Chapter 15: Quantum Information Theory

In this lecture you will learn:

- **Classical Information Theory and Entropy**
- **Classical Information Compression and Information Communication**
- **Von Neumann Entropy and Quantum Information Theory**
- **Holevo's Theorem and Accessible Information**
- **HSW Theorem and Quantum Communication**
- **Classical Communication with Quantum States of Light**
- **Entanglement and Entropy**

Classical Coding Theory

Consider the following possible values of a random variable *X* that is to be measured and the corresponding a-priori probabilities:

Value of *X*	Probability	Coding #1
a	1/32	000
b	132	001
c	1/8	010
d	1/4	011
e	1/8	100
f	1/8	101
g	1/16	110
h	1/4	111

Suppose you make the measurement *N* times

After you are done, you wish to tell your friend about ALL the measurement results

How many bits do you need to do this?

Coding Scheme #1:
There are 8 possible outcomes of every measurement, so we need 3 bits to encode all the outcomes of a single measurement, and all *N* outcomes can be encoded using 3*N* bits

Classical Coding Theory and Information Compression

The coding scheme #1 does not take into account that some measurement outcomes are very unlikely and some are much more likely

Value of *X*	Probability	Coding #1	Coding #2
a	1/32	000	00000
b	1/32	001	00001
c	1/8	010	011
d	1/4	011	10
e	1/8	100	001
f	1/8	101	010
g	1/16	110	0001
h	1/4	111	11

Assign shorter codes to more likely outcomes and longer codes to less likely outcomes

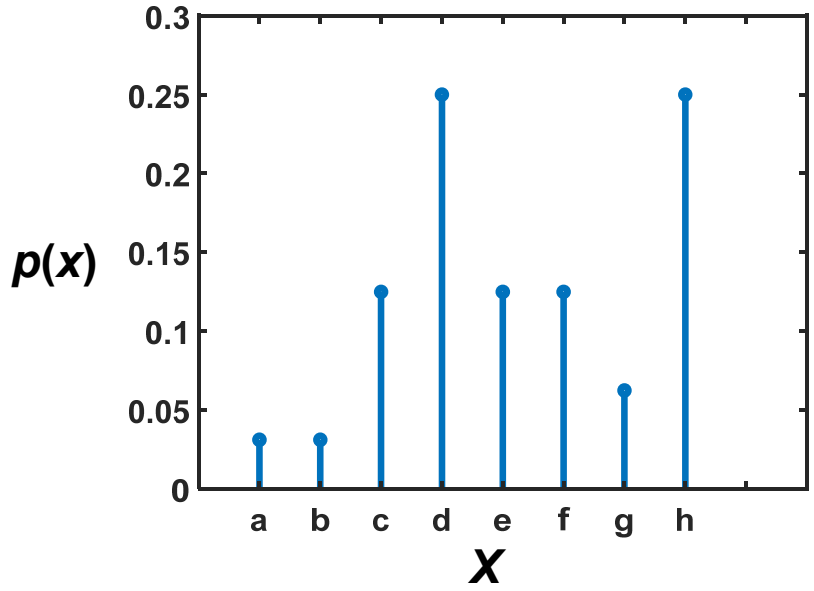
…..but such that any chain of bits representing the *N* outcomes is uniquely decodable!

Average number of bits required per outcome
= 2x(1/32)x5 + 1x(1/16)x4 + 3x(1/8)x3 + 2x(1/4)x2 = 2.69 bits !

Bits required to transmit the results of *N* measurements is 2.69*N* < 3*N* bits !

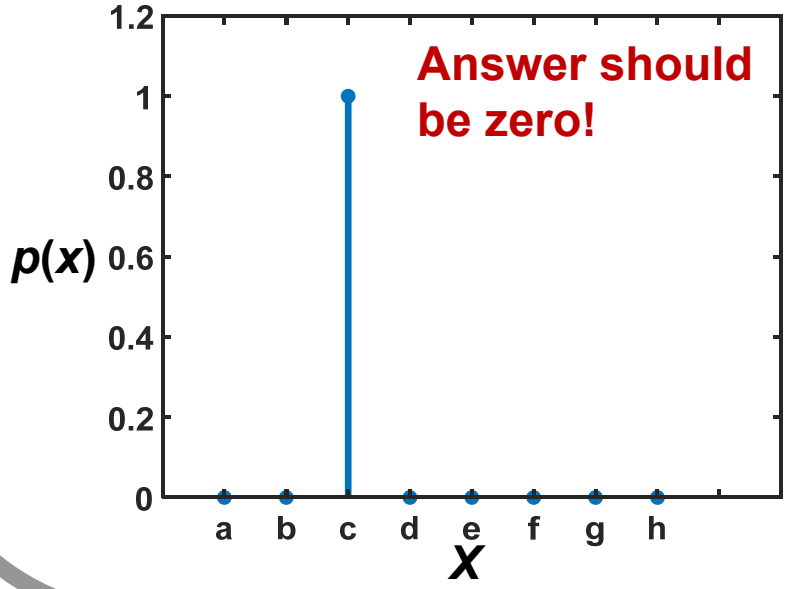
Entropy and Information

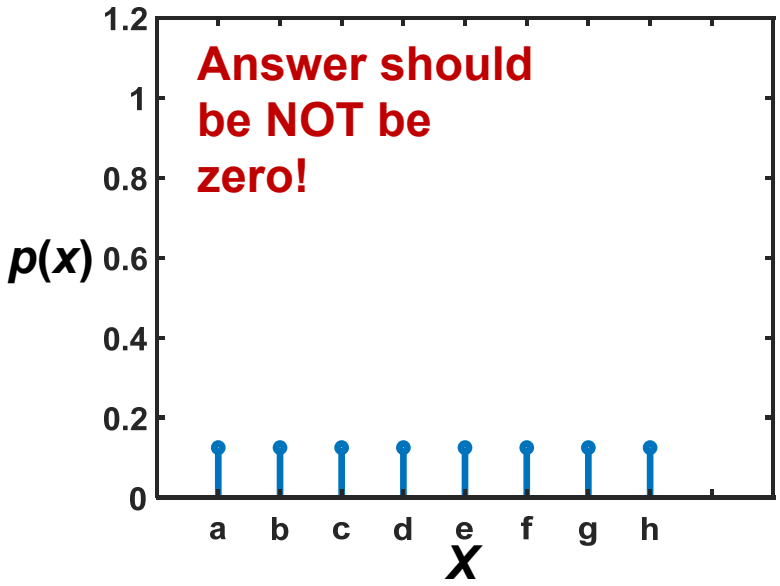
Consider a random variable *X* with the following probability distribution:



Question: If you were to find out the outcome of the random variable (say after making a measurement) then how much information did you acquire?

Hint: How many bits do you need, on average, to convey the result of the measurement to your friend?





Information and Asymptotic Equipartition Property

Consider a random variable X with the probability distribution $P(x)$

$$X = \{a, b, c, d, e, f, g, h\}$$

Suppose we measure X exactly N times (N is very large) and we plan to send the results to a friend

The results of these measurements is the sequence: $x_1, x_2, x_3, \ldots\ldots x_N$

The joint a-priori probability for this sequence is: $P(x_1, x_2, x_3, \ldots\ldots x_N) = \prod_{i=1}^{N} P(x_i)$

$$\Rightarrow -\frac{1}{N}\log_2 P(x_1, x_2, x_3, \ldots\ldots x_N) = -\frac{1}{N}\sum_{i=1}^{N} \log_2 P(x_i)$$

As $N \to \infty$ then:

$$\underset{N \to \infty}{\text{Limit}} \; -\frac{1}{N}\sum_{i=1}^{N} \log_2 P(x_i) = -\sum_{x} P(x)\log_2 P(x) = H(X)$$

Asymptotic Equipartition Property (AEP)

This means that as $N \to \infty$,

$$P(x_1, x_2, x_3, \ldots\ldots x_N) \to 2^{-NH(X)}$$

- This means that as $N \to \infty$ there can only be $2^{NH(X)}$ different result sequences that are probabilistically likely (and each one of them has the same a-priori probability)
- Therefore, we only need $NH(X)$ bits to encode any result sequence that is likely to occur
- This means on average we need only $H(X)$ bits per result to encode it

Entropy and Information

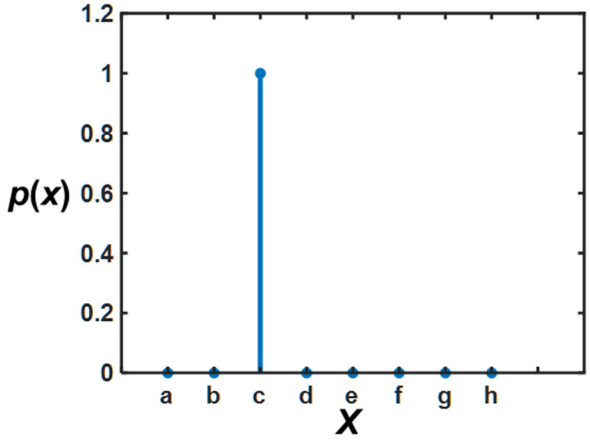
Entropy:

The amount of information (in bits) that is gained by learning about the outcome of a measurement of a random variable is given by the entropy function:

$$H(X) = -\sum_{x} p(x)\log_2\left[p(x)\right]$$

Equivalently, entropy is the minimum number of bits required on average to transmit reliably the outcome of a measurement of the random variable

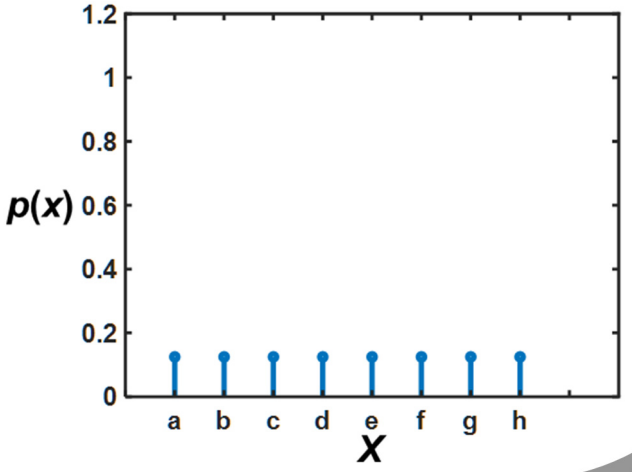
Case 1:



Completely deterministic scenario!

$$H(X) = 0$$

Case 2;



Completely random scenario!

$$p(x) = \frac{1}{8} \qquad \{\ x = a, b, \ldots\ldots h$$

$$H(X) = \log_2 8 = 3 \text{ bits}$$

Entropy and Information

Value of *X*	Probability	Coding #1	Coding #2
a	1/32	000	00000
b	1/32	001	00001
c	1/8	010	011
d	1/4	011	10
e	1/8	100	001
f	1/8	101	010
g	1/16	110	0001
h	1/4	111	11

$$H(X) = -\sum_{x} p(x) \log_2 \left[p(x) \right]$$

$$= \mathbf{2.69} \quad \mathbf{bits!}$$

Entropy and Data Compression

A classical *message* M consists of a very long sequence of *letters* y_i :

$$M = \{y_1, y_2, y_3, \ldots\ldots\ldots y_N\}$$

in which each letter belongs to an *alphabet* A of k letters:

$$A = \{a_1, a_2, a_3, \ldots\ldots\ldots a_k\}$$

In the message, each letter a_i occurs with an a-priori probability p_i

The *entropy of the message* is then:

$$H(C) = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

Shannon's Source Coding Theorem:

A classical message of N letters, as described above, can be reliably compressed to just $NH(C)$ bits and recovered with an error probability that approaches zero as the message length N becomes large

C. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, July and October 1948

Entropy Maximizing Distributions

Continuous random variable:

What probability distribution maximizes *H*(*X*) subject to the constrains:

$$\langle x \rangle = x_o \qquad \left\langle (x - x_o)^2 \right\rangle = \sigma^2 \qquad \{-\infty < x < \infty\}$$

Answer: A Gaussian (or Normal) distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_o)^2}{2\sigma^2}} = \mathbb{N}(x_o, \sigma^2) \quad \Rightarrow \quad H(X) = \frac{1}{2}\log_2\left(2\pi e\sigma^2\right)$$

Discrete random variable:

What probability distribution maximizes *H*(*N*) subject to the constrains:

$$\langle n \rangle = n_o \qquad \{n = 0, 1, 2, 3 \ldots\ldots\ldots \infty\}$$

Answer: A Thermal (or Bose-Einstein) distribution

$$P(n) = \frac{1}{1+n_o}\left(\frac{n_o}{1+n_o}\right)^n \quad \Rightarrow \quad H(N) = \log_2(1+n_o) + n_o \log_2\left(1+\frac{1}{n_o}\right)$$

Conditional Entropy

How much information can be obtained on average from learning about the outcome of a measurement of a random variable *Y* if the outcome of the measurement of another random variable *X* is known?

$$H(Y \mid X) = -\sum_{x} p(x) \sum_{y} p(y \mid x) \log_2 \left[p(y \mid x) \right]$$

$$= \sum_{x} p(x)\, H(Y \mid X = x)$$

$$= -\sum_{x,y} p(x,y) \log_2 \left[p(y \mid x) \right]$$

Cases:

$H(Y \mid X) = H(Y)$ **Iff *X* and *Y* are independent random variables**

$H(Y \mid X) = 0$ **iff *X* completely determines *Y***

Mutual Information

Difference between the information obtained on average from learning about the outcome of a measurement of a random variable Y and the information obtained on average from learning about the outcome of a measurement of a random variable Y if the outcome of the measurement of another random variable X is known

$$I(Y:X) = H(Y) - H(Y \mid X) = I(X:Y) = H(X) - H(X \mid Y)$$

$$= \sum_{x,y} p(x,y) \log_2 \left[\frac{p(x,y)}{p(x)p(y)} \right]$$

Mutual information quantifies how much information one random variable conveys about another random variable

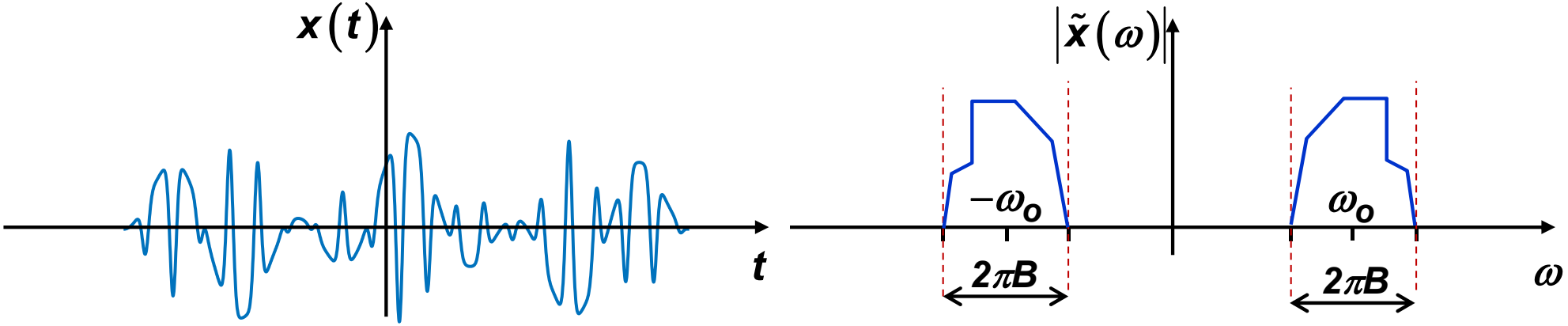
Cases:

$I(Y:X) = H(Y) - H(Y \mid X) = H(Y)$ **iff X completely determines Y**

$I(Y:X) = H(Y) - H(Y \mid X) = 0$ **Iff X and Y are independent random variables**

Classical Signals and Degrees of Freedom

How many degrees of freedom do bandwidth-limited real classical signals have?



Answer: 2*B* real degrees of freedom per second (Nyquist Theorem) where *B* is the single-sided signal bandwidth in Hertz (not radians)

Recall from Chapter 5 that real narrowband signals can always be written as:

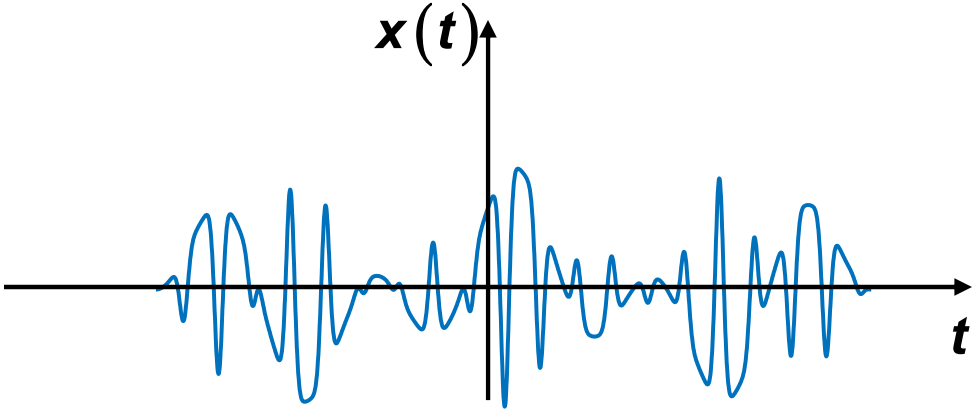
$$x(t) = \mathrm{Re}\left\{a(t)e^{-i\omega_o t}\right\} \longrightarrow a(t) = x_1(t) + ix_2(t)$$

$$x(t) = x_1(t)\cos(\omega_o t) + x_2(t)\sin(\omega_o t)$$

So each time-domain sample of the signal carries information on two real degrees of freedom

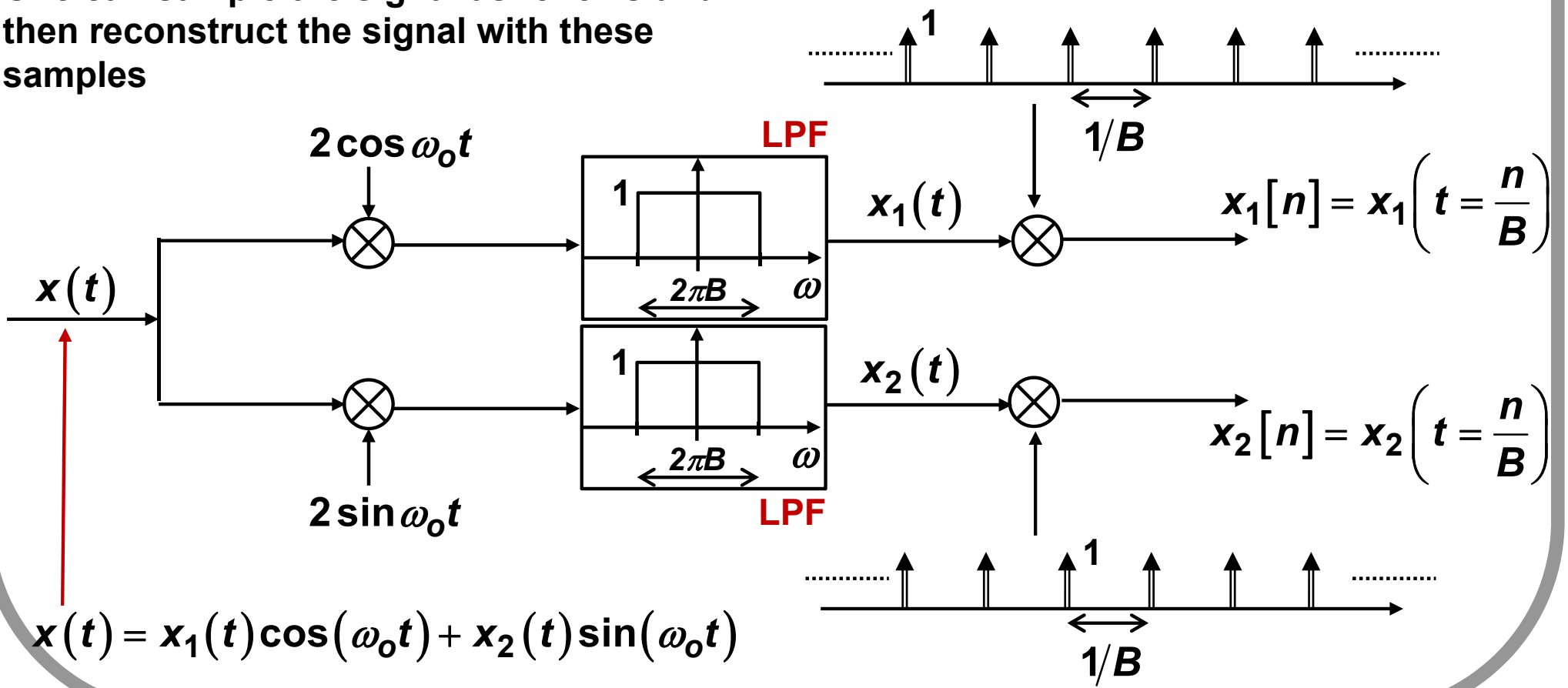
And by Nyquist theorem, a band-limited signal can have at max *B* independent samples per second (where *B* is the single-sided bandwidth in Hertz)

Classical Signals and Degrees of Freedom



$$x[n] = x\left(t = \frac{n}{B}\right)$$

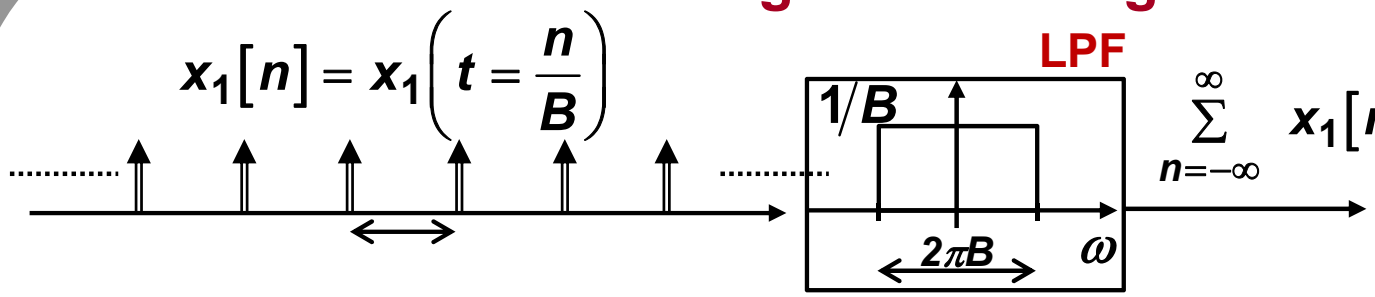
One can sample the signal as follows and then reconstruct the signal with these samples



$$x(t) = x_1(t)\cos(\omega_o t) + x_2(t)\sin(\omega_o t)$$

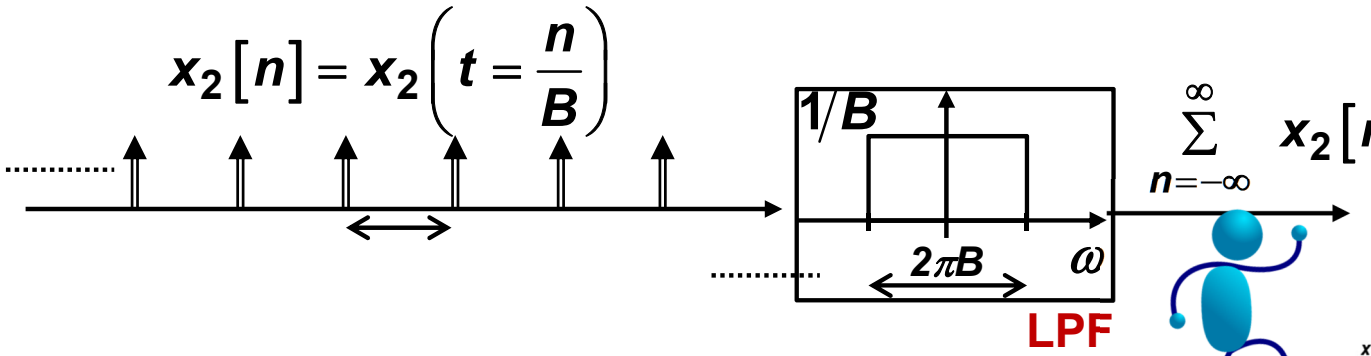
Classical Signals and Degrees of Freedom

$$x_1[n] = x_1\left(t = \frac{n}{B}\right)$$



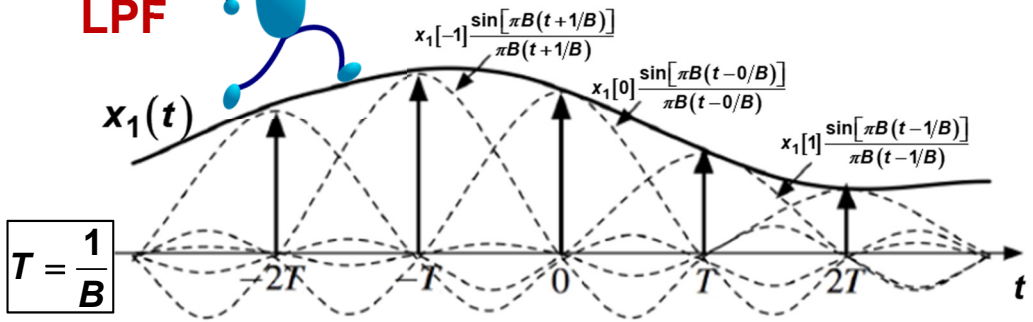
$$\sum_{n=-\infty}^{\infty} x_1[n]\frac{\sin\left[\pi B(t-n/B)\right]}{\pi B(t-n/B)} = x_1(t)$$

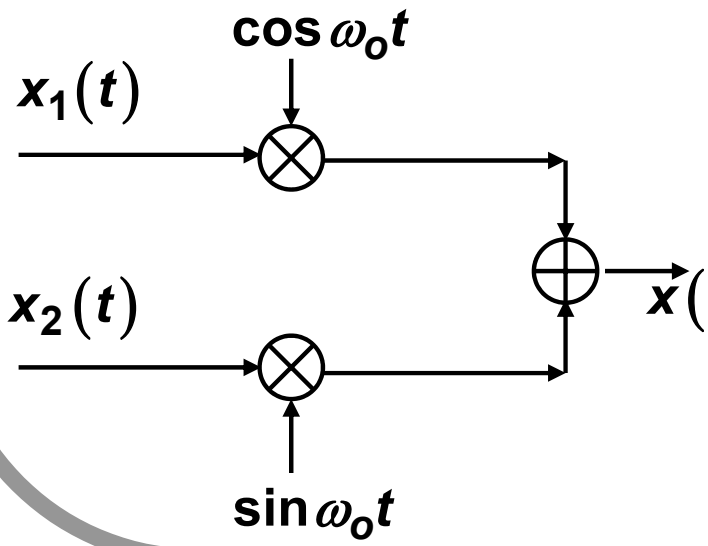
$$x_2[n] = x_2\left(t = \frac{n}{B}\right)$$



$$\sum_{n=-\infty}^{\infty} x_2[n]\frac{\sin\left[\pi B(t-n/B)\right]}{\pi B(t-n/B)} = x_2(t)$$

Construction of $x_1(t)$:





$$x(t) = \sum_{n=-\infty}^{\infty} x_1[n]\frac{\sin\left[\pi B(t-n/B)\right]}{\pi B(t-n/B)}\cos(\omega_o t)$$

$$+ \sum_{n=-\infty}^{\infty} x_2[n]\frac{\sin\left[\pi B(t-n/B)\right]}{\pi B(t-n/B)}\sin(\omega_o t)$$

Time Domain Basis

Note that the signal can be expended in an orthogonal time-domain basis set:

$$x(t) = \sum_{n=-\infty}^{\infty} x_1[n] \frac{\sin\left[\pi B(t - n/B)\right]}{\pi B(t - n/B)} \cos(\omega_o t) + \sum_{n=-\infty}^{\infty} x_2[n] \frac{\sin\left[\pi B(t - n/B)\right]}{\pi B(t - n/B)} \sin(\omega_o t)$$

The time-domain and time-localized functions,

$$\frac{\sin\left[\pi B(t - n/B)\right]}{\pi B(t - n/B)} \cos(\omega_o t) \qquad \frac{\sin\left[\pi B(t - n/B)\right]}{\pi B(t - n/B)} \sin(\omega_o t)$$

form a complete orthogonal set that can be used to expand any band-limited signal centered at frequencies $\pm\omega_o$

Classical Signals and Degrees of Freedom

$$x(t) = \mathrm{Re}\left\{a(t)e^{-i\omega_o t}\right\} \longrightarrow a(t) = x_1(t) + ix_2(t)$$

$$x(t) = x_1(t)\cos(\omega_o t) + x_2(t)\sin(\omega_o t)$$

Power of a narrowband signal:

$$P(t) = \frac{1}{2}x_1^2(t) + \frac{1}{2}x_2^2(t) = \frac{1}{2}\left|a(t)\right|^2$$

Total energy of a narrowband signal:

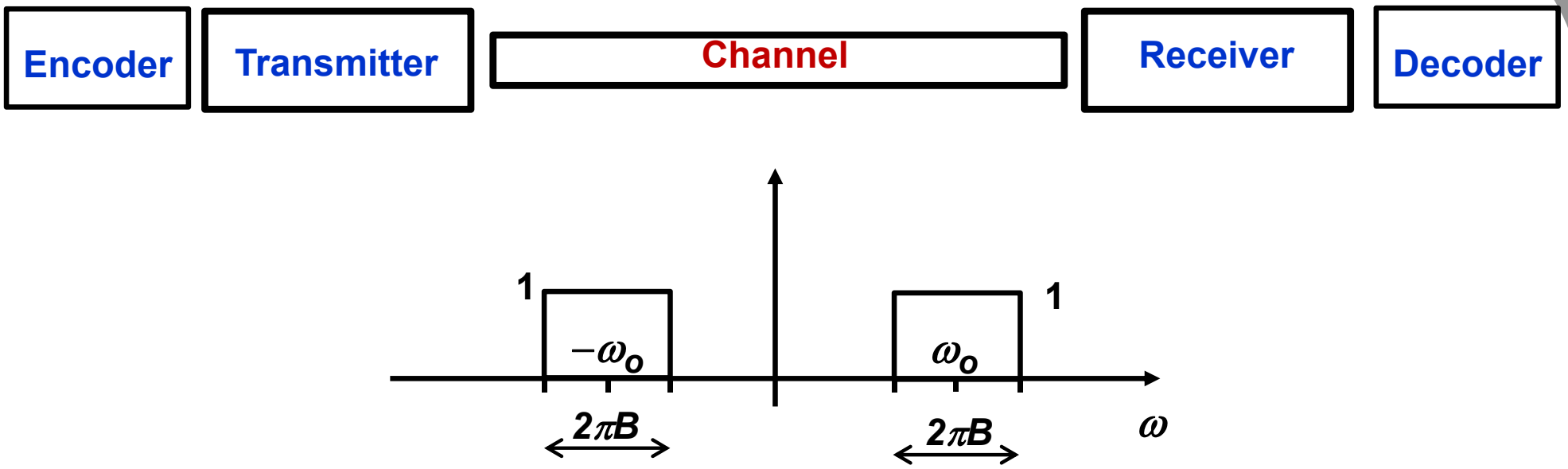
$$E = \int_{-\infty}^{\infty} dt\,\langle P(t)\rangle = \frac{1}{2}\int_{-\infty}^{\infty} dt\; x_1^2(t) + \frac{1}{2}\int_{-\infty}^{\infty} dt\; x_2^2(t)$$

$$= \frac{1}{2}\left\{\sum_{n=-\infty}^{\infty} x_1^2[n]\frac{1}{B} + \sum_{n=-\infty}^{\infty} x_2^2[n]\frac{1}{B}\right\}$$

Total energy is just half the energy of all the orthogonal sinc pulses in the signal:

$$x(t) = \sum_{n=-\infty}^{\infty} x_1[n]\frac{\sin\left[\pi B(t-n/B)\right]}{\pi B(t-n/B)}\cos(\omega_o t) + \sum_{n=-\infty}^{\infty} x_2[n]\frac{\sin\left[\pi B(t-n/B)\right]}{\pi B(t-n/B)}\sin(\omega_o t)$$

Classical Communication and Channel Capacity: AWGN Channel



Suppose one needs to send a message over a narrowband communication channel

- **One can map the message to the amplitudes of the two quadratures**
- **Note that one can send only $2B$ different quadrature values per second**

$$x(t) = \sum_{n=-\infty}^{\infty} x_1[n] \frac{\sin\left[\pi B(t - n/B)\right]}{\pi B(t - n/B)} \cos(\omega_o t) + \sum_{n=-\infty}^{\infty} x_2[n] \frac{\sin\left[\pi B(t - n/B)\right]}{\pi B(t - n/B)} \sin(\omega_o t)$$

Classical Communication and Channel Capacity: AWGN Channel

Suppose one sends *N* different quadratures through the channel in time *N*/2*B*:

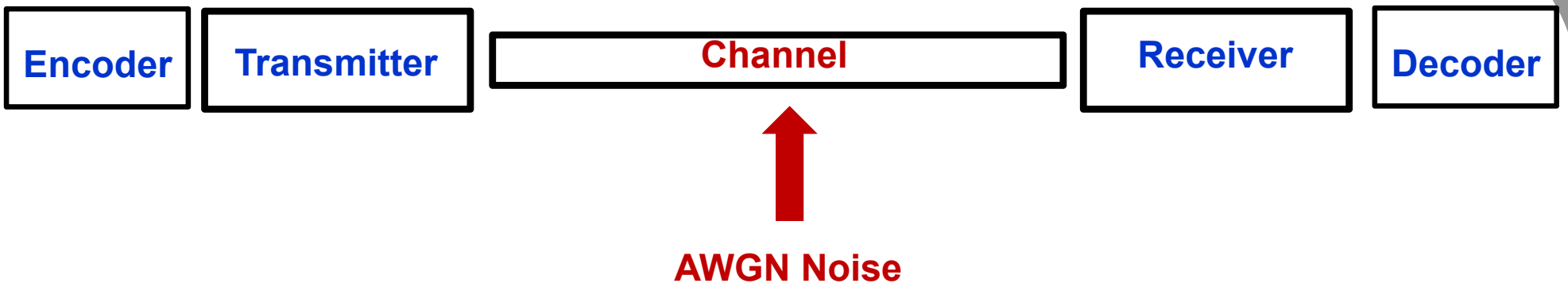
$$y[1], y[2], y[3], y[4] \ldots\ldots\ldots\ldots y[N]$$

The data to be transmitted and the mapping process will impart an a-priori probability distribution *P*(*y*) for the quadrature amplitudes

We assume there is also an energy/power constrain on the input:

$$\frac{1}{N}\sum_{n=1}^{N} y^2[n] \approx \int_{-\infty}^{\infty} y^2 p_{in}(y)\, dy \le BE = P = \text{average power}$$

Classical Communication and Channel Capacity: AWGN Channel



The **channel adds noise** so that the received quadrature is:

$$z[n] = y[n] + \Delta f[n]$$

Where $f[n]$ represents **zero-mean white Gaussian noise**:

$$\langle \Delta f[n] \rangle = 0 \qquad \langle \Delta f[n] \Delta f[m] \rangle = M\delta_{n,m}$$

$$M = S_{\Delta f \Delta f}(\omega_o) B$$

$$P(\Delta f) = \mathbb{N}(0, M)$$

Question: how much information (in bits) can be communicated over this channel using these N quadratures?

Classical Communication and Channel Capacity: AWGN Channel

$$z[n] = y[n] + \Delta f[n]$$

$$\Rightarrow \left\langle z^2[n] \right\rangle = \left\langle y^2[n] \right\rangle + \left\langle \Delta f^2[n] \right\rangle$$

$$\Rightarrow \left\langle z^2[n] \right\rangle \leq P + M$$

Represent each possible value of the received quadrature as a point in an *N*-dimensional space of radius:

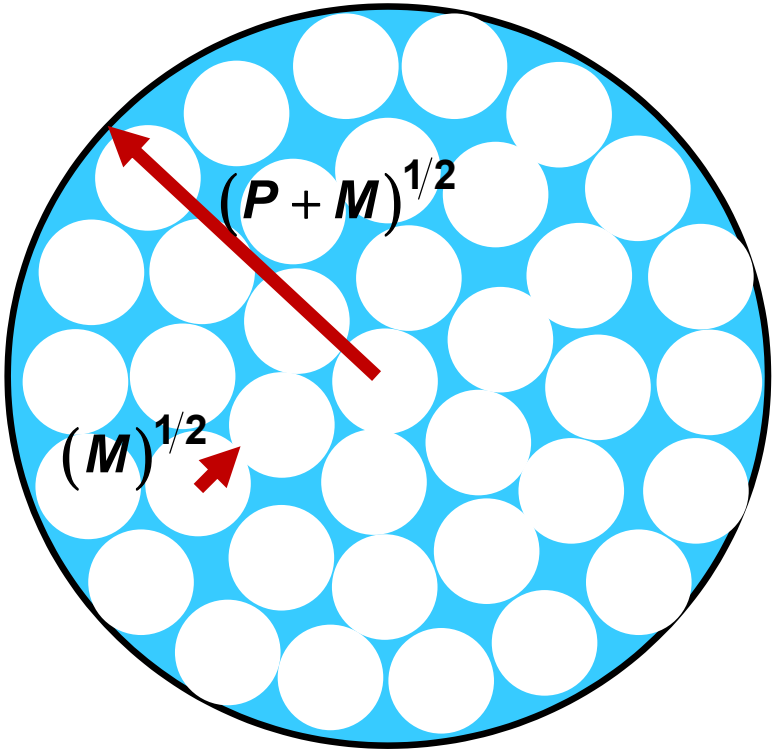
$$(P+M)^{1/2}$$

The noise is represented by an error region of radius:

$$(M)^{1/2}$$

around each quadrature value that can be received

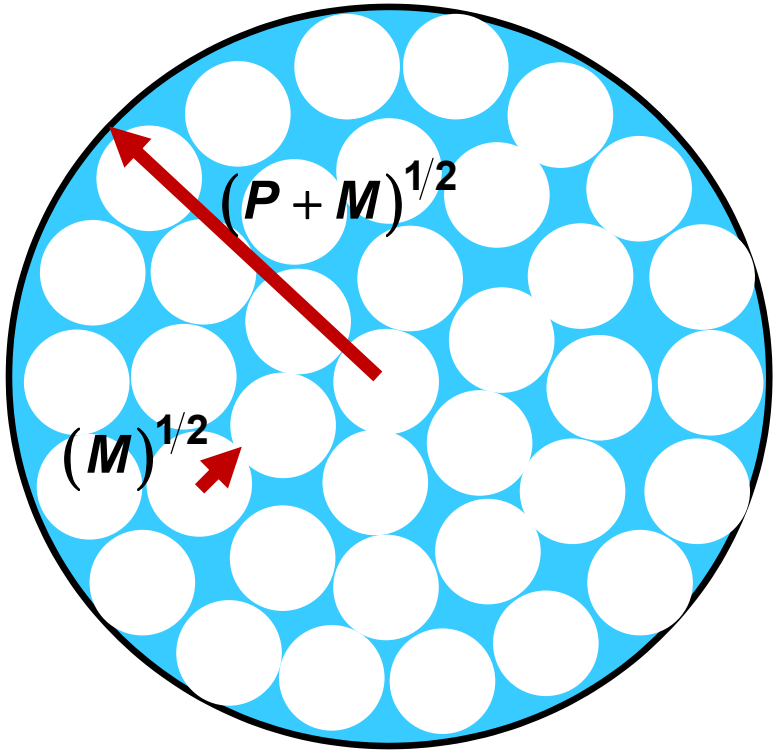
Now consider an *N*-dimensional space



Classical Communication and Channel Capacity: AWGN Channel

Encoder	Transmitter	Channel	Receiver	Decoder

The number of distinct quadratures that can be received and distinguished from each other in the presence of noise is equal to the number of non-overlapping *N*-spheres of radius $\sqrt{M}$ that can be packed in a *N*-sphere of radius $\sqrt{P+M}$



Which equals = $$\frac{(P+M)^{N/2}}{(M)^{N/2}} = \left(1+\frac{P}{M}\right)^{N/2}$$

So the information in bits that can be transferred using *N*-quadratures is:

$$\log_2\left(1+\frac{P}{M}\right)^{N/2} = \frac{N}{2}\log_2\left(1+\frac{P}{M}\right)$$

Classical Communication and Channel Capacity: AWGN Channel

- So the information in bits that can be transferred using *N*-quadratures is:

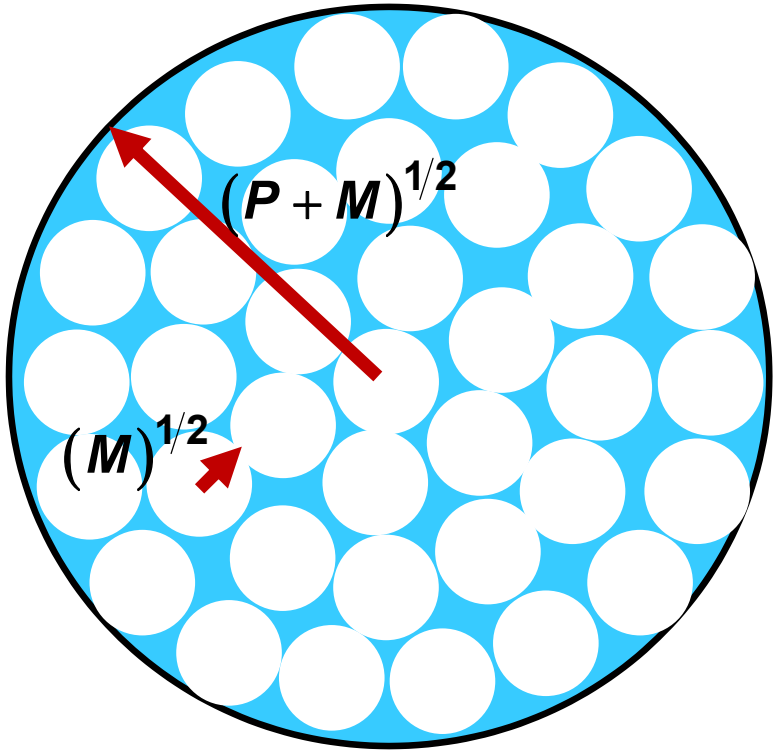
$$\log_2\left(1+\frac{P}{M}\right)^{N/2} = \frac{N}{2}\log_2\left(1+\frac{P}{M}\right)$$

- Then the information in bits that can be transferred using *one* quadrature is:

$$= \frac{1}{2}\log_2\left(1+\frac{P}{M}\right)$$

- Since we can send $2B$ quadratures per second through the channel, the information in bits that can be transferred per second is:

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



C. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, July and October 1948

***C* is called the capacity of the classical AWGN channel**

Mutual Information and Channel Capacity

The channel capacity (per usage) with input *Y* and output *Z* is defined as the maximal of the mutual information over all possible input distributions taking into account all realistic constrains (such as the power/energy constrain):

$$C = \max_{p_{in}(y)} I(Z:Y) = \max_{p_{in}(y)} H(Z) - H(Z \mid Y)$$

Shannon's Noisy Channel Coding Theorem

Any amount of information (in bits) less than or equal to *C* can be reliably transmitted and recovered per usage of a noisy channel with an error probability that approaches zero as the number of uses of the channel becomes large

C. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, July and October 1948

Mutual Information and Channel Capacity: AWGN Channel

The channel capacity (per usage) is more formally defined as the maximal of the mutual information over all possible input distributions taking into account the power/energy constrain:

$$C = \max_{p_{in}(y)} I(Z:Y) = \max_{p_{in}(y)} H(Z) - H(Z \mid Y) \qquad \left\{ \int_{-\infty}^{\infty} y^2 p_{in}(y)\, dy \le P \right.$$

For AWGN Channel:

$$z[n] = y[n] + \underbrace{\Delta f[n]}_{\text{AWGN}}$$

$$\begin{cases} \langle \Delta f[n] \rangle = 0 \\ \langle \Delta f[n] \Delta f[m] \rangle = M \delta_{n,m} \end{cases}$$

$$C = \max_{p_{in}(y)} H(Z) - \int_{-\infty}^{\infty} dy\; p_{in}(y) H(Z \mid Y = y) = \max_{p_{in}(y)} H(Z) - \frac{1}{2}\log_2(2\pi e M)$$

Mutual information will be maximized if the output *Z* is Gaussian, and *Z* will be Gaussian if the input *Y* is Gaussian

Mutual Information and Channel Capacity: AWGN Channel

$$C = \max_{p_{in}(y)} H(Z) - \frac{1}{2}\log_2(2\pi e M)$$

Mutual information will be maximized if the output *Z* is Gaussian, and *Z* will be Gaussian if the input *Y* is Gaussian

$$z[n] = y[n] + \Delta f[n]$$

If:

$$p(\Delta f) = \mathbb{N}(0, M)$$

Then:

$$p_{out}(z \mid y) = \mathbb{N}(y, M)$$

And then if:

$$p_{out}(y) = \mathbb{N}(0, P)$$

Then:

$$p_{out}(z) = \int_{-\infty}^{\infty} dy \; p_{out}(z \mid y)\, p_{in}(y) = \mathbb{N}(0, P + M)$$

Mutual Information and Channel Capacity: AWGN Channel

For AWGN Channel:

$$C = \max_{p_{in}(y)} H(Z) - \frac{1}{2}\log_2(2\pi e M)$$

So if we assume for input *Y* the a-priori probability distribution:

$$p_{in}(y) = \frac{1}{\sqrt{2\pi P}} e^{-\frac{y^2}{2P}} = \mathbb{N}(0, P)$$

Satisfies the constrain:

$$\int_{-\infty}^{\infty} y^2 p_{in}(y)\, dy \le P$$

Then the output *Z* will have the probability distribution:

$$p_{out}(z) = \frac{1}{\sqrt{2\pi(P+M)}} e^{-\frac{z^2}{2(P+M)}} = \mathbb{N}(0, P+M)$$

And the channel capacity (per quadrature) becomes:

$$C = \frac{1}{2}\log_2(2\pi e(P+M)) - \frac{1}{2}\log_2(2\pi e M) = \frac{1}{2}\log_2\left(1+\frac{P}{M}\right)$$

Same as before!!

Quantum Information: The Basics

The unit of quantum information is a "qubit" (not a bit):

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

Unlike the classical bit, a qubit can be in a superposition of the two logical states at the same time

The density operator:

The state of a quantum system is represented by a density operator $\hat{\rho}$

Density operator for a pure state: $\hat{\rho} = |\phi\rangle\langle\phi|$

Density operator for a mixed state (i.e. an ensemble of pure states): $\hat{\rho} = \sum_i p_i |\phi_i\rangle\langle\phi_i|$

Density operator for an ensemble of mixed states: $\hat{\rho} = \sum_i p_i \hat{\rho}_i$

Quantum Information: Von Neumann Entropy

The "information" content of a quantum state is related to the Von Neumann entropy:

$$S(\hat{\rho}) = -\mathrm{Tr}\left[\hat{\rho}\log_2(\hat{\rho})\right]$$

The Von Neumann entropy plays three roles (that we know of so far):

1) It *quantifies* the quantum information content in qubits of a quantum state (i.e. the minimum number of qubits needed to reliably encode the quantum state)

2) It also *quantifies* the classical information in bits that can be gained about the quantum state by making the best possible measurement

3) It also quantifies the amount of entanglement in bipartite pure states

As you will see, the Von Neumann entropy will not always give the answer to the question we will ask!

Von Neumann Entropy: Some Properties

$$S(\hat{\rho}) = -\mathrm{Tr}\left[\hat{\rho}\log_2(\hat{\rho})\right]$$

1) Suppose:

$$\hat{\rho} = |\phi\rangle\langle\phi| \longrightarrow$$ **A pure state**

$$\Rightarrow S(\hat{\rho}) = 0$$

2) Suppose:

$$\hat{\rho} = \sum_i p_i |\phi_i\rangle\langle\phi_i| \longrightarrow$$ **An ensemble of pure ORTHOGONAL states**

$$\Rightarrow S(\hat{\rho}) = -\sum_i p_i \log_2\left[p_i\right] = H$$ **= Shannon entropy of the ensemble**

If the states in the ensemble were not all completely orthogonal then: $S(\hat{\rho}) < H$

3) Suppose:

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i \longrightarrow$$ **An ensemble of mixed states but the mixed states in the ensemble have support on ORTHOGONAL spaces**

Von Neumann Entropy: Some Properties

3) Suppose:

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i \longrightarrow$$

An ensemble of mixed states but the mixed states in the ensemble have support on ORTHOGONAL spaces

$$\Rightarrow S(\hat{\rho}) = -\mathrm{Tr}\left[\hat{\rho}\log_2(\hat{\rho})\right] = -\mathrm{Tr}\left[\left(\sum_i p_i \hat{\rho}_i\right)\log_2\left(\sum_j p_j \hat{\rho}_j\right)\right]$$

$$= -\mathrm{Tr}\left\{\sum_i \left[(p_i \hat{\rho}_i)\log_2(p_i \hat{\rho}_i)\right]\right\}$$

$$= -\mathrm{Tr}\left\{\sum_i \left[(p_i \hat{\rho}_i)\log_2(p_i) + (p_i \hat{\rho}_i)\log_2(\hat{\rho}_i)\right]\right\}$$

$$= -\sum_i p_i \log_2[p_i] + \sum_i p_i S(\hat{\rho}_i) = H + \sum_i p_i S(\hat{\rho}_i)$$

4) Change of basis:

Entropy is invariant under a unitary transformation (or change of basis)

$$S\left(U\hat{\rho}U^{+}\right) = S(\hat{\rho})$$

Quantum Messages

A quantum *message M* consists of a very long sequence of *letters (or quantum states)* $\hat{\sigma}_i$:

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\}$$

in which each letter belongs to an *alphabet A* of *k* letters:

$$A = \{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \ldots\ldots\ldots \hat{\rho}_k\}$$

In the message, each letter $\hat{\rho}_i$ occurs with an a-priori probability p_i

The density operator for each letter in the message is then:

$$\hat{\sigma} = \sum_{i=1}^{k} p_i \, \hat{\rho}_i$$

The density operator for the entire message of *N* letters is then:

$$\hat{\sigma}^N = \hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes \ldots\ldots\ldots \otimes \hat{\sigma}$$

Question: is it possible to compress this long message to a smaller Hilbert space requiring fewer qubits without comprising the fidelity of the message?

Quantum Fidelity

How can we tell if two quantum states are identical, similar, not so similar, etc?

Example: in classical information theory we can judge the similarity or difference between random variables *Y* and *X* by the mean square difference:

$$\left\langle (x-y)^2 \right\rangle = \int dx\, dy\, (x-y)^2\, P(x,y)$$

This is not the only measure used

Quantum Fidelity:

Given two quantum states, $\hat{\rho}$ and $\hat{\sigma}$, the fidelity *F*, a measure of the *closeness* between them, is generally defined as the quantity:

$$F(\hat{\rho},\hat{\sigma}) = \left(\mathrm{Tr}\left\{ \sqrt{\sqrt{\hat{\rho}}\,\hat{\sigma}\,\sqrt{\hat{\rho}}} \right\} \right)^2 = F(\hat{\sigma},\hat{\rho})$$

This is not the only measure used

Example: Suppose,

$$\hat{\rho} = |\phi\rangle\langle\phi|$$

$$\hat{\sigma} = |\psi\rangle\langle\psi|$$

$$F(\hat{\rho},\hat{\sigma}) = |\langle\phi|\psi\rangle|^2$$

Quantum Messages and Quantum Data Compression

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\} \qquad \hat{\sigma} = \sum_{i=1}^{k} p_i \ \hat{\rho}_i \qquad A = \{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \ldots\ldots\ldots \hat{\rho}_k\}$$

A quantum *message M* consisting of a very long sequence of *letters (or quantum states)* $\hat{\sigma}_i$:

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\}$$

can be compressed to *NC* qubits, in the limit of large *N*, without loss of fidelity, where:

$$S(\hat{\sigma}) \geq C \geq I(\hat{\sigma}) = S(\hat{\sigma}) - \sum_{i=1}^{k} p_i S(\hat{\rho}_i)$$

The lower limit is achievable if the alphabet *C* represents pure states (not necessarily orthogonal), or if the different letters in the alphabet commute

B. Schumacher, Phys. Rev. A 51, 2738 (1995)
M. Horodecki, Phys. Rev. A 57, 3364 (1998)

Classical Information from Quantum Messages

Question: How much classical information in bits can be obtained from a quantum message by making the best possible measurement?

Generalized quantum measurements and POVMs:

The most general measurements to obtain classical information from quantum states can be described in terms of a complete set of positive Hermitian operators $\hat{F}_j$ which provide a resolution of the identity operator,

$$\sum_j \hat{F}_j = \hat{1}$$

These generalized measurements constitute a positive operator valued measure (POVM). The probability p_k that the outcome of a measurement on a quantum state $\hat{\rho}$ will be k is given as,

$$p_k = \mathrm{Tr}\left\{\hat{\rho}\,\hat{F}_k\right\}$$

Example: For a photon number measurement on a quantum state of light in a cavity, the POVM is formed by the operators $|n\rangle\langle n|$ and the probabilities are given as:

$$p(n) = \mathrm{Tr}\left\{\hat{\rho}\,|n\rangle\langle n|\right\}$$

Classical Information from Quantum Messages

Suppose a quantum message is made up of qubits:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

The quantum state is specified by two complex numbers and each can take an value

But the classical information that can be extracted from the above qubit is just one bit!!

Suppose the sender send the following two states with a-priori probability 1/2 each:

$$|0\rangle \qquad |1\rangle \qquad \hat{\sigma} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

One can use the following POVM:

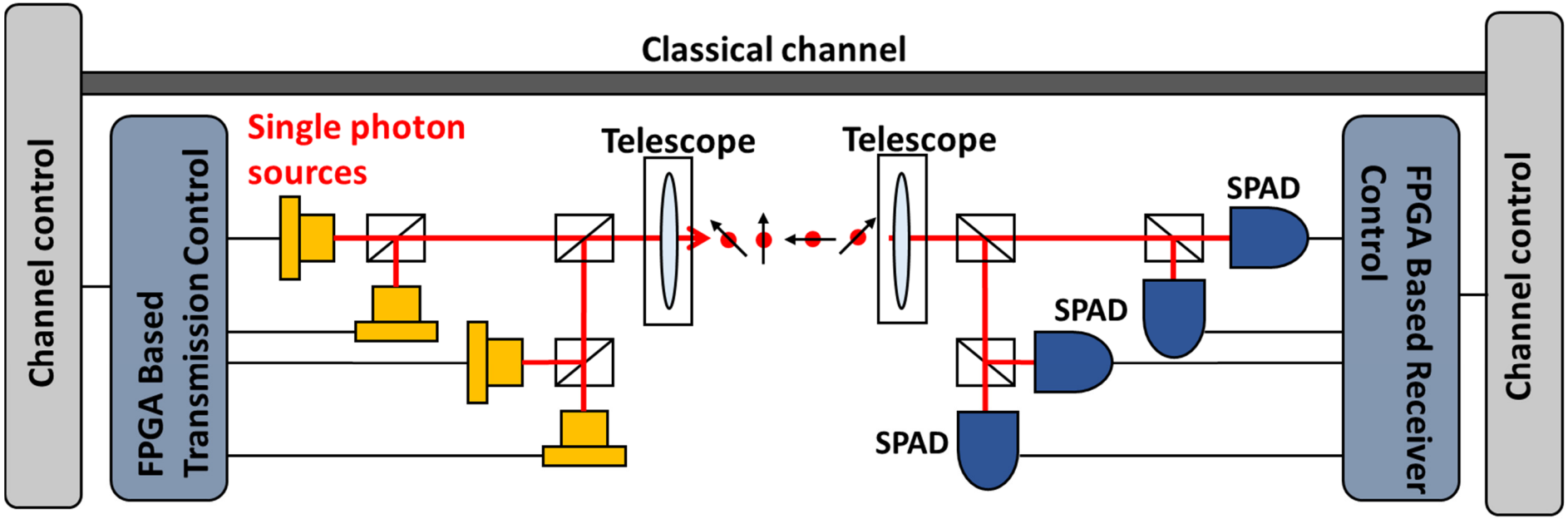
$$\hat{F}_0 = |0\rangle\langle 0| \qquad \hat{F}_1 = |1\rangle\langle 1| \qquad \sum_j \hat{F}_j = \hat{1}$$

And

$$S(\hat{\sigma}) = 1 \text{ bit}$$

Accessible information is only 1 bit!

Quantum Cryptography: The BB84 Protocol



The Holevo Bound: Case of Separate Measurements

A theorem, stated by Jim Gordon (without proof in 1964), and proved by Holevo (1973), gives an upper bound on the amount of classical information (in bits) that can be gained from a quantum message of *N* letters by making the best possible measurement on each letter individually:

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\} \qquad \hat{\sigma} = \sum_{i=1}^{k} p_i \, \hat{\rho}_i \qquad A = \{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \ldots\ldots\ldots \hat{\rho}_k\}$$

$$\hat{\sigma}^N = \underbrace{\hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes \ldots\ldots\ldots \otimes \hat{\sigma}}_{N \text{ terms}}$$

For the above quantum message, the obtained classical information $I(M:P)$ per letter about the preparation of the message, using the optimal measurement scheme, is bounded as follows:

$$\max_{\hat{F}} I(M:P) \le I(\hat{\sigma}) = S(\hat{\sigma}) - \sum_{i=1}^{k} p_i S(\hat{\rho}_i)$$

The upper limit in Holevo's theorem can be achieved if and only if the quantum states of all the letters in the alphabet *C* commute, i.e. $[\hat{\rho}_i, \hat{\rho}_k] = 0$

J. P. Gordon, in Quantum Electronics and Coherent Light, edited by P. A. Miles, Academic Press (1964).

A. S. Kholevo, Probl. Peredachi Inf. 9, 177 (1973).

Proof of the Holevo Bound: Case of Separate Measurements

If all the letters in the alphabet commute then they can all be diagonalized using a common orthonormal basis $\{|\phi_\alpha\rangle\}$, and:

$$\hat{\rho}_i = \sum_\alpha f_{i,\alpha} |\phi_\alpha\rangle\langle\phi_\alpha|$$

$$S(\hat{\rho}_i) = -\sum_\alpha f_{i,\alpha} \log_2 (f_{i,\alpha})$$

What this means is that even if the quantum letter or state is known, there is still entropy related to the outcome of measurements performed on it since it is not a pure state

So if we choose the POVM to be:

$$\hat{F}_\alpha = |\phi_\alpha\rangle\langle\phi_\alpha|$$

Then the probability p_α of measuring α is:

$$p_\alpha = \mathrm{Tr}\{\sigma \hat{F}_\alpha\} = \mathrm{Tr}\left\{\left(\sum_i p_i \hat{\rho}_i\right)|\phi_\alpha\rangle\langle\phi_\alpha|\right\} = \mathrm{Tr}\left\{\left(\sum_{i,\beta} p_i f_{i,\beta} |\phi_\beta\rangle\langle\phi_\beta|\right)|\phi_\alpha\rangle\langle\phi_\alpha|\right\} = \sum_i p_i f_{i,\alpha}$$

The entropy comes out to be:

$$S(\hat{\sigma}) = -\sum_\alpha p_\alpha \log_2 (p_\alpha) = H(M)$$

The Holevo Bound: Case of Separate Measurements

The maximum classical information is the mutual information between the measurement result and the preparation of the quantum state of each letter in the message:

$$\max_{\hat{F}} I(M:P) = H(M) - H(M|P)$$

$$= -\sum_{\alpha} p_{\alpha} \log_2(p_{\alpha}) - \sum_{i} p_i \left[-\sum_{\alpha} f_{i,\alpha} \log_2(f_{i,\alpha}) \right]$$

$$= S(\hat{\sigma}) - \sum_{i=1}^{k} p_i S(\hat{\rho}_i)$$

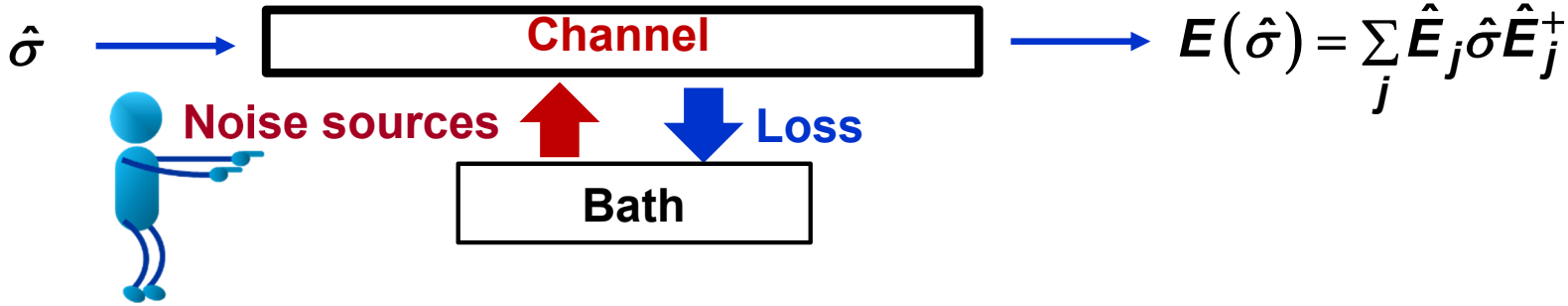
The Holevo Bound: Case of Block Measurements

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, ...\hat{\sigma}_N\} \qquad \hat{\sigma} = \sum_{i=1}^{k} p_i \ \hat{\rho}_i \qquad A = \{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, ...\hat{\rho}_k\}$$

$$\hat{\sigma}^N = \underbrace{\hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes ..... \otimes \hat{\sigma}}_{N \text{ terms}}$$

If instead of separate measurements on each letter of the message, if one is allowed to make optimal measurements on all *N* letters of the message at the same time then the upper limit in Holevo's theorem can be achieved even if the letters in the alphabet *C* do not commute, i.e. $\left[\hat{\rho}_i, \hat{\rho}_k\right] \neq 0$

Classical Information Over Quantum Channel



The channel is described by a trace preserving linear quantum operation *E* such that the density operator of each letter at the output of the channel is related to the density operator at the input of the channel by the relation:

$$\hat{\sigma} \to E(\hat{\sigma}) = \sum_j \hat{E}_j \hat{\sigma} \hat{E}_j^+ \qquad \left\{ \sum_j \hat{E}_j^+ \hat{E}_j = \hat{1} \right.$$

What is really happening is the following:

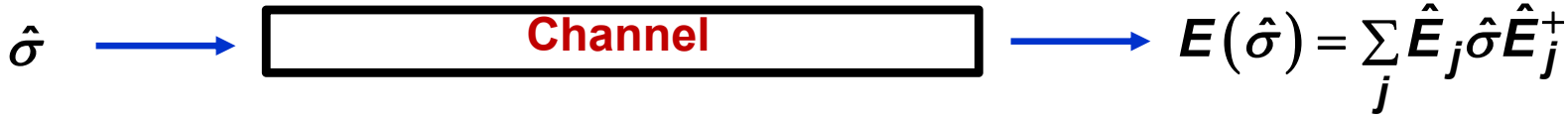
Initial state of the channel input and the bath → $\hat{\rho}_{\text{initial}} = \hat{\sigma} \otimes |B_0\rangle\langle B_0|$

After propagation through the channel; unitary time evolution → $\hat{\rho}_{\text{final}} = \hat{U}\left[\hat{\sigma} \otimes |B_0\rangle\langle B_0|\right]\hat{U}^+ \qquad \left\{ \hat{U} = e^{-i\frac{\hat{H}}{\hbar}t} \right.$

After trace over all bath degrees the channel output is: → $\hat{\rho}_{\text{output}} = \text{Tr}_{\text{Bath}}\left\{\hat{U}\left[\hat{\sigma} \otimes |B_0\rangle\langle B_0|\right]\hat{U}^+\right\}$

$$= \sum_j \hat{E}_j \hat{\sigma} \hat{E}_j^+$$

Classical Information Over Quantum Channel



The channel is described by a trace preserving linear quantum operation *E* such that the density operator of each letter at the output of the channel is related to the density operator at the input of the channel by the relation:

$$\hat{\sigma} \rightarrow E(\hat{\sigma}) = \sum_j \hat{E}_j \hat{\sigma} \hat{E}_j^+ \qquad \left\{ \sum_j \hat{E}_j^+ \hat{E}_j = \hat{1} \right.$$

Classical information over the channel is encoded in a quantum *message* *M* consists of a very long sequence of *letters (or quantum states)* $\hat{\sigma}_i$:

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\}$$

in which each letter belongs to an *alphabet A* of *k* letters:

$$A = \{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \ldots\ldots\ldots \hat{\rho}_k\}$$

In the message, each letter $\hat{\rho}_i$ occurs with an a-priori probability p_i

The density operators for each letter in the message and of the full message are then:

$$\hat{\sigma} = \sum_{i=1}^{k} p_i \, \hat{\rho}_i \qquad \hat{\sigma}^N = \hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes \ldots\ldots\ldots \otimes \hat{\sigma}$$

Question: How much classical information in bits can be communicated over the channel per letter?

Classical Information Over Quantum Channel: Channel Capacity

$\hat{\sigma}$ → Channel → $E(\hat{\sigma}) = \sum_j \hat{E}_j \hat{\sigma} \hat{E}_j^+$

The Holevo-Schumacher-Westmoreland (HSW) Theorem:

The classical capacity of this quantum channel per letter is:

$$C = \max_{p_i} S\left(E(\hat{\sigma})\right) - \sum_{i=1}^{k} p_i S\left(E(\hat{\rho}_i)\right)$$

The classical capacity of the quantum channel is achievable (even for non-commuting letters in the message) if the receiver is allowed to make block measurements on all received letters

Note:

This capacity is also called the fixed-alphabet product-state capacity, since 1) the optimization is not performed over the choice of input letters $\hat{\sigma}_i$, and 2) the input letters are not assumed to be entangled over multiple uses of the channel and therefore the input density operator is in a tensor product form

Classical Information Over a Photonic Channel: Photon Number States and Photon Number Detection

Channel

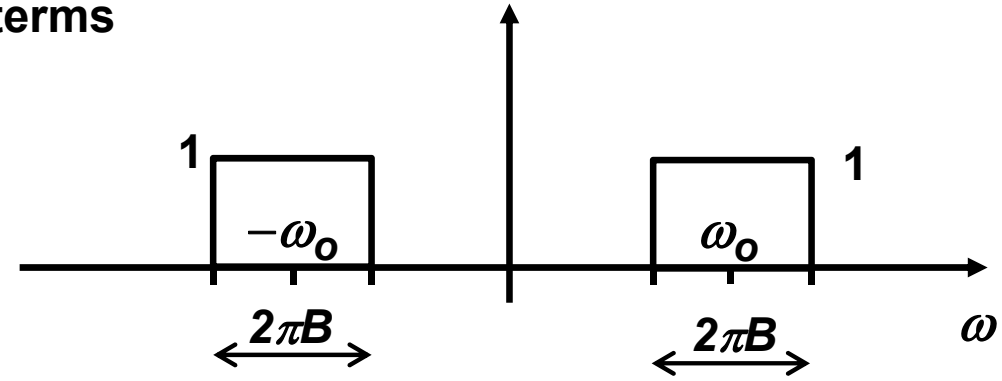
$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\}$$

$$A = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|, \ldots\ldots\ldots |n\rangle\langle n| \ldots\ldots\ldots\}$$

$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n|$$

$$\hat{\sigma}^N = \underbrace{\hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes \ldots\ldots\ldots \otimes \hat{\sigma}}_{N \text{ terms}}$$

- The channel is bandlimited
- Optical pulses are used and the classical information is encoded in the number of photons in each optical pulse



From previous discussion, at max B such pulses can be sent per second

Power Constrain:

$$\sum_{n=0}^{\infty} p_{in}(n)\, nB\hbar\omega_o = P$$

$$\Rightarrow \sum_{n=0}^{\infty} p_{in}(n)\, n = n_o = \frac{P}{B\hbar\omega_o}$$

Classical Information Over a Photonic Channel: Photon Number States and Photon Number Detection

$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n)\,|n\rangle\langle n| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n)\,|n\rangle\langle n|$$

The maximum classical information transmitted over the channel per letter is then:

$$C = \max_{p_i} S\left(E(\hat{\sigma})\right) - \sum_{i=1} p_i S\left(E(\hat{\rho}_i)\right)$$

$$= \max_{p_i} S(\hat{\sigma}) - \sum_{i=1} p_i S(\hat{\rho}_i)$$

$$= \max_{p_{in}(n)} S\left(\sum_{n=0}^{\infty} p_{in}(n)|n\rangle\langle n|\right) - \sum_{n=0}^{\infty} p_{in}(n) S\left(|n\rangle\langle n|\right)$$

$$= \max_{p_{in}(n)} S\left(\sum_{n=0}^{\infty} p_{in}(n)|n\rangle\langle n|\right)$$

The ideal POVM is:

$$\hat{F}_n = |n\rangle\langle n|$$

$$\sum_n \hat{F}_n = \hat{1}$$

The entropy is maximized for a thermal distribution of photons in every pulse!

$$p_{in}(n) = \frac{1}{1+n_o}\left(\frac{n_o}{1+n_o}\right)^n$$

$$\sum_{n=0}^{\infty} p_{in}(n)\, n = n_o = \frac{P}{B\hbar\omega_o}$$

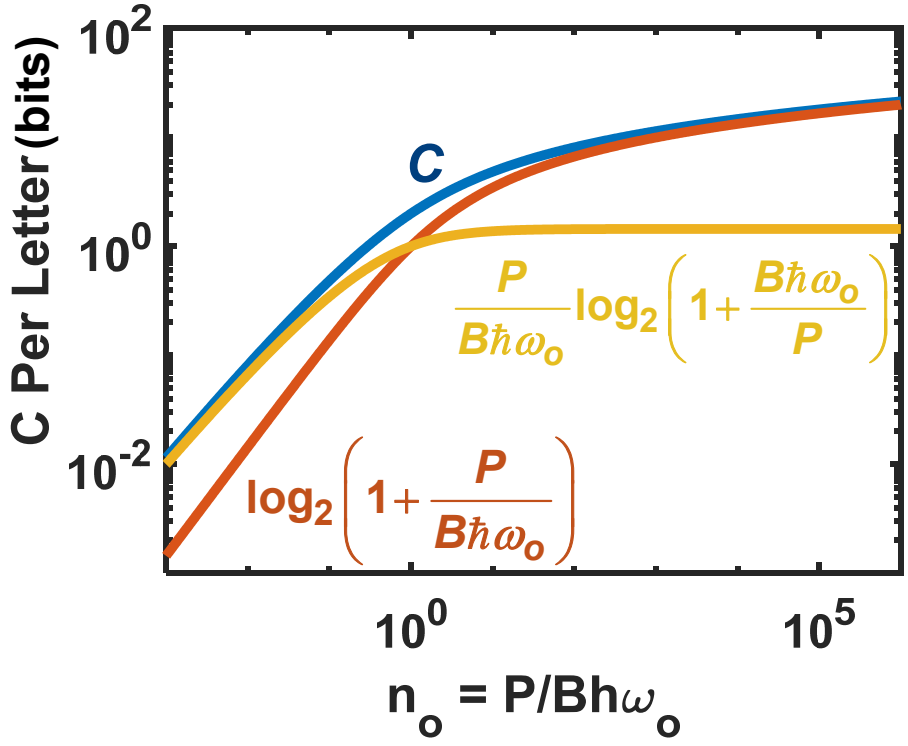
$$C = \log_2\left(1+\frac{P}{B\hbar\omega_o}\right) + \frac{P}{B\hbar\omega_o}\log_2\left(1+\frac{B\hbar\omega_o}{P}\right)$$

Classical Information Over a Photonic Channel: Photon Number States and Photon Number Detection

$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n)\,|n\rangle\langle n| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n)\,|n\rangle\langle n|$$

The capacity in bits per letter is:

$$C = \log_2\left(1+\frac{P}{B\hbar\omega_o}\right) + \frac{P}{B\hbar\omega_o}\log_2\left(1+\frac{B\hbar\omega_o}{P}\right)$$



The capacity in bits per second is:

$$C = B\log_2\left(1+\frac{P}{B\hbar\omega_o}\right) + \frac{P}{\hbar\omega_o}\log_2\left(1+\frac{B\hbar\omega_o}{P}\right)$$

Classical Information Over a Photonic Channel: Photon Number States and Photon Number Detection

$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n)\,|n\rangle\langle n| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n)\,|n\rangle\langle n|$$

The High Power Limit:

In the limit $P >> B\hbar\omega_o$ the capacity (bits/s) becomes:

$$C = B\log_2\left(1+\frac{P}{B\hbar\omega_o}\right)$$

Compare the above to the classical AWGN channel result:

$$C = B\log_2\left(1+\frac{P}{BS_{\Delta f\Delta f}(\omega_o)}\right)$$

In the limit $P >> B\hbar\omega_o$, the quantum channel result is as if it were a classical AWGN channel with added white noise with a noise power spectral density of $\hbar\omega_o$

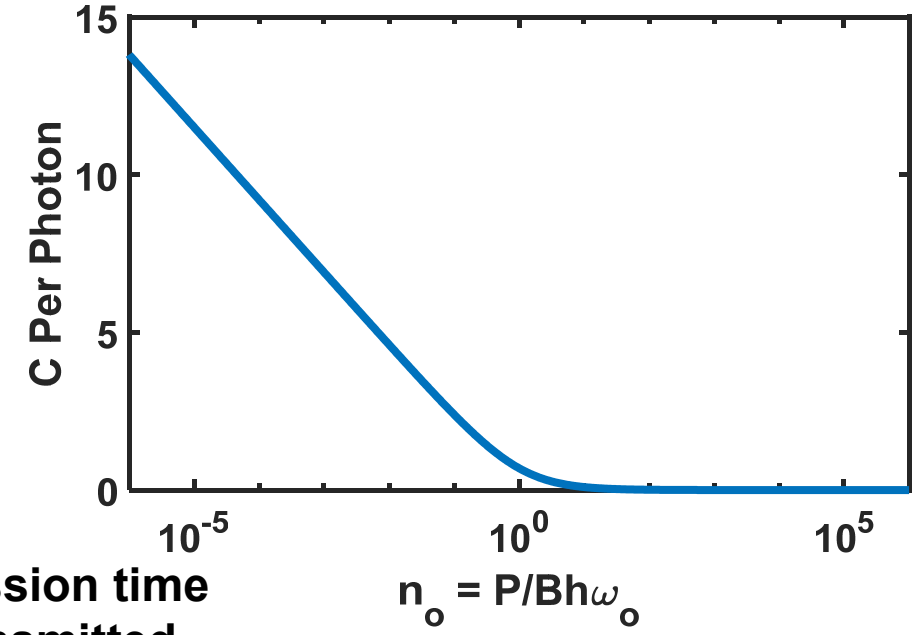
WHY???

Classical Information Over a Photonic Channel: Photon Number States and Photon Number Detection

The Low Power Limit:

In the limit $P << B\hbar\omega_o$ the capacity (bits/s) becomes:

$$C \approx \frac{P}{\hbar\omega_o}\log_2\left(\frac{B\hbar\omega_o}{P}\right)$$



How do we understand the above result?

For small signal powers P, choose a transmission time T long enough such that one photon gets transmitted in time T . Then:

$$P = \frac{\hbar\omega_o}{T}$$

If the channel bandwidth is B then the transmission time T can be divided into BT time slots.

The transmitted photon can occupy any one of these time slots. The information in bits transmitted per second by that one photon, and therefore the channel capacity, becomes:

$$C = \frac{\log_2(BT)}{T} = \frac{P}{\hbar\omega_o}\log_2\left(\frac{B\hbar\omega_o}{P}\right)$$

Classical Information Over a Photonic Channel: Coherent States and Photon Number Detection

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha|$$

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\}$$

$$A = \{|\alpha\rangle\langle\alpha|\}$$

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha|$$

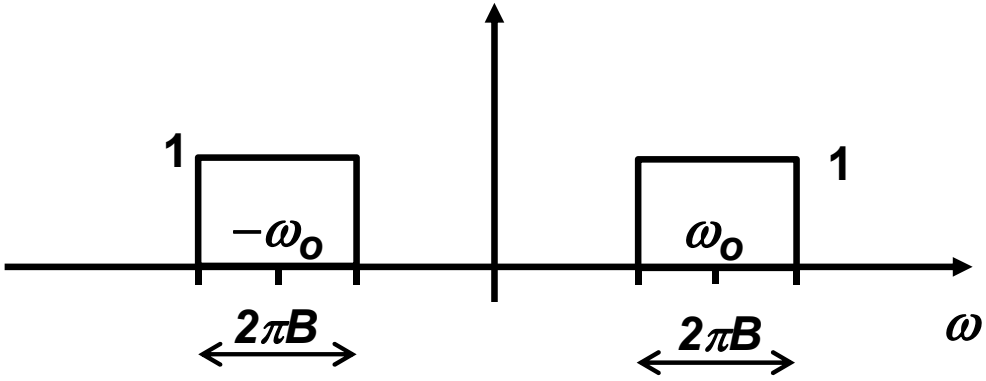
$$\hat{\sigma}^N = \underbrace{\hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes \ldots\ldots\ldots \otimes \hat{\sigma}}_{N \text{ terms}}$$

POVM

$$\hat{F}_n = |n\rangle\langle n|$$

$$\sum_n \hat{F}_n = \hat{1}$$

- The channel is bandlimited
- Optical pulses are used and the classical information is encoded in the amplitude quadrature of each optical pulse



From previous discussion, at max B such pulses can be sent per second

Power Constrain:

$$B\hbar\omega_o \int d^2\alpha \; p_{in}(\alpha) |\alpha|^2 = P$$

Classical Information Over a Photonic Channel: Coherent States and Photon Number Detection

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) \left|\alpha\right\rangle\left\langle\alpha\right| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) \left|\alpha\right\rangle\left\langle\alpha\right|$$

The chosen POVM, given below, for detection is (possibly) not the optimal POVM for the coherent state alphabet (as we will see later)

$$\hat{F}_n = \left|n\right\rangle\left\langle n\right| \qquad \sum_n \hat{F}_n = \hat{1}$$

Since channel capacity definition includes use of the optimal POVM, which we are (possibly) not using, we just calculate the mutual information between channel input and the detector output

The conditional probability of detecting *n* photons, given the input $\left|\alpha\right\rangle$, is:

$$p_{out}(n \mid \alpha) = \text{Tr}\left\{\left|\alpha\right\rangle\left\langle\alpha\right| \hat{F}_n\right\} = e^{-|\alpha|^2} \frac{|\alpha|^{2n}}{n!} \quad \Rightarrow \quad p_{out}(n) = \text{Tr}\left\{\hat{\sigma}\, \hat{F}_n\right\} = \int d^2\alpha \; p_{in}(\alpha)\, e^{-|\alpha|^2} \frac{|\alpha|^{2n}}{n!}$$

The optimal mutual information between the channel input *I* and the detector output *O* is:

$$I(O:I) = \max_{p_{in}(\alpha)} H(O) - H(O \mid I)$$

$$= \max_{p_{in}(\alpha)} - \sum_{n=0}^{\infty} p_{out}(n) \log_2\left[p_{out}(n)\right] + \int d^2\alpha \; p_{in}(\alpha) \sum_{n=0}^{\infty} p_{out}(n \mid \alpha) \log_2\left[p_{out}(n \mid \alpha)\right]$$

Classical Information Over a Photonic Channel: Coherent States and Photon Number Detection

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha|$$

$$I(O:I) = \max_{p_{in}(\alpha)} - \sum_{n=0}^{\infty} p_{out}(n) \log_2 \left[p_{out}(n) \right] + \int d^2\alpha \; p_{in}(\alpha) \sum_{n=0}^{\infty} p_{out}(n|\alpha) \log_2 \left[p_{out}(n|\alpha) \right]$$

The above needs to be maximized over $p(\alpha)$ under the power constrain:

$$B\hbar\omega_o \int d^2\alpha \; p_{in}(\alpha) |\alpha|^2 = P$$

The maximizing procedure turns out to be analytically cumbersome, but results in the low power and high power limits are known

The Low Power Limit: $P << B\hbar\omega_o$

$$I(O:I) \approx \frac{P}{\hbar\omega_o} \log_2 \left(\frac{B\hbar\omega_o}{P} \right)$$

Same as in the case of using photon number states

The High Power Limit: $P >> B\hbar\omega_o$

$$I(O:I) \approx \frac{B}{2} \log_2 \left(\frac{P}{B\hbar\omega_o} \right)$$

One half of the result in the case of using photon number states

WHY?!?

Classical Information Over a Photonic Channel: Coherent States and Balanced Heterodyne Detection

$$M = \{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \ldots\ldots\ldots \hat{\sigma}_N\}$$

$$A = \{|\alpha\rangle\langle\alpha|\}$$

POVM

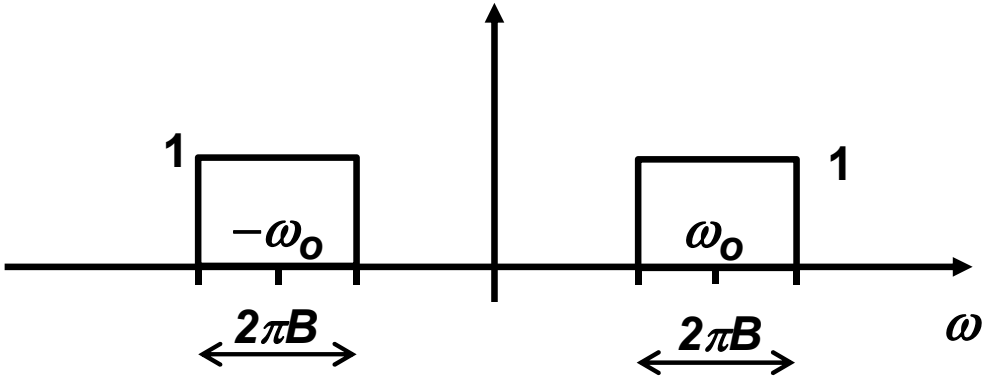
$$\hat{F}_\beta = \frac{1}{\pi}|\beta\rangle\langle\beta|$$

$$\int d^2\beta \;\; \hat{F}_\beta = \hat{1}$$

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha)\,|\alpha\rangle\langle\alpha|$$

$$\hat{\sigma}^N = \underbrace{\hat{\sigma} \otimes \hat{\sigma} \otimes \hat{\sigma} \otimes \ldots\ldots\ldots \otimes \hat{\sigma}}_{N \text{ terms}}$$

- The channel is bandlimited
- Optical pulses are used and the classical information is encoded in the two quadratures of each optical pulse



From previous discussion, at max B such pulses can be sent per second

Power Constrain:

$$B\hbar\omega_o \int d^2\alpha \; p_{in}(\alpha)\,|\alpha|^2 = P$$

Classical Information Over a Photonic Channel: Coherent States and Balanced Heterodyne Detection

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha|$$

POVM:

$$\hat{F}_\beta = \frac{1}{\pi} |\beta\rangle\langle\beta| \qquad \int d^2\beta \;\; \hat{F}_\beta = \hat{1}$$

The chosen POVM implies that both the field quadratures are measured simultaneously!

Since channel capacity definition includes use of the optimal POVM, which we are (possibly) not using, we just calculate the mutual information between channel input and the detector output

The conditional probability of detecting $|\beta\rangle$, given the input $|\alpha\rangle$, is:

$$p_{out}(\beta \mid \alpha) = \text{Tr}\left\{ |\alpha\rangle\langle\alpha| \hat{F}_\beta \right\} = \frac{1}{\pi} |\langle\beta|\alpha\rangle|^2 \Rightarrow p_{out}(\beta) = \text{Tr}\left\{ \hat{\sigma} \, \hat{F}_\beta \right\} = \int d^2\alpha \; p_{in}(\alpha) \; p_{out}(\beta \mid \alpha)$$

The optimal mutual information between the channel input *I* and the heterodyne detector output *O* is:

$$I(O:I) = \max_{p_{in}(\alpha)} H(O) - H(O \mid I)$$

$$= \max_{p_{in}(\alpha)} - \int d^2\beta \; p_{out}(\beta) \log_2 \left[p_{out}(\beta) \right] + \int d^2\alpha \; p_{in}(\alpha) \int d^2\beta \;\; p_{out}(\beta \mid \alpha) \log_2 \left[p_{out}(\beta \mid \alpha) \right]$$

Classical Information Over a Photonic Channel: Coherent States and Balanced Heterodyne Detection

$$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha| \longrightarrow \boxed{\text{Channel}} \longrightarrow \hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha|$$

$$p_{out}(\beta \mid \alpha) = \frac{1}{\pi} \left| \langle \beta | \alpha \rangle \right|^2 = \frac{1}{\pi} e^{-|\alpha - \beta|^2}$$

$$p_{out}(\beta \mid \alpha) = \frac{1}{\sqrt{2\pi(1/2)}} e^{-\frac{(\beta_r - \alpha_r)^2}{2(1/2)}} \frac{1}{\sqrt{2\pi(1/2)}} e^{-\frac{(\beta_i - \alpha_i)^2}{2(1/2)}}$$

Gaussian!!
With added noise having a variance of 1/2 in each quadrature!!

$$p_{out}(\beta) = \int d^2\alpha \; p_{in}(\alpha) \, p_{out}(\beta \mid \alpha)$$

To maximize *I(O:I)*, we need $p_{out}(\beta)$ to be Gaussian as well, and this is possible if $p_{in}(\alpha)$ is Gaussian and satisfies the power constrain:

$$\int d^2\alpha \; p_{in}(\alpha) |\alpha|^2 = \frac{P}{B\hbar\omega_o}$$

The maximizing yields (just like in the case of AWGN) the capacity per letter (consisting of two quadratures):

$$C = 2x \frac{1}{2} \log_2 \left(\frac{P/2B\hbar\omega_o + 1/2}{1/2} \right) = \log_2 \left(1 + \frac{P}{B\hbar\omega_o} \right)$$

Classical Information Over a Photonic Channel: Coherent States and Balanced Heterodyne Detection

$\hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha| \longrightarrow$ **Channel** $\longrightarrow \hat{\sigma} = \int d^2\alpha \; p_{in}(\alpha) |\alpha\rangle\langle\alpha|$

The max information per letter (consisting of two quadratures) is:

$$C = \log_2\left(1 + \frac{P}{B\hbar\omega_o}\right)$$

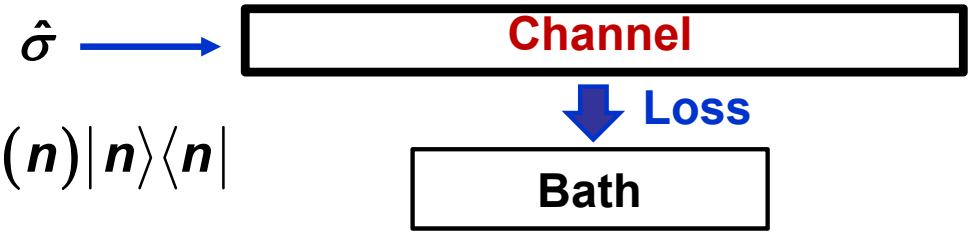
And since one can send *B* letters per second, the capacity in bits/s is:

$$C = B\log_2\left(1 + \frac{P}{B\hbar\omega_o}\right)$$

The result, although identical to the one obtained using photon number states and photon number detection (in the high power limit), has more similarities with the classical AWGN result if:

1) each quadrature of the input coherent state is assumed to be a classical variable and,
2) the channel adds white Gaussian noise to each quadrature and,
3) the power spectral density of the added white Gaussian noise is assumed to be 1/2($\hbar\omega_o$)

Classical Information Over a Lossy Photonic Channel: Number States and Photon Number Detection



$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n|$$

$$E(\hat{\sigma}) = \sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n|$$

Input Power Constrain:

$$\sum_{n=0}^{\infty} p_{in}(n)\, nB\hbar\omega_o = P_{in} \quad \Rightarrow \quad \sum_{n=0}^{\infty} p_{in}(n)\, n = n_{in} = \frac{P_{in}}{B\hbar\omega_o}$$

POVM

$$\hat{F}_n = |n\rangle\langle n|$$

How do we model photon loss in the channel?

$$\hat{\sigma} \rightarrow E(\hat{\sigma}) = \sum_j \hat{E}_j \hat{\sigma} \hat{E}_j^+$$

Channel Power Transmissivity: T **Channel Power Loss:** $1\text{-}T$

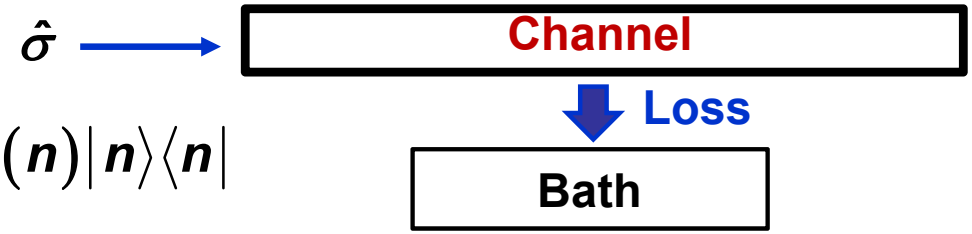
We know how a photon number state behaves in the presence of loss

$$|n\rangle \otimes |0\rangle_B \rightarrow \sum_{m=0}^{n} \sqrt{\frac{n!}{m!(n-m)!}} \left(\sqrt{T}\right)^m \left(\sqrt{1-T}\right)^{n-m} |m\rangle |n-m\rangle_B$$

Binomial distribution of photons

$$|n\rangle\langle n| \rightarrow E(|n\rangle\langle n|) = \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} T^m (1-T)^{n-m} |m\rangle\langle m|$$

Classical Information Over a Lossy Photonic Channel: Number States and Photon Number Detection



$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n|$$

$$E(\hat{\sigma}) = \sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n|$$

The conditional probability of detecting *m* photons at the output, given the input $|n\rangle$, is:

$$p_{out}(m \mid n) = \frac{n!}{m!(n-m)!} T^m (1-T)^{n-m} \qquad (m \le n, \ \ 0 \quad \text{otherwise})$$

$$\Rightarrow p_{out}(m) = \sum_{n=0}^{\infty} p_{out}(m \mid n)\, p_{in}(n)$$

Output photon number and power:

$$\Rightarrow \sum_{n=0}^{\infty} p_{out}(n)\, n = n_{out} = n_{in} T = \frac{P_{in} T}{B\hbar\omega_o} = \frac{P_{out}}{B\hbar\omega_o}$$

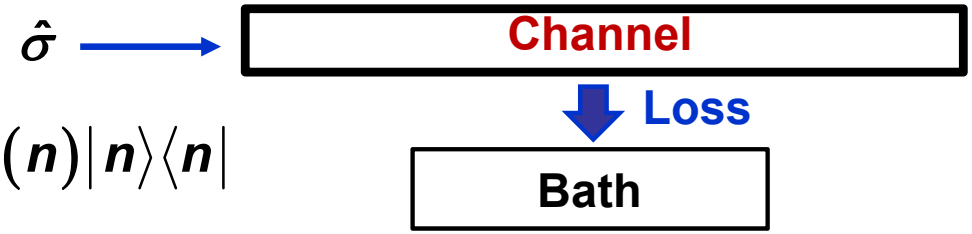
Input and output states:

$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n| \rightarrow E\left(\sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n| \right)$$

$$= \sum_{n=0}^{\infty} p_{in}(n) \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} T^m (1-T)^{n-m} |m\rangle\langle m|$$

$$= \sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n|$$

Classical Information Over a Lossy Photonic Channel: Number States and Photon Number Detection



$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n|$$

$$E(\hat{\sigma}) = \sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n|$$

The channel capacity is (the POVM is optimal):

$$C = \max_{p_i} S\left(E(\hat{\sigma})\right) - \sum_{i=1} p_i S\left(E(\hat{\rho}_i)\right)$$

$$= \max_{p_{in}(n)} S\left(\sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n| \right) - \sum_{n=0}^{\infty} p_{in}(n) S\left(\sum_{m=0}^{n} p_{out}(m \mid n) |m\rangle\langle m| \right)$$

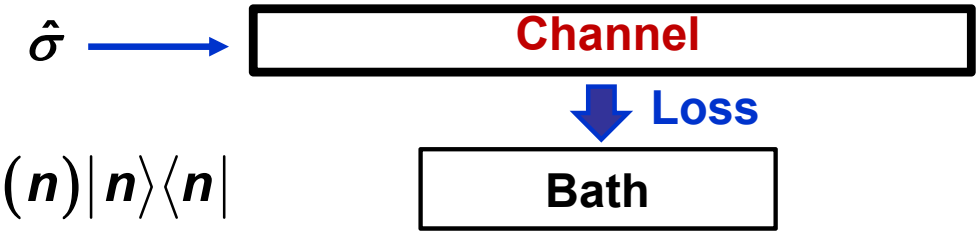
It is not difficult to evaluate:

$$S\left(\sum_{m=0}^{n} p_{out}(m \mid n) |m\rangle\langle m| \right) = -\sum_{m=0}^{n} p_{out}(m \mid n) \log_2 \left[p(m \mid n) \right] \approx \frac{1}{2} \log_2 \left[1 + 2\pi e n T (1-T) \right]$$

The channel capacity becomes:

$$C = \max_{p_{in}(n)} - \sum_{n=0}^{\infty} p_{out}(n) \log_2 \left[p_{out}(n) \right] - \sum_{n=0}^{\infty} p_{in}(n) \frac{1}{2} \log_2 \left[1 + 2\pi e n T (1-T) \right]$$

Classical Information Over a Lossy Photonic Channel: Number States and Photon Number Detection



$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n|$$

$$E(\hat{\sigma}) = \sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n|$$

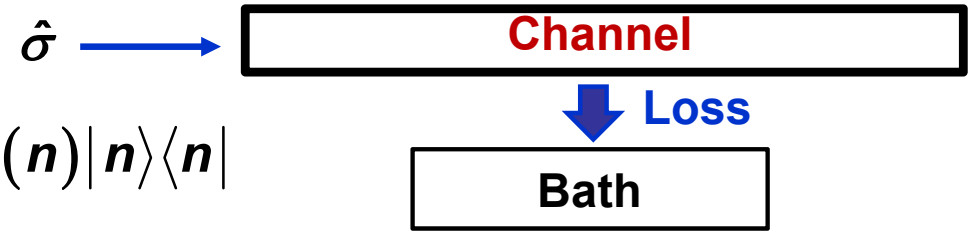
$$C = \max_{p_{in}(n)} - \sum_{n=0}^{\infty} p_{out}(n) \log_2 \left[p_{out}(n) \right] - \sum_{n=0}^{\infty} p_{in}(n) \frac{1}{2} \log_2 \left[1 + 2\pi e n T (1-T) \right]$$

We want to maximize the capacity ⟹ we want output distribution to be thermal

The output distribution can be thermal if the input distribution is also thermal

$$p_{in}(n) = \frac{1}{1+n_{in}} \left(\frac{n_{in}}{1+n_{in}} \right)^n \Longrightarrow p_{out}(n) = \frac{1}{1+n_{in}T} \left(\frac{n_{in}T}{1+n_{in}T} \right)^n$$

Classical Information Over a Lossy Photonic Channel: Number States and Photon Number Detection



$$\hat{\sigma} = \sum_{n=0}^{\infty} p_{in}(n) |n\rangle\langle n|$$

$$E(\hat{\sigma}) = \sum_{n=0}^{\infty} p_{out}(n) |n\rangle\langle n|$$

The optimal mutual information (bits per use) between the input *I* and the output *O* is:

$$I(O:I) \approx \max_{p_{in}(n)} - \sum_{n=0}^{\infty} p_{out}(n) \log_2 \left[p_{out}(n) \right] - \frac{1}{2} \log_2 \left[1 + \mu \pi e n_{in} T (1-T) \right]$$

$$= \log_2 (1 + n_{in} T) + n_{in} T \log_2 \left(1 + \frac{1}{n_{in} T} \right) - \frac{1}{2} \log_2 \left[1 + \mu \pi e n_{in} T (1-T) \right]$$

$$= \log_2 \left(1 + \frac{P_{out}}{B \hbar \omega_o} \right) + \frac{P_{out}}{B \hbar \omega_o} \log_2 \left(1 + \frac{B \hbar \omega_o}{P_{out}} \right) - \frac{1}{2} \log_2 \left(1 + \mu \pi e \frac{P_{out}}{B \hbar \omega_o} (1-T) \right)$$

Here, μ is a number with values between 1 and 2 and depends on the value of T

The Low Power Limit: $P_{out} << B\hbar\omega_o$

$$I(O:I) \approx \frac{P_{out}}{\hbar \omega_o} \log_2 \left(\frac{B \hbar \omega_o}{P_{out}} \right)$$

(Now bits/s)

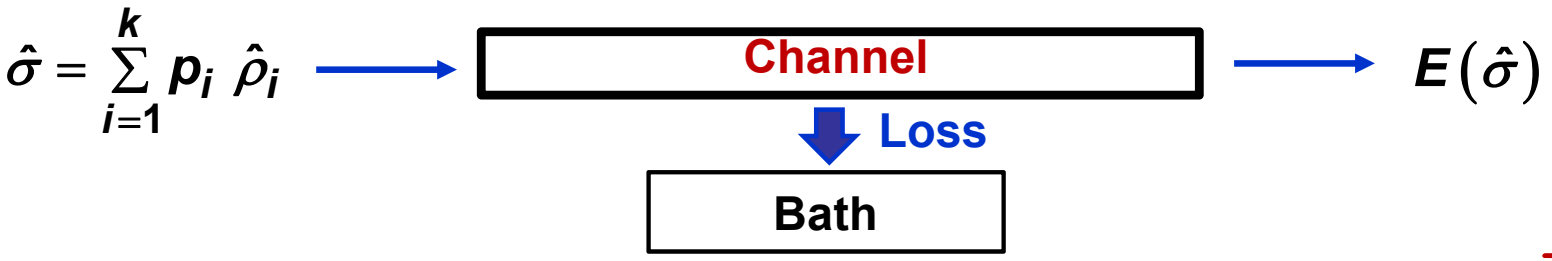
The High Power Limit: $P_{out} >> B\hbar\omega_o$ ($T << 1$)

$$I(O:I) \approx \frac{B}{2} \log_2 \left(\frac{P_{out}}{B \hbar \omega_o} \right)$$

(Now bits/s)

WHY?!?

Classical Information Over a Lossy Photonic Channel: Capacity



POVM

$\hat{F}_j = ?$

The capacity of this quantum channel per letter is:

$$C = \max_{p_i} S\left(E\left(\hat{\sigma}\right)\right) - \sum_{i=1}^{k} p_i S\left(E\left(\hat{\rho}_i\right)\right)$$

Channel Power Transmissivity: T

Channel Power Loss: $1\text{-}T$

Input Power Constrain:

$$\mathrm{Tr}\left\{\hat{\sigma}\,\hat{n}\right\} B\hbar\omega_o = P_{in} \quad \Rightarrow \quad P_{out} = P_{in}T$$

$$C = \log_2\left(1+\frac{P_{out}}{B\hbar\omega_o}\right) + \frac{P_{out}}{B\hbar\omega_o}\log_2\left(1+\frac{B\hbar\omega_o}{P_{out}}\right)$$

Classical Information Over a Lossy Photonic Channel: Capacity

VOLUME 92, NUMBER 2 PHYSICAL REVIEW LETTERS week ending 16 JANUARY 2004

Classical Capacity of the Lossy Bosonic Channel: The Exact Solution

V. Giovannetti,[1] S. Guha,[1] S. Lloyd,[1,2] L. Maccone,[1] J. H. Shapiro,[1] and H. P. Yuen[3]

[1]*Massachusetts Institute of Technology–Research Laboratory of Electronics, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA*

[2]*Massachusetts Institute of Technology–Department of Mechanical Engineering, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA*

[3]*Northwestern University–Department of Electrical and Computer Engineering, 2145 North Sheridan Road, Evanston, Illinois 60208-3118, USA*

(Received 6 August 2003; published 15 January 2004)

The classical capacity of the lossy bosonic channel is calculated exactly. It is shown that its Holevo information is not superadditive, and that a coherent-state encoding achieves capacity. The capacity of far-field, free-space optical communications is given as an example.

Quantum Information: Von Neumann Entropy

The "information" content of a quantum state is related to the Von Neumann entropy:

$$S(\hat{\rho}) = -\mathrm{Tr}\left[\hat{\rho}\log_2(\hat{\rho})\right]$$

The Von Neumann entropy plays three roles (that we know of so far):

1) **It *quantifies* the quantum information content in qubits of a quantum state (i.e. the minimum number of qubits needed to reliably encode the quantum state)**

2) **It also *quantifies* the classical information in bits that can be gained about the quantum state by making the best possible measurement**

3) **It also quantifies the amount of entanglement in bipartite states**

As you will see, the Von Neumann entropy will not always give the answer to the question we will ask!

Quantifying Entanglement of Bipartite Pure States

Consider the following two states of two 2-level systems:

$$|\phi_a\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right]$$

$$|\phi_b\rangle = \frac{1}{2}\left[\sqrt{3}|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right]$$

They are both entangled

But which one is more entangled??

How can we quantify the level of entanglement of states?

The answer for at least pure states of bipartite systems seems to be available

Entanglement as a Resource

$$|\phi_a\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right]$$

Alice

Bob

Entanglement between qubits possessed by Alice and Bob cannot be generated by any local operations or measurements or performed by Alice or Bob on their respective qubit or by classical communications between Alice and Bob (LOCC)

Entanglement can only be generated by a joint operation on both the qubits

Entanglement is a resource

Bell States:

$$|\phi_a\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right] \qquad |\phi_c\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right]$$

$$|\phi_b\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B\right] \qquad |\phi_d\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|0\rangle_B - |1\rangle_A|1\rangle_B\right]$$

Quantifying Entanglement of Bipartite Pure States

Suppose Alice and Bob would like to prepare *n* copies of an entangled state:

$$|\psi\rangle = \alpha|0\rangle_A|0\rangle_B + \beta|0\rangle_A|1\rangle_B + \gamma|1\rangle_A|0\rangle_B + \delta|1\rangle_A|1\rangle_B$$

Alice

Bob

But what they already have in their possession are multiple copies of a Bell state (doesn't matter which one)

Suppose Alice and Bob use a minimum of k_{min} Bell states in their possession, and lots of local operations on their respective qubits and classical communication between each other (LOCC), and are able to generate *n* copies of the desired state $|\psi\rangle$

Then can we use the ratio k_{min}/n as a measure of entanglement in the state $|\psi\rangle$??

i.e. how many Bell states does one need to use to generate one copy?

Quantifying Entanglement of Bipartite Pure States

Suppose Alice and Bob have *n* copies of an entangled state:

$$|\psi\rangle = \alpha|0\rangle_A|0\rangle_B + \beta|0\rangle_A|1\rangle_B + \gamma|1\rangle_A|0\rangle_B + \delta|1\rangle_A|1\rangle_B$$

Alice

Bob

But what they want are multiple copies of a Bell state (doesn't matter which one)

Suppose Alice and Bob are able to prepare a maximum of k_{max} Bell states from the *n* copies of the state $|\psi\rangle$ in their possession, with only local operations on their respective qubits and classical communication between each other (LOCC)

Then can we use the ratio k_{max}/n as a measure of entanglement in the state $|\psi\rangle$??

i.e. how many Bell states does one generate per one copy?

Quantifying Entanglement of Bipartite Pure States

Alice

Bob

$$|\psi\rangle = \alpha|0\rangle_A|0\rangle_B + \beta|0\rangle_A|1\rangle_B + \gamma|1\rangle_A|0\rangle_B + \delta|1\rangle_A|1\rangle_B$$

It can be shown that in the limit $n \to \infty$,

$$\lim_{n\to\infty} \frac{k_{max}}{n} = \frac{k_{min}}{n} = S(\hat{\rho}_A) = S(\hat{\rho}_B) = E(|\psi\rangle)$$

This many Bell states go into or come out of the above state

Where:

$$\hat{\rho}_A = \mathrm{Tr}_B\{|\psi\rangle\langle\psi|\}$$

$$\hat{\rho}_B = \mathrm{Tr}_A\{|\psi\rangle\langle\psi|\}$$

The above expression for bipartite entanglement works even when the qubits involved are not 2-level systems but any arbitrary multilevel systems

Quantifying Entanglement of Bipartite Pure States

Consider the following two states of two 2-level systems:

$$|\phi_a\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right]$$

$$|\phi_b\rangle = \frac{1}{2}\left[\sqrt{3}|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right]$$

They are both entangled

But which one is more entangled??

Answer:

$$E\left(|\phi_a\rangle\right) = 1.0$$

$$E\left(|\phi_b\rangle\right) \approx 0.81$$

All four Bell states are maximally entangled

Quantifying Entanglement of Bipartite Mixed States

Alice

Bob

What is Alice and Bob share a mixed entangled state?

$$\hat{\rho}_{AB}$$

What is the entanglement of this state?

How many Bell states can Alice and Bob distill from $\hat{\rho}_{AB}$?

How many Bell states are needed to prepare $\hat{\rho}_{AB}$?

We don't know the general answers to the above questions!!!

The Last Slide

That's All Folks