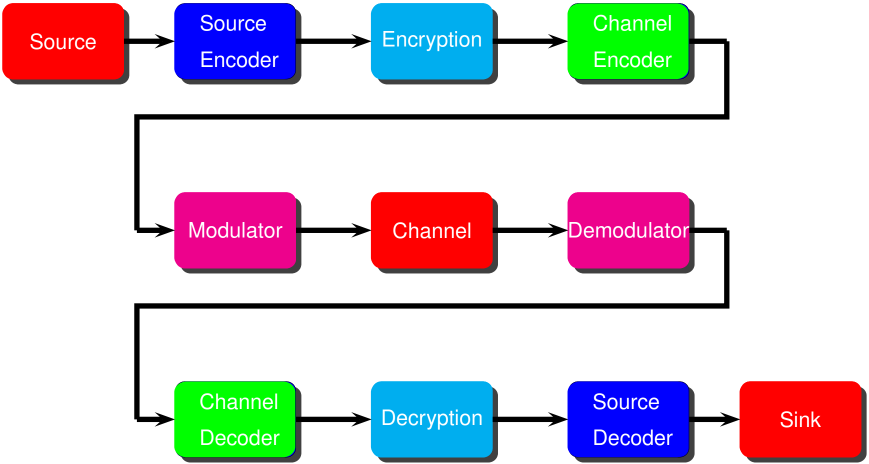
Lecture 2: Goals

- Understand the basic characteristics of signals as vectors: rate, minimum square Euclidean distance
- Know the fundamental tradeoff between data rate, bandwidth, signal power and noise power in communicating binary information (bits) from a source to a destination
- Know the fundamental tradeoff between data rate, and distortion in representing a source signal in binary form (bits)
- Know the fundamental tradeoff between source sample rate, distortion, energy, bandwidth in a communication system.

Communication System Block Diagram



Signals as Vectors

- In digital communication we can communicate 1 bit of information by transmitting one of 2 signals.
- We can communicate $\log_2(M)$ bits of information by transmitting one of M signals.
- A collection of M signals is called a signal set.
- Signals can be represented by a set of N orthonormal waveforms and M vectors of length N (more on this later).

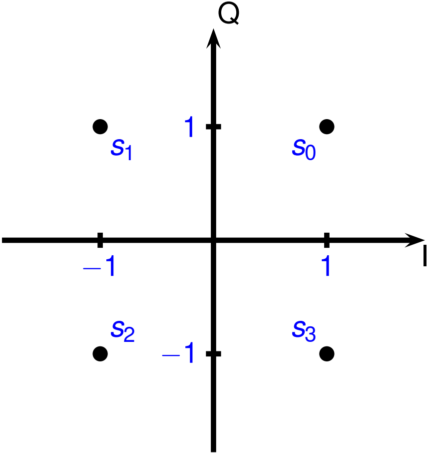
Signal Set Parameters

- Rate of communications is the number of bits we can communicate per second (bps).
- Rate of communication per unit bandwidth is sometimes called the bandwidth efficiency. The units are bps/Hz.
- Energy used per signal, E, average energy used, $\bar{E}$, and average energy used per bit of information ($\bar{E}_b$).
- Euclidean distance between signals, $d_E^2(s_i, s_j)$.

Signals as Vectors

- A signal set with M signals is often represented with M vectors of length N (dimension N).
- Often a vector of length 2 has an I component and a Q component (for in-phase and quadrature phase).
- The energy of a signal is the squared length of the vector.
- The minimum squared Euclidean distance $d^2_{E,min}$ of a signal set is the minimum square distance between any two distinct vectors.
- The rate in units of bits/dimension is $r = \log_2(M)/N$.
- The normalized rate in bits/second/Hz is $R/W = 2\log_2(M)/N$ since there are $2W$ possible dimensions per second in a bandwidth of W Hz (more on this later).

Example 1: Quadriphase Shift Keying (QPSK)

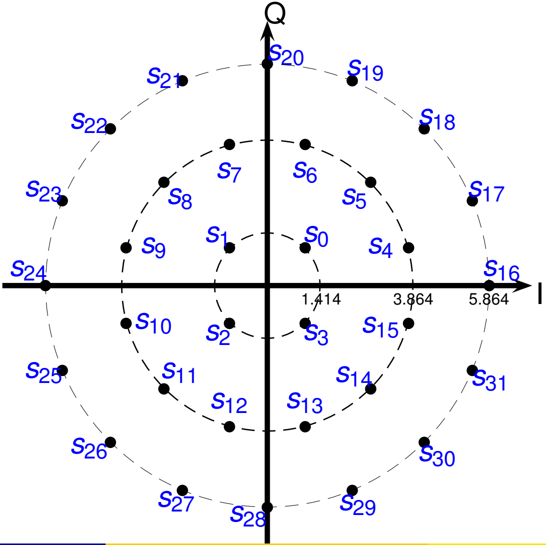


- Energy per signal is 2 for each of the four signals.
- Energy per bit is $E_b = 1$.
- Minimum squared Euclidean distance between signals is $d^2_{E,min} = 4$.
- Rate is $r = 2$ bits/2 dimensions or $R = 2$ bps/Hz.

Bits/dimension vs. bits/second/Hz

- Suppose we have a signal with bandwidth (highest frequency) W.
- We know we can represent such a signal with samples at rate (very slightly greater than) $2W$ samples/second.
- We can perfectly reconstruct the signal from the samples.
- Any set of samples $2W$ samples per second can be turned into a waveform with bandwidth W.
- So a vector of samples of length $N = 2WT$ can be turned into a waveform with bandwidth W and duration T.
- The conclusion is that there are $2WT$ possible dimensions in a bandwidth W and duration T or 2 dimensions/sec/Hz
- So r bits/dimension corresponds to $R/W = 2r$ bits/second/Hz.

Example 2: 32 Amplitude and Phase Shift Keying (32APSK)



Example 2: 32 APSK

- Energy per signal is 2 for four of the signals.
- Energy per signal is 14.928 for 12 of the signals.
- Energy per signal is 34.383 for 16 of the signals.
- Average energy per bit is $\bar{E}_b = 4.608$.
- Minimum squared Euclidean distance between signals is $d^2_{E,min} = 4$.
- Rate is $r = 5$ bits $/2$ dimensions, $R = 5$ bps/Hz.

Example 3

$N = 8, M = 8$.

$$\begin{aligned}
000 \to s_0 &= (+1,+1,+1,+1,+1,+1,+1,+1)\\
001 \to s_1 &= (+1,-1,+1,-1,+1,-1,+1,-1)\\
010 \to s_2 &= (+1,+1,-1,-1,+1,+1,-1,-1)\\
011 \to s_3 &= (+1,-1,-1,+1,+1,-1,-1,+1)\\
100 \to s_4 &= (+1,+1,+1,+1,-1,-1,-1,-1)\\
100 \to s_5 &= (+1,-1,+1,-1,-1,+1,-1,+1)\\
110 \to s_6 &= (+1,+1,-1,-1,-1,-1,+1,+1)\\
111 \to s_7 &= (+1,-1,-1,+1,-1,+1,+1,-1)
\end{aligned}$$

Example 3

- $M = 8$ signals in $N = 8$ dimensions
- Energy per signal is $E = 8$.
- Average energy per bit is $\bar{E}_b = 8/3 = 2.66$.
- Minimum squared Euclidean distance between signals is $d^2_{E,min} = 16$.
- Rate is $r = 3$ bits/8 dimensions $= 0.375$ bits/dimension, $R = 0.75$ bps/Hz.

Comparison

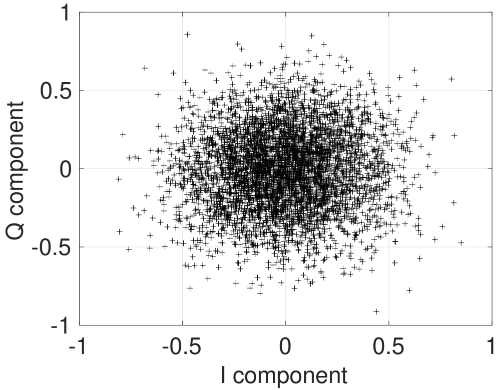
Example	Rate (bits/dimension)	$d^2_{E,min}/\bar{E}_b$
1	1	4.00
2	2.5	0.868
3	0.375	6.00

- Better distinguishability per unit energy (larger $d^2_{E,min}/E_b$) can be achieved with smaller rate.
- Higher rate means signals are less distinguishable per unit energy.

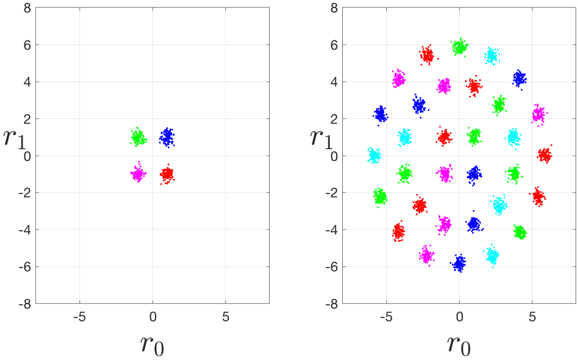
Noise

- Noise is present in virtually all communication system
- Noise is due, in part, to thermal motion of electrons in the receiver electronics
- The noise power is proportional to the temperature of the receiver.
- Thermal noise is white: same power at all frequencies.
- Thermal noise is measured by the power spectral density denoted by $N_0/2$ where $N_0 = kT_0$ and k is Boltzmann's constant and T_0 is the temperature in Kelvin.
- The noise power in a bandwidth $-W$ to W is $N_0/2(2W) = N_0W$.
- In addition to noise, the propagation characteristics can cause the received signal to differ significantly from the transmitted signal in power but can also cause distortion.

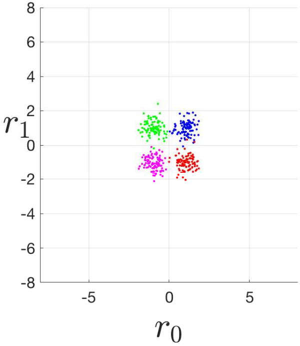
Noise

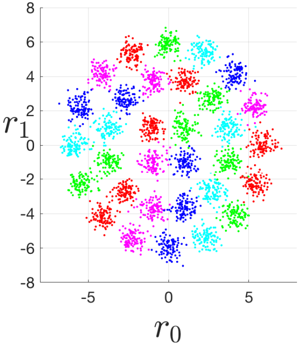


Signal and Noise: High SNR



Signal and Noise: Low SNR

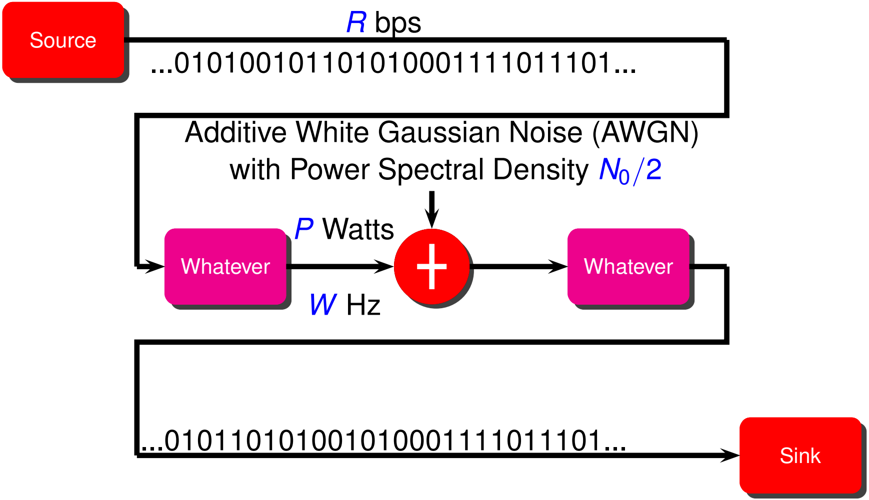




Two Fundamental Tradeoffs

1. At what rate R in bits/second can we reliably (very very small error probability) communication data when the received power is P, the available bandwidth is W and there is a certain amount of noise N_0 (the noise power spectral density)?
2. At what rate R in bits/second can a source of information (e.g. an analog source) be compressed into bits and then the original source be reconstructed with distortion D from the original?

Fundamental Tradeoff One



Assumptions

- The source produces equally likely data bits (0s and 1s) at rate R bits/second.
- We transmit a signal (waveform) such that the *received* power is P.
- The transmitted signal has bandwidth W (Hz).
- Noise is added to the transmitted signal. The noise is white (power at all frequencies of interest), Gaussian and has power spectral density $N_0/2$ Watts/Hz. This is called an additive white Gaussian noise channel.
- We can allow any delay or complexity.

Fundamental Tradeoff

- In 1948 Claude Shannon (U of M EE/Math graduate) published a paper in which he determined the tradeoff between data rate, bandwidth, signal power and noise power for reliable communications for an additive white Gaussian noise channel.
- W be the bandwidth (in Hz),
- R be the data rate (in bits per second),
- P be the *received* signal power (in Watts),
- $N_0/2$ the noise power spectral density (in Watts/Hz).
- Then reliable communication is possible provided

$$\boxed{R < W \log_2(1 + \frac{P}{N_0 W}).}$$

Formula Search

Find this equation on bust of Shannon outside the EECS building.

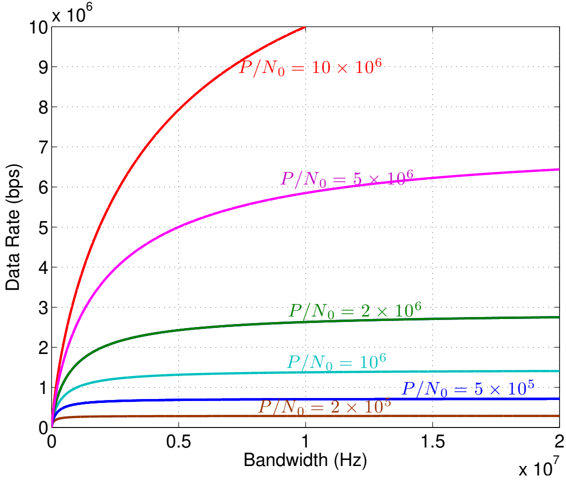
$$C = W \log_2\left(\frac{P+N}{N}\right)$$

Here $N = N_0 W$ is the noise power.

Soni and Goodman: A mind at play: How Claude Shannon invented the information age

Before Shannon's 1948 "A Mathematical Theory of Communication," a century of common sense and engineering trial and error said that noise–the physical world's tax on our messages–had to be lived with. Shannon proved that noise could be defeated, that information sent from Point A cound be received with perfection at Point B, not just often, but essentially always. He gave engineers the conceptual tools to digitize information and send it flawlessly, a result considered hopelessly utopian up until the moment Shannon proved it was not.

Fundamental Tradeoff



Capacity

For large values of W the maximum rate (capacity) approaches

$$\lim_{W\to\infty} W\log_2(1+\frac{P}{N_0W}) = \frac{P}{N_0\ln(2)} = 1.4426(P/N_0).$$

Let E_b be the energy transmitted per bit of information. Then

$$E_b = P/R \quad \text{or} \quad P = E_bR.$$

Using this relation we can express the capacity formula as

$$R/W < \log_2(1+\frac{E_b}{N_0}\frac{R}{W}).$$

Inverting this we obtain

$$\boxed{E_b/N_0 > \frac{2^{R/W}-1}{R/W}.}$$

Capacity

$$\boxed{E_b/N_0 > \frac{2^{R/W}-1}{R/W}.}$$

- The interpretation is that reliable communication is possible with *bandwidth efficiency* R/W provided that the *signal-to-noise ratio*, E_b/N_0, is larger than the right hand side of the above equation.
- For small values of R/W the smallest value of E_b/N_0 where reliable communication is possible is $\ln(2) = 0.693$. That is,

$$\lim_{R/W \to 0} \frac{2^{R/W} - 1}{R/W} = \ln(2).$$

dB or not dB?

When the range of values for energy or power are vast we usually employ a dB scale. The conversion is

$$E_b/N_0(dB) = 10\log_{10}(E_b/N_0).$$

The smallest signal-to-noise ratio for reliable communication (at low rates) is

$$E_b/N_0 > \log(2) = 0.693$$
$$E_b/N_0(dB) > 10\log_{10}(0.693) = -1.59dB.$$

dBW, dBm

Sometimes **absolute** power levels are also expressed in dB's by referencing them to either 1W or 1mW. When referencing to 1W the dB units are written as dBW. When referencing to 1mW the dB units are written as dBm. So, for example

$$\begin{aligned} \text{100Watts} &= 10\log_{10}(\text{100Watts/1Watt}) = \text{20dBW} \\ &= 10\log_{10}(\text{100Watts/1mWatt}) = \text{50dBm} \end{aligned}$$

$$\begin{aligned} \text{10 Watts} &= 10\log_{10}(\text{10Watts/1Watt}) = \text{10dBW} \\ &= 10\log_{10}(\text{10Watts/1mWatt}) = \text{40dBm} \end{aligned}$$

dBW, dBm

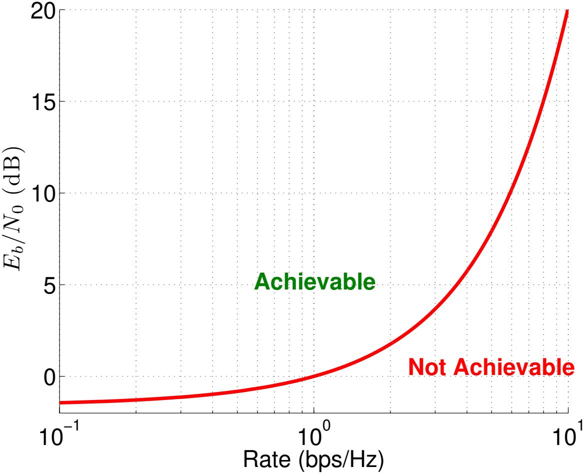
$$\begin{aligned} 1\text{Watt} &= 10\log_{10}(1\text{Watt}/1\text{Watt}) = 0\text{dBW} \\ &= 10\log_{10}(1\text{Watt}/1\text{mWatt}) = 30\text{dBm} \end{aligned}$$

$$\begin{aligned} .1\text{Watts} &= 10\log_{10}(.1\text{Watt}/1\text{Watt}) = -10\text{dBW} \\ &= 10\log_{10}(.1\text{Watt}/1\text{mWatt}) = 20\text{dBm} \end{aligned}$$

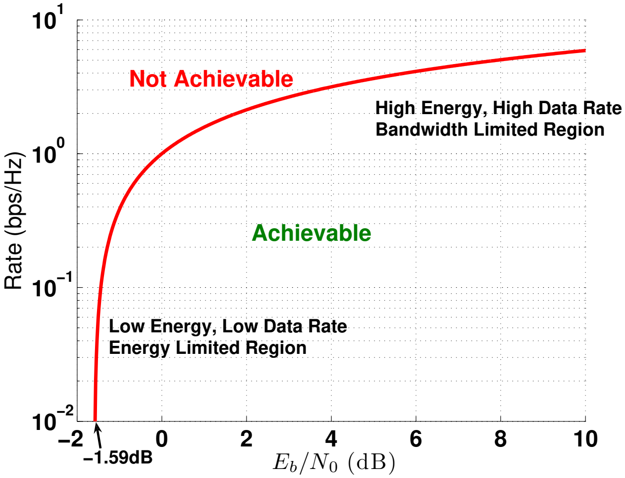
$$\begin{aligned} .01\text{Watts} &= 10\log_{10}(.01\text{Watt}/1\text{Watt}) = -20\text{dBW} \\ &= 10\log_{10}(.01\text{Watt}/1\text{mWatt}) = 10\text{dBm} \end{aligned}$$

$$\begin{aligned} .001\text{Watts} &= 10\log_{10}(.001\text{Watt}/1\text{Watt}) = -30\text{dBW} \\ &= 10\log_{10}(.001\text{Watt}/1\text{mWatt}) = 0\text{dBm} \end{aligned}$$

Capacity

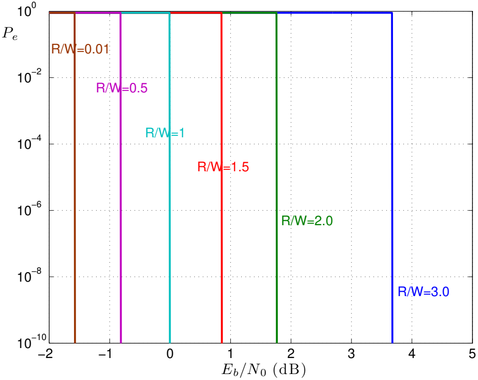

20
15
10
5
0
E_b/N_0 (dB)
Achievable
Not Achievable
10^{-1}
10^{0}
10^{1}
Rate (bps/Hz)

Capacity



Rate (bps/Hz)
10^1
10^0
10^{-1}
10^{-2}
−2
0
2
4
6
8
10
E_b/N_0 (dB)
−1.59dB
Not Achievable
Achievable
High Energy, High Data Rate
Bandwidth Limited Region
Low Energy, Low Data Rate
Energy Limited Region

Capacity



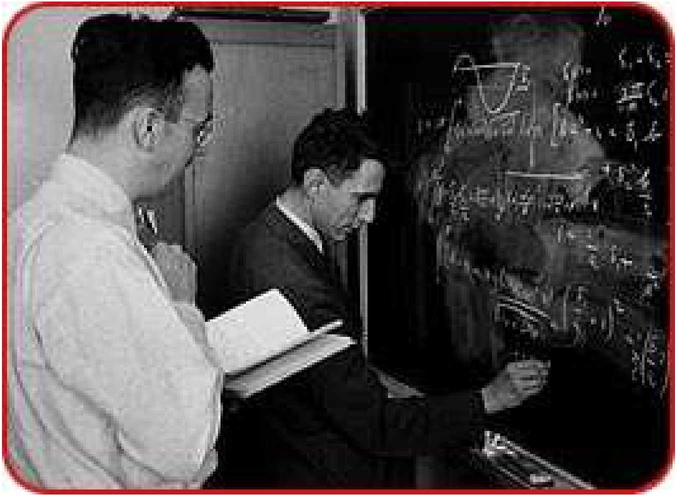
- Each curve represents the achievable error probability for a given rate.
- The error probability achievable becomes zero when the signal-to-noise ratio (E_b/N_0) is above some threshold that depends on R/W.

Notes

- The capacity formula only provides a tradeoff between energy efficiency and bandwidth efficiency. Complexity is essentially infinite, as is delay. The model of the channel is rather benign in that no signal fading is assumed to occur.
- The capacity theorem says that we can communicate with error probability near zero at rates below the capacity or equivalently at values of E_b/N_0 above a threshold.

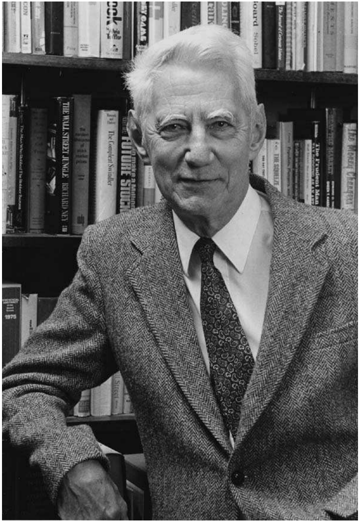
Claude Elwood Shannon (1916-2001)

Claude Elwood Shannon



Claude Elwood Shannon

Claude Elwood Shannon



Claude Elwood Shannon

Example

ZigBee radios (IEEE 802.15.4) use about 2MHz of bandwidth and have a data rate of 250kbps. The transmitted power is 1mW. The relation between transmitted power and received power as a function of distance in certain environments is

$$P_{r|dBW} = P_{t|dBW} - 50.3 - 40\log_{10}(d)$$

Determine the maximum possible distance for reliable communications.

Example

Solution: Known quantities:

$$\begin{aligned} R &= 250\text{kbps} \\ W &= 2\text{MHz} \\ N_0 &= 4 \times 10^{-21}\text{Watts/Hz} \end{aligned}$$

The capacity formula tells us that

$$R < W \log_2(1 + \frac{P_r}{N_0 W})$$
$$(2^{R/W} - 1) N_0 W < P_r$$

So from the smallest possible received power we can find the largest possible distance using

$$P_{r|dBW} = P_{t|dBW} - 50.3 - 40 \log_{10}(d)$$

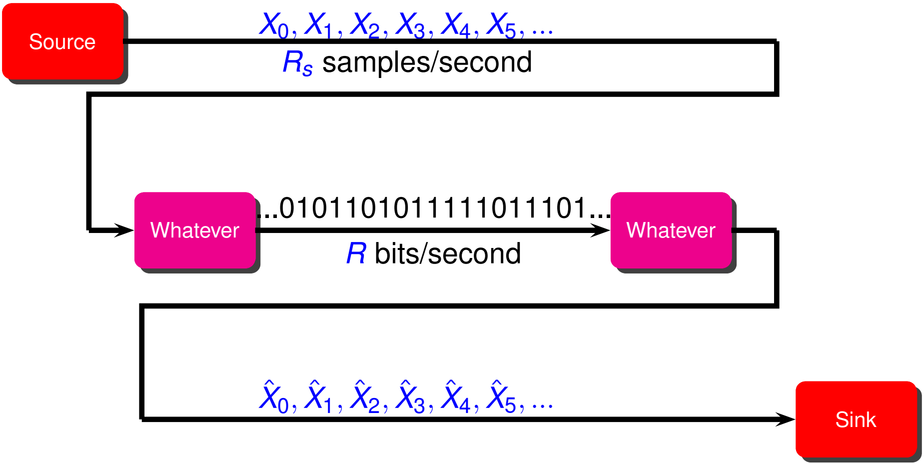
$$
\begin{aligned}
(2^{(250\times10^3)/(2\times10^6)} - 1)(4\times10^{-21})\times(2\times10^6) &< P_r \\
7.24\times10^{-16} &< P_r \\
-151\text{dBW} &< P_{r|dBW} \\
40\log_{10}(d) &< P_{t|dB} - P_{r|dB} - 50.3 \\
40\log_{10}(d) &< -30 + 151 - 50.3 \\
40\log_{10}(d) &< 71.18 \\
\log_{10}(d) &< 1.78 \\
d &< 59.9m
\end{aligned}
$$

Example

- In practice ZigBee radios are far from optimal. They are designed for low power (low complexity).
- A typical ZigBee radio might have a receiver sensitivity of between -90dBm and -100 dBm (-120dBW and -130dBW) [1].
- Part of this difference is due to the noise factor in the receiver which can be from 2-10 dB.
- Part of this difference is due to the nonideal channel characteristics (multipath fading).
- Much of this is due to the system having low complexity.

[1] https://www.eetimes.com/a-zigbee-radio-tutorial-for-the-non-rf

Second Fundamental Tradeoff



Second Fundamental Tradeoff

- Suppose a source produces samples at a rate of R_s samples/second.
- Each sample is a Gaussian distributed random variable (mean 0, variance σ^2).
- The samples are independent. For example

$$X_0 = -0.4326,\ X_1 = -1.6656,\ X_2 = +0.1253,$$
$$X_3 = +0.2877,\ X_4 = -1.1465,\ X_5 = +1.1909, ...$$

- We want to represent these samples with a sequence of bits at rate R bps.
- The goal is to as accurately as possible reproduce the source samples with the fewest possible bits or the lowest bit rate R.
- The distortion is the mean-squared-error; that is $D = E[(X_i - \hat{X}_i)^2]$.

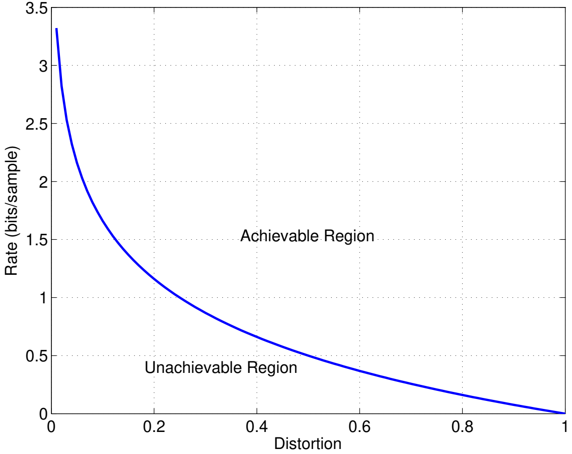
Fundamental Tradeoff Two

- Reproduction of the source with distortion D is possible provided

$$R/R_s > \begin{cases} \frac{1}{2}\log_2(\frac{\sigma^2}{D}), & D < \sigma^2 \\ 0, & D > \sigma^2. \end{cases}$$

- One extra bit per sample improves the distortion by a factor of 4 (6dB).

Rate-Distortion Function

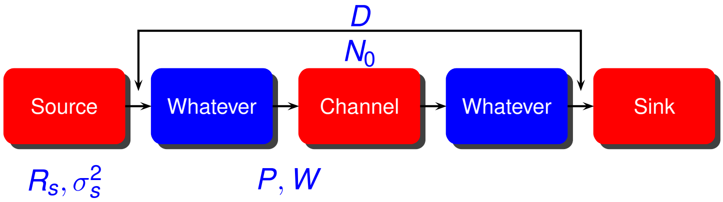


Joint Source-Channel Coding Problem: Two examples

- Gaussian source, additive Gaussian noise channel
- Binary source, additive Gaussian noise channel

Joint Source-Channel Coding Problem

- Suppose we have a Gaussian source with (independent) samples at rate R_s, with variance σ_s^2.
- Suppose we have an additive white Gaussian noise channel with power spectral density $N_0/2$.
- Suppose we have signals with power P and bandwidth W.
- What distortion D is possible between the source and the sink?

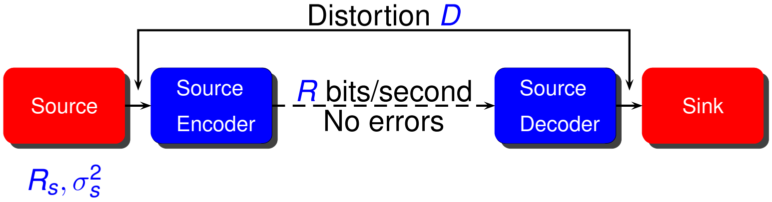


Joint Source-Channel Coding Problem

The source can be represented (compressed) by a sequence of bits with rate (bits/second)

$$R = \frac{R_s}{2} \log_2\left(\frac{\sigma_s^2}{D}\right)$$

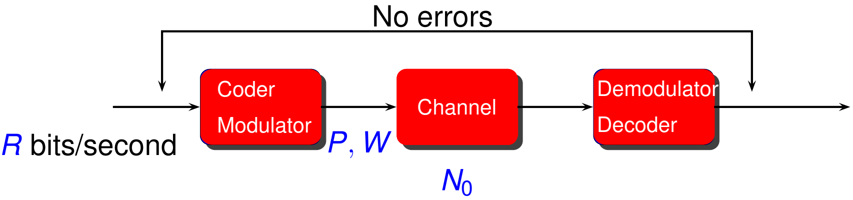
and reproduced from those bits with distortion (mean squared error) of D.



Joint Source-Channel Coding Problem

Those bits, produced by the source encoder, can be reliably (essentially no errors) communicated from the transmitter to the receiver with power P and bandwidth W over an additive white Gaussian noise channel with noise power spectral density $N_0/2$ provided the rate (bits/second) is less than the capacity

$$R < W \log_2(1 + \frac{P}{N_0 W}).$$



Joint Source-Channel Coding Problem

So we can communicate the source to the destination with distortion D provided

$$
\begin{aligned}
\frac{R_s}{2}\log_2\left(\frac{\sigma_s^2}{D}\right) &< W\log_2\left(1+\frac{P}{N_0W}\right) \\
D &> \frac{\sigma_s^2}{\left(1+\frac{P}{N_0W}\right)^{2W/R_s}}
\end{aligned}
$$

Joint Source-Channel Coding Problem

- This is achieved by first compressing the source to a rate of

$$R_s/2\log_2(\frac{\sigma^2}{D})$$

  and then transmitting these over a channel with capacity (maximum achievable rate) of

$$W\log_2(1+\frac{P}{N_0 W}).$$

- The transmission of the bits generated by the source encoder is reliable. So the only distortion comes from the source encoder.
- The channel limits the rate at which reliable communication is possible.
- This puts a constraint on the source encoder in terms of the maximum rate that the source can compress the information into.

Binary Source-Gaussian Channel

- If we start with a sequence of bits at rate R_s bits/second, equally likely, then we can compress these bits to a rate of R_1 bits if we allow errors with probability $P_{e,b}$ where

$$
\begin{aligned}
R &= R_s(1 - H_2(P_{e,b})) \\
&= R_s(1 + P_{e,b}\log_2(P_{e,b}) + (1 - P_{e,b})\log_2(1 - P_{e,b})) \\
&= R_s[1 - H_2(P_{e,b})]
\end{aligned}
$$

  where

$$H_2(x) = -x\log_2(x) - (1 - x)\log_2(1 - x)).$$

- These compressed bits can be reliably communicated over a channel with bandwidth W Hz using power P watts provided

$$R < W\log_2(1 + \frac{P}{N_0 W})$$

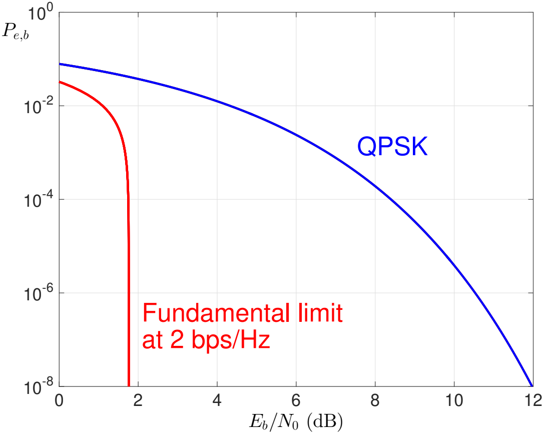
- Thus error probability $P_{e,b}$ is possible provided

$$R_s[1 - H_2(P_{e,b})] < W \log_2(1 + \frac{P}{N_0 W})$$

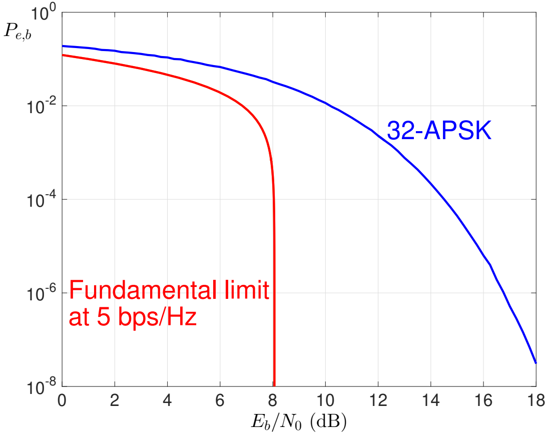
- We can express this in terms of energy per information bit by writing $E_b = P/R_s$. Using this relation the condition for achieving error probability $P_{e,b}$ is

$$\begin{aligned} 1 - H_2(P_{e,b}) &< \frac{W}{R_s} \log_2(1 + \frac{E_b}{N_0} \frac{R_s}{W}) \\ E_b/N_0 &> \frac{2^{R_s/W(1-H_2(P_{e,b}))} - 1}{R_s/W}. \end{aligned}$$

Fundamental limit on error probability and signal-to-noise ratio for bandwidth efficiency 2 bps/Hz compared to QPSK



Fundamental limit on error probability and signal-to-noise ratio for bandwidth efficiency 5 bps/Hz compared to 32-APSK



Key Point

Channel Coding Theorem (Shannon)

There exists communication systems that have arbitrarily small error probability if and only if the rate of communication is less than the channel capacity.

$$R < W \log_2(1 + \frac{P}{N_0 W})$$

$$E_b/N_0 > \frac{2^{R/W} - 1}{R/W}$$

This is a fundamental tradeoff between bandwidth efficiency, R/W (bits/second/Hz) and energy efficiency, E_b/N_0.