_n) = \prod_{k=1}^{n} P(Y_k = v_k \,|\, X_k = u_k)$$

The conditional probabilities $P(y_i \,|\, x_j)$ can be arranged in the matrix form $\mathbf{P} = [p_{ji}]$, P is called

the probability transition matrix for the channel.

3. Discrete-input, continuous-output channels:

Suppose the output of the channel encoder has q-ary symbols as above, but the output of the detector is un quantized ($Q = \infty$) . The conditional probability density functions

$$p\ (y \,|\, X = x_k),\ k = 0,1,\ 2,...,\ q-1$$

AWGN is the most important channel of this type.

$$Y = X + G$$

where $G \square N(0, \sigma^2)$. Accordingly

$$P(y/X = \text{xk}) = \frac{1}{\sqrt{2\pi}} e^{(-x-y)^2/2\sigma^2} \quad \text{k= 0,1,2,q-1}$$

For any given sequence X_i, $i = 1,\ 2,...,\ n$, the corresponding output is Y_i, $i = 1,\ 2,...,\ n$

$$Y_i = X_i + G_i,\ i = 1,\ 2,...,\ n$$

If, further, the channel is memory less, then the joint conditional pdf of the detector‘s output is

$$p(y_1, y_2, ..., y_n \,|\, X_1 = u_1, X_2 = u_2, ..., X_n = u_n) = \prod_{i=1}^{n} p(y_i \,|\, X_i = u_i)$$

4. Waveform channels:

If such a channel has bandwidth W with ideal frequency response $C(f) = 1$, and if the bandwidth-limited input signal to the channel is $x(t)$, and the output signal, $y(t)$ of the channel is corrupted by AWGN, then

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

The channel can be described by a complete set of orthonormal functions:

$$y(t) = \sum_i y_i f_i(t), \quad x(t) = \sum_i x_i f_i(t), \quad n(t) = \sum_i n_i f_i(t)$$

where

$$y_i = \int_0^T y(t) f_i^*(t) dt = \int_0^T [x(t) + n(t)] f_i^*(t) dt = x_i + n_i$$

The functions $\{f_i(t)\}$ form a complete orthonormal set over (0,T)

The statistical description in such a system is

$P(y/X = xk) = \frac{1}{\sqrt{2\pi}} e^{(-x-y)^2/2\sigma^2}$ k= 0,1,2,q-1

Since $\{n_i\}$ are uncorrelated and are Gaussian, therefore, statistically independent. So

$$p(y_1, y_2, \ldots, y_N \mid x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(y_i \mid x_i)$$

Channel Capacity:

Channel model: DMC

Input alphabet: $X = \{x_0, x_1, x_2, \ldots, x_{q-1}\}$

Output alphabet: $Y = \{y_0, y_1, y_2, \ldots, y_{q-1}\}$

Suppose x_j is transmitted, y_i is received, then

The mutual information (MI) provided about the event $\{X = x_j\}$ by the occurrence of the event

$$\{Y = y_j\} \text{ is } \log\left[P(y_i \mid x_j)/P(y_i)\right] \text{ with } P(y_i) = P(Y = y_i) = \sum_{k=0}^{q-1} P(x_k) P(y_i \mid x_k)$$

Hence, the average mutual information (AMI) provided by the output Y about the input X is

$$I(X,Y)=\sum_{j=0}^{q-1}\sum_{i=0}^{Q-1}P(x_j)P(y_i|x_j)\log\left[P(y_i|x_j)/P(y_i)\right]$$

To maximize the AMI, we examine the above equation:

(1). $P(y_i)$ represents the jth output of the detector;
(2). $P(y_i|x_j)$ represents the channel characteristic, on which we cannot do anything;

(3). $P(x_j)$ represents the probabilities of the input symbols, and we may do something or control them. Therefore, the channel capacity is defined by

$$C=\max_{P(x_j)}\sum_{j=0}^{q-1}\sum_{i=0}^{Q-1}P(x_j)P(y_i|x_j)\log\left[P(y_i|x_j)/P(y_i)\right]$$

with two constraints: $P(x_j)\geq 0$; $\sum_{j=0}^{q-1}P(x_j)=1$

- **Unit of C:**
  - bits/channel use when $\log=\log_2$; and
  - nats/input symbol when $\log=\log_e=\ln$
  - If a symbol enters the channel every τ_s seconds (seconds/channel use)
  - Channel capacity: C/τ_s (bits/s or nats/s).

UNIT-1

INFORMATION THEORY

Information theory is a branch of science that deals with the analysis of a communications system. We will study digital communications – using a file (or network protocol) as the channel Claude Shannon Published a landmark paper in 1948 that was the beginning of the branch of information theory. The messages will be a sequence of binary digits Does anyone know the term for a binary digit.

Fig 1.1 Block Diagram of a typical communication system

(Source:https://www.google.com/search?q=Block+Diagram+of+a+typical+communication+system)

One detail that makes communicating difficult is noise noise introduces uncertainty Suppose I wish to transmit one bit of information what are all of the possibilities tx 0, rx 0 - good tx 0, rx 1 - error tx 1, rx 0 - error tx 1, rx 1 - good Two of the cases above have errors – this is where probability fits into the picture In the case of steganography, the noise may be due to attacks on the hiding algorithm. Claude Shannon introduced the idea of self-information.

$$I(Xj) = \log\frac{1}{p(xj)} = \log\frac{1}{p(j)} = -\log pj$$

Suppose we have an event X, where Xi represents a particular outcome of the

Consider flipping a fair coin, there are two equiprobable outcomes: say X0 = heads, P0 = 1/2, X1 = tails, P1 = 1/2 The amount of self-information for any single result is 1 bit. In other words, the number of bits required to communicate the result of the event is 1 bit. When outcomes are equally likely, there is a lot of information in the result. The higher the likelihood of a particular outcome, the less information that outcome conveys However, if the coin is biased such that it lands with heads up 99% of the time, there is not much information conveyed when we flip the coin and it lands on heads. Suppose we have an event X, where Xi represents a particular outcome of the event. Consider flipping a coin, however, let‘s say there are 3 possible outcomes: heads (P = 0.49), tails (P=0.49), lands on its side (P = 0.02) – (likely much higher than in

reality).

Information

There is no some exact definition, however Information carries new specific knowledge, which is definitely new for its recipient; Information is always carried by some specific carrier in different forms (letters, digits, different specific symbols, sequences of digits, letters, and symbols , etc.); Information is meaningful only if the recipient is able to interpret it. According to the Oxford English Dictionary, the earliest historical meaning of the word information in English was the act of informing, or giving form or shape to the mind. The English word was apparently derived by adding the common "noun of action" ending "-action the information materialized is a message.

Information is always about something (size of a parameter, occurrence of an event, etc). Viewed in this manner, information does not have to be accurate; it may be a truth or a lie. Even a disruptive noise used to inhibit the flow of communication and create misunderstanding would in this view be a form of information. However, generally speaking, if the amount of information in the received message increases, the message is more accurate. Information Theory How we can measure the amount of information? How we can ensure the correctness of information? What to do if information gets corrupted by errors? How much memory does it require to store information? Basic answers to these questions that formed a solid background of the modern information theory were given by the great American mathematician, electrical engineer, and computer scientist Claude E. Shannon in his paper ―A Mathematical Theory of Communication‖ published in ―The Bell System Technical Journal in October, 1948.

A noiseless binary channel 0 0 transmits bits without error, What to do if we have a noisy channel and you want to send information across reliably? Information Capacity Theorem (Shannon Limit) The information capacity (or channel capacity) C of a continuous channel with bandwidth BHertz can be perturbed by additive Gaussian white noise of power spectral density N0/2, C=B log2(1+P/N0B) bits/sec provided bandwidth B satisfies where P is the average transmitted power P = Eb Rb (for an ideal system, Rb= C). Eb is the transmitted energy per bit, Rb is transmission rate.

ENTROPY:

Entropy is the average amount of information contained in each message received.

Here, message stands for an event, sample or character drawn from a distribution or data stream. Entropy thus characterizes our uncertainty about our source of information. (Entropy is best understood as a measure of uncertainty rather than certainty as entropy is larger for more random sources.) The source is also characterized by the probability distribution of the samples drawn from it.

Formula for entropy:

Information strictly in terms of the probabilities of events. Therefore, let us suppose that we have a set of probabilities (a probability distribution) P = {p1, p2, . . . , pn}. We define entropy of the distribution P by

$$H(P) = \sum_{i=1}^{n} pi * log(1/pi).$$

Shannon defined the entropy of the a discrete random variable X with possible values {x1, ..., xn} and probability mass function P(X) as: Here E is the expected value operator, and I is the information content of X. I(X) is itself a random variable. One may also define the conditional entropy of two events X and Y taking values xi and yj respectively, as

$$H(X|Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(y_j)}{p(x_i, y_j)}$$

where $p(x_i,y_j)$ is the probability that $X=x_i$ and $Y=y_j$.

Properties:

- If *X* and *Y* are two independent experiments, then knowing the value of *Y* doesn't influence our knowledge of the value of *X* (since the two don't influence each other by independence):

$$H(X|Y) = H(X).$$

- The entropy of two simultaneous events is no more than the sum of the entropies of each individual event, and are equal if the two events are independent. More specifically, if*X* and *Y* are two random variables on the same probability space, and *(X,Y)* denotes their Cartesian product, then

$$H[(X,Y)] \leq H(X) + H(Y).$$

SHANNON–HARTLEY THEOREM

In information theory, the Shannon–Hartley theorem tells the maximum rate at which information can be transmitted over a communications channel of a specified bandwidth in the presence of noise. It is an application of the noisy channel coding theorem to the archetypal case of a continuous-time analog communications channel subject to Gaussian noise. The theorem establishes Shannon's channel capacity for such a communication link, a bound on the maximum amount of error-free digital data (that is, information) that can be transmitted with a specified bandwidth in the presence of the noise interference, assuming that the signal power is bounded, and that the Gaussian noise process is characterized by a known power or power spectral density. The law is named after Claude Shannon and Ralph Hartley.

$$C = B\log_2\left(1+\frac{S}{N}\right) = \frac{B\log_{10}\left(1+S/N\right)}{\log_{10}(2)}$$

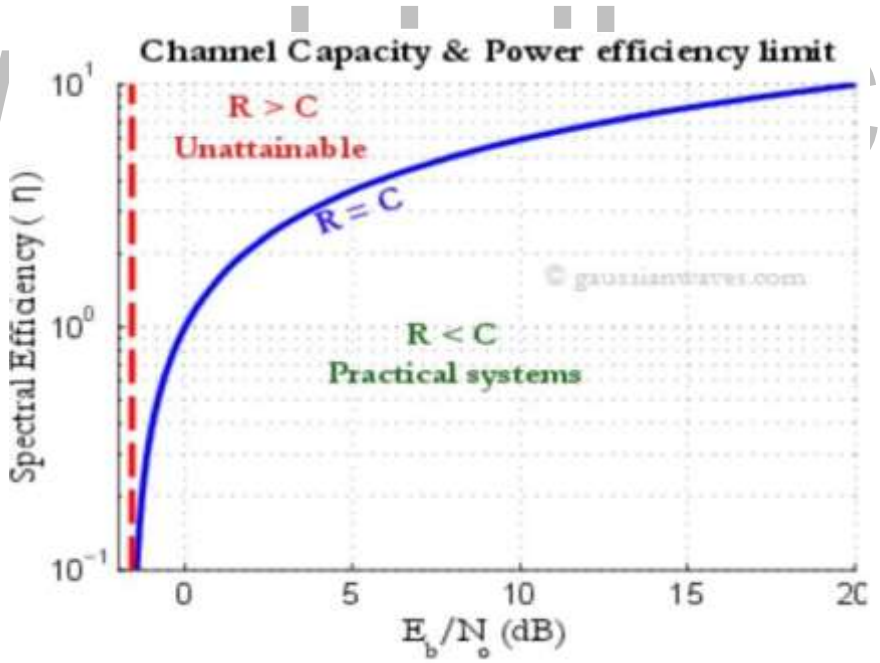

Fig 1.5 Shannon–Hartley Theorem

(Source:https://www.google.com/search?q=Shannon%E2%80%93Hartley+Theorem&tbm)

Considering all possible multi-level and multi-phase encoding techniques, the Shannon–Hartley theorem states the channel capacity C, meaning the theoretical tightest upper bound on the information rate (excluding error correcting codes)

of clean (or arbitrarily low bit error rate) data that can be sent with a given average signal power S through an analog communication channel subject to

additive white Gaussian noise of power N, is:

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

Where C is the channel capacity in bits per second;

B is the bandwidth of the channel in hertz (pass band bandwidth in case of a modulated signal);

S is the average received signal power over the bandwidth (in case of a modulated signal, often denoted C, i.e. modulated carrier), measured in watts (or volts squared);

N is the average noise or interference power over the bandwidth, measured in watts (or volts squared); and

S/N is the signal-to-noise ratio (SNR) or the carrier-to-noise ratio (CNR) of the communication signal to the Gaussian noise interference expressed as a linear power ratio (not as logarithmic decibels).

APPLICATION & ITS USES:

1. Huffman coding is not always optimal among all compression methods.

2. Discrete memory less channels.

3. To find 100% of efficiency using these codings.

Shannon-Fano Algorithm:

- Arrange the character set in order of decreasing probability
- While a probability class contains more than one symbol:

✓ Divide the probability class in two so that the probabilities in the two halves are as nearly as possible equal.

✓ Assign a '1' to the first probability class, and a '0' to the second

Character	Probability			code
X6	0.25		1/0	11
X3	0.2	1		10
X4	0.15	1	1/0	011
X5	0.15			010
X1	0.1	0	1/0	001
X7	0.1			0001
X2	0.05	0	1/0	0000

Huffman Encoding:

Statistical encoding To determine Huffman code, it is useful to construct a binary tree Leaves are characters to be encoded Nodes carry occurrence probabilities of the characters belonging to the subtree Example: How does a Huffman code look like for symbols with statistical symbol occurrence probabilities: P(A) = 8/20, P(B) = 3/20, P(C) = 7/20, P(D) = 2/20? Step 1 : Sort all Symbols according to their probabilities (left to right) from Smallest to largest these are the leaves of the Huffman tree Step 2: Build a binary tree from left to Right Policy: always connect two smaller nodes together (e.g., P(CE) and P(DA) had both Probabilities that were smaller than P(B), Hence those two did connect first Step 3: label left branches of the tree With 0 and right branches of the tree With 1 Step 4: Create Huffman Code Symbol A = 011 Symbol B = 1 Symbol C = 000 Symbol D = 010 Symbol E = 001

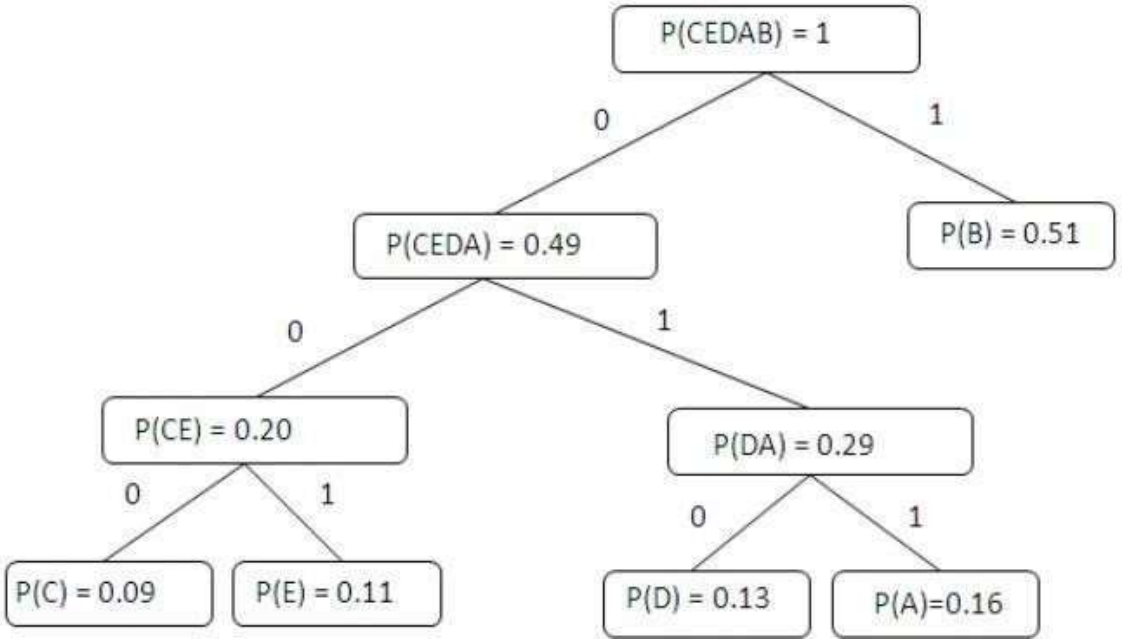

Fig1.4 Huffman Coding

(Source:https://www.google.com/search?q=huffman+coding&safe)

SHANNON-FANO CODING

This is a basic information theoretic algorithm. A simple example will be used to illustrate the algorithm:

Symbol	A	B	C	D	E
Count	15	7	6	6	5

Encoding for the Shannon-Fano Algorithm:

A top-down approach

1. Sort symbols according to their frequencies/probabilities, e.g., ABCDE.
2. Recursively divide into two parts, each with approx. same number of counts.

Procedure for shannon fano algorithm:

A Shannon–Fano tree is built according to a specification designed to define an effective code table. The actual algorithm is simple:

1. For a given list of symbols, develop a corresponding list of probabilities or frequency counts so that each symbol‘s relative frequency of occurrence is known.

2. Sort the lists of symbols according to frequency, with the most frequently occurring symbols at the left and the least common at the right.

3. Divide the list into two parts, with the total frequency counts of the left part being as close to the total of the right as possible.

4. The left part of the list is assigned the binary digit 0, and the right part is assigned the digit 1. This means that the codes for the symbols in the first part will all start with 0, and the codes in the second part will all start with 1.

5. Recursively apply the steps 3 and 4 to each of the two halves, subdividing groups and adding bits to the codes until each symbol has become a corresponding code leaf on the tree.

Table For Shannon Fano Coding:

✓ **Table For Shannon Fano Coding:**

Symbol	Count	log(1/p)	Code	Subtotal (# of bits)
A	15	1.38	00	30
B	7	2.48	01	14
C	6	2.70	.10	12
D	6	2.70	110	18
E	5	2.96	111	15

TOTAL NO OF BITS: 89

HUFFMAN CODING

The Shannon–Fano algorithm doesn't always generate an optimal code. In 1952, David A. Huffman gave a different algorithm that always

produces an optimal tree for any given probabilities. While the Shannon–Fano tree is created from the root to the leaves, the Huffman

Procedure for Huffman Algorithm:

1. Create a leaf node for each symbol algorithm works from leaves to the root in the opposite direction and add it to frequency of occurrence.

2. While there is more than one node in the queue:

- Remove the two nodes of lowest probability or frequency from the queue
- Prepend 0 and 1 respectively to any code already assigned to these nodes

- Create a new internal node with these two nodes as children and with probability equal to the sum of the two nodes' probabilities.
- Add the new node to the queue.

3. The remaining node is the root node and the tree is complete.

CHANNEL CODING THEOREM

The noisy-channel coding theorem (sometimes Shannon's theorem), establishes that for any given degree of noise contamination of a communication channel, it is possible to communicate discrete data (digital information) nearly error-free up to a computable maximum rate through the channel. This result was presented by Claude Shannon in 1948 and was based in part on earlier work and ideas of Harry Nyquist and Hartley. The Shannon limit or Shannon capacity of a communications channel is the theoretical maximum information transfer rate of the channel, for a particular noise level.

The theorem describes the maximum possible efficiency of error-correcting methods versus levels of noise interference and data corruption. Shannon's theorem has wide-ranging applications in both communications and data storage. This theorem is of foundational importance to the modern field of information theory. Shannon only gave an outline of the proof. The first rigorous proof for the discrete case is due to Amiel Feinstein in 1954.

The Shannon theorem states that given a noisy channel with channel capacity C and information transmitted at a rate R, then if R<C there exist codes that allow the probability of error at the receiver to be made arbitrarily small. This means that, theoretically, it is possible to transmit information nearly without error at any rate below a limiting rate, C.

The converse is also important. If R>C , an arbitrarily small probability of error is not achievable. All codes will have a probability of error greater than a certain positive minimal level, and this level increases as the rate increases. So, information cannot be guaranteed to be transmitted reliably across a channel at rates beyond the channel capacity. The theorem does not address the rare situation in which rate and capacity are equal.

The channel capacity C can be calculated from the physical properties of a channel; for a band-limited channel with Gaussian noise, using the Shannon–Hartley theorem.

For every discrete memory less channel, the channel capacity has the following property. For any ε > 0 and R < C, for large enough N, there exists a code of length N and rate ≥ R and a decoding algorithm, such that the maximal probability of block error is ≤ ε.

2. I ıievable,

whe $R(p_b) = \dfrac{C}{1 - H_2(p_b)}$. and $H_2(p_b)$ is the binary entropy function

$$H_2(p_b) = -[p_b \log_2 p_b + (1 - p_b) \log_2 (1 - p_b)]$$

SOURCE CODING

A code is defined as an n-tuple of q elements. Where q is any alphabet. Ex. 1001 n=4, q={1,0} Ex.

2389047298738904 n=16, q={0,1,2,3,4,5,6,7,8,9} Ex. (a,b,c,d,e) n=5, q={a,b,c,d,e,…,y,z} The most common code is when q={1,0}. This is known as a binary code. The purpose A message can become distorted through a wide range of unpredictable errors.

- Humans
- Equipment failure
- Lighting interference
- Scratches in a magnetic tape

Error-correcting code:

To add redundancy to a message so the original message can be recovered if it has be garbled. e.g. message = 10 code = 1010101010

Send a message:

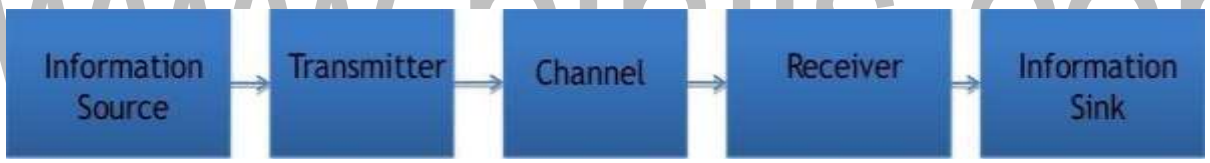

Fig 1.3 Block Diagram of Transmission

(Source:https://www.google.com/search?q=Block+Diagram+of+a+typical+communication+system)

Source Coding loss:

It may consider semantics of the data depends on characteristics of the data e.g. DCT, DPCM, ADPCM, color model transform A code is distinct if each code word can be distinguished from every other (mapping is one-to-one) uniquely decodable if every code word is identifiable when immersed in a sequence of code words e.g., with previous table, message 11 could be defined as either ddddd or bbbbbb Measure of Information Consider symbols si and the probability of occurrence of each symbol p(si)

Example Alphabet = {A, B} p(A) = 0.4; p(B) = 0.6 Compute Entropy (H) - 0.4*log2 0.4 + - 0.6*log2 0.6 = .97 bits Maximum uncertainty (gives largest H) occurs when all probabilities are equal Redundancy Difference between avg. code word length (L) and avg. information content (H) If H is constant, then can just use L Relative to the optimal value