Classical and Quantum Information Theory

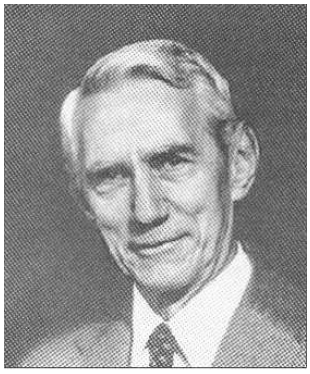
from Discrete to Continuous Variables

Nicolas J. Cerf

Centre for Quantum Information and Communication (QuIC), Ecole Polytechnique, Université Libre de Bruxelles, Belgium

QUANTUM INFORMATION, COMPUTATION, AND COMPLEXITY
Institut Henri Poincaré (IHP)
Paris, January 4 -- April 7, 2006

Shannon's information theory



Claude E. Shannon
1916–2001

A mathematical theory of communication, Bell System Technical Journal **27** (1948) 379-423 and 623-656.

Communication theory of secrecy systems, Bell System Technical Journal **28** (1949) 656-715.

This work has laid out the entire foundation of today's information technology era.

1. Every kind of information – text, image, sound – can be associated with an information content, which quantifies how efficiently it can be *represented* with 0's and 1's.

2. Any *imperfect* communication channel – telephone line, radio channel, satellite link – has a capacity, which quantifies how much information can be transmitted *reliably* over the channel.

Source coding theorem
(noiseless coding)

Q: What is the highest possible data compression rate?

A: Shannon's entropy $H = f(\text{source statistics})$.

0 1 1 0 1 0 0 1 1 1 0 1 0 0 1 $\longrightarrow$ 0 1 0 1 0 1 1 0 0 1

... suppressing redundancy: "source code"

Channel coding theorem
(noisy channel coding)

Q: What is the highest possible transmission rate?

A: Shannon's capacity $C = f(\text{noise statistics})$.

0 1 0 1 0 1 1 0 0 1 $\longrightarrow$ 0 1 1 0 1 0 0 1 1 1 0 1 0 0 1

... adding redundancy: "error correcting code"

Example of source coding

Source: random variable X distributed as $p(x)$

Entropy: $H(X) \equiv -\sum_{x=1}^{S} p(x) \log_2 p(x)$

where S is the alphabet size

$$p(x) = \begin{cases} 1/2 & \text{if } x = \text{"A"} \\ 1/4 & \text{if } x = \text{"B"} \\ 1/8 & \text{if } x = \text{"C"} \\ 1/8 & \text{if } x = \text{"D"} \end{cases}$$

Simple code:

"A" $\rightarrow$ 00
"B" $\rightarrow$ 01
"C" $\rightarrow$ 10
"D" $\rightarrow$ 11

$L = 2$ bits/symbol

Average code length $L \geq H(X)$

$H(X) = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} = \frac{7}{4}$ bits < 2 bits

Clever code:

"A" $\rightarrow$ 0
"B" $\rightarrow$ 10
"C" $\rightarrow$ 110
"D" $\rightarrow$ 111

$L = \frac{7}{4}$ bits/symbol

Typical sequences

Entropy $H(X)$ is thus the average length of the shortest description of X. But can we always reach this irreducible compression?

YES, using "typical" sequences of size $n \to \infty$.

Sequence $(X_1, \ldots X_n)$ of independent random variables distributed each as $p(x)$

$$-\tfrac{1}{n}\log_2 p(x_1, \ldots x_n) = -\tfrac{1}{n}\log_2 \prod_{i=1}^{n} p(x_i) = -\tfrac{1}{n}\sum_{i=1}^{n} \log_2 p(x_i)$$
$$\simeq \langle -\log_2 p(X)\rangle_{X\sim p(x)} \equiv H(X)$$

$\Rightarrow p(x_1, \ldots x_n) \simeq 2^{-nH(X)}$ for typical sequences.

Normalization: $\sum_{\bar{x}\in S_{\text{typ}}} p(\bar{x}) \simeq 1 \;\Rightarrow |S_{\text{typ}}| \simeq 2^{nH(X)}$

About $nH(X)$ bits are sufficient, on average, to code a typical n-symbol sequence, so that $L \simeq H(X)$ bits/symbol.

Example: binary source

$p(0) = 1 - p = 3/4$
$p(1) = p = 1/4$

$$\underbrace{00001010000100110000}_{n=20 \text{ bits}}$$

Typical n-bit sequence contains
$\sim n(1-p) = 15\times$ 0's
$\sim np = 5\times$ 1's.

Probability of emission

$p(x_1, \ldots x_n) = (1-p)^{\text{nr. of 0's}}\ p^{\text{nr. of 1's}}$

$\simeq (1-p)^{n(1-p)} p^{np} = 2^{n(1-p)\log_2(1-p) + np\log_2 p} \equiv 2^{-nH(X)}$

Number of typical sequences

$$|S_{\text{typ}}| \simeq \binom{n}{np} = \frac{n!}{(np)!(n(1-p))!} \qquad \text{use } n! \simeq n^n = 2^{n\log_2 n}$$

$$\simeq 2^{n\log_2 n - np\overbrace{\log_2(np)}^{\log_2 n + \log_2 p} - n(1-p)\overbrace{\log_2(n(1-p))}^{\log_2 n + \log_2(1-p)}} \equiv 2^{nH(X)}$$

Probability of emitting any typical sequences
$\sim 2^{-nH(X)} \times 2^{nH(X)} \sim 1$

Note: most frequent sequence 00000000000000000000 is generally not typical !

Remarkably, S_{typ} is only an exponentially small part of the set of all sequences, while it has arbitrarily large probability.

Since $\quad 0 \leq H(X) \leq \log_2 S$

generally $\quad |S_{\text{typ}}| \simeq 2^{nH(X)} \ll 2^{n \log_2 S} = S^n$

• Typical set is sufficiently large to contain most of the probability

$\forall \epsilon > 0, \delta > 0; \exists n$ and S_{typ} :

$|S_{\text{typ}}| < 2^{n[H(X)+\delta]}$ and $\sum_{\bar{x} \in S_{\text{typ}}} p(\bar{x}) > 1 - \epsilon$

• Typical set is the smallest set that contains most of the probability

$\forall \epsilon > 0, \delta > 0; \forall n$:

$|S| < 2^{n[H(X)-\delta]} \quad \Rightarrow \quad \sum_{\bar{x} \in S} p(\bar{x}) < \epsilon$

Instantaneous (variable-length) codes

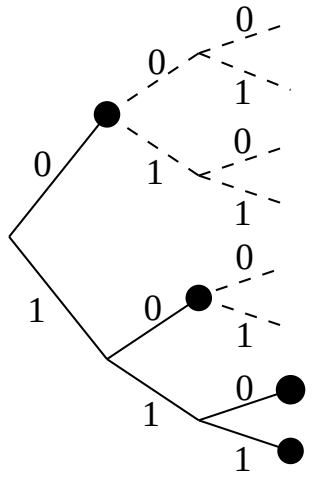
Code C: symbol $x \rightarrow$ codeword $C(x)$ of length $l(x)$

C is instantaneous code if no $C(x)$ is a prefix of another $C(x')$ with $x' \neq x$

"A" $\rightarrow 0$
"B" $\rightarrow 10$
"C" $\rightarrow 110$
"D" $\rightarrow 111$

$10, 0, 0, 110, 0, 111, 10, \ldots$

Kraft inequality $\sum_{x=1}^{S} 2^{-l(x)} \leq 1$



"A" uses $2^{-1} = 1/2$ of the decoding tree
"B" $2^{-2} = 1/4$
"C" $2^{-3} = 1/8$
"D" $2^{-3} = 1/8$

Necessary and sufficient condition for the existence of instantaneous code

Optimal instantaneous code

$$\min_{l(x)} \sum_{x=1}^{S} p(x) l(x) \quad \text{with constraint} \quad \sum_{x=1}^{S} 2^{-l(x)} = 1$$

using Lagrange multiplier

$$\mathcal{L} = \sum_x p(x) l(x) + \lambda \sum_x 2^{-l(x)}$$

$$\forall x, \ \frac{\partial \mathcal{L}}{\partial l(x)} = 0 \ \Rightarrow \ p(x) - \lambda 2^{-l(x)} \ln 2 = 0 \ \Rightarrow \ \overbrace{\sum 2^{-l(x)}}^{\sum = 1} = \overbrace{\frac{\sum p(x)}{\lambda \ln 2}}^{\sum = 1}$$

$\Rightarrow l_{\text{opt}}(x) = -\log_2 p(x)$ (assuming it is an integer)

$\Rightarrow L_{\text{opt}} = \sum_x p(x) l_{\text{opt}}(x) = -\sum_x p(x) \log_2 p(x) \equiv H(X)$

Bloc coding: typical size-n sequences become almost equiprobable $p(x_1, \ldots x_n) \simeq 2^{-nH(X)}$

$l_{\text{opt}}(x_1, \ldots x_n) \simeq nH(X) \simeq$ integer

$L_{\text{opt}} \simeq nH(X)$ bits/bloc

No further compression?

Relative entropy (Kullback "distance")

$$D(p||q) = \sum_{x=1}^{S} p(x) \log_2 \frac{p(x)}{q(x)}$$

with $p(x), q(x)$ probability distributions.

$$\begin{aligned} D(p||q) &= \langle -\log_2 \frac{q(x)}{p(x)} \rangle_{X\sim p(x)} \\ &\geq -\log_2 \langle \frac{q(x)}{p(x)} \rangle_{X\sim p(x)} && \text{concavity of log} \\ &= -\log_2 \textstyle\sum_x p(x) \frac{q(x)}{p(x)} \\ &= -\log_2 \textstyle\sum_x q(x) && = 0 \end{aligned}$$

$D(p||q) \geq 0$

$D(p||q) = 0$ iff $\forall x,\ p(x) = q(x)$

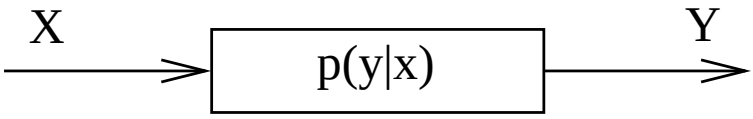
$$
\begin{aligned}
L - H(X) &= \textstyle\sum_x p(x) l(x) + \sum_x p(x) \log_2 p(x) \\
&= \textstyle\sum_x p(x) \log_2 \left[2^{l(x)} p(x) \right] \\
&= \textstyle\sum_x p(x) \log_2 \left[\dfrac{\sum_y 2^{-l(y)}}{2^{-l(x)}} \times \dfrac{p(x)}{\sum_y 2^{-l(y)}} \right]
\end{aligned}
$$

Let $q(x) = \dfrac{2^{-l(x)}}{\sum_y 2^{-l(y)}}$ normalized probability distribution

$$
\begin{aligned}
L - H(X) &= \textstyle\sum_x p(x) \log_2 \left[\dfrac{p(x)}{q(x)} \times \dfrac{1}{\sum_y 2^{-l(y)}} \right] \\
&= \underbrace{D(p||q)}_{\geq 0} - \log_2 \underbrace{\sum_y 2^{-l(y)}}_{\leq 1 \text{ Kraft}} \\
&\geq 0
\end{aligned}
$$

The average length L of any instantaneous code cannot be lower than $H(X)$

Noisy channel coding



$X =$ input symbol
$Y =$ output symbol

$p(y|x) =$ transition probability

Naive picture:
to reduce the error probability P_e
- either increase signal power S
- or decrease noise power N

Shannon theory:
as soon as the rate $R \leq C$, an arbitrarily low P_e is achievable in the limit of large blocs $n \to \infty$.

capacity $C = B \log_2(1 + S/N)$ ($B =$ bandwidth)

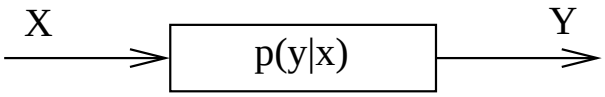
Conditional and mutual entropy

$H(X) \equiv -\sum_x p(x) \log_2 p(x) =$ uncertainty of X

$H(X|Y) \equiv \sum_y p(y) \underbrace{H(X|Y=y)}_{-\sum_x p(x|y) \log_2 p(x|y)} =$ unc. of X if Y known

$$\Rightarrow H(X|Y) = -\sum_{x,y} p(y)p(x|y) \log_2 \frac{p(x,y)}{p(y)}$$

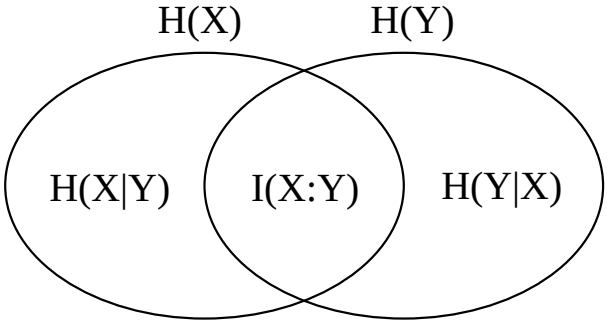
$$= \underbrace{-\sum_{x,y} p(x,y) \log_2 p(x,y)}_{H(X,Y)} + \underbrace{\sum_y p(y) \log_2 p(y)}_{-H(Y)}$$



$H(X|Y) = H(X,Y) - H(Y) =$ loss of information

perfect channel $H(X|Y) = 0$
random channel $H(X|Y) = H(X)$

$I(X{:}Y) \equiv H(X) - H(X|Y) =$ mutual information
$= H(X) + H(Y) - H(X,Y)$



$I(X{:}Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$
$= D\big(p(x,y)||p(x)p(y)\big) \geq 0$

$$I(X{:}Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$
$$= D\Big(p(x,y) || p(x)p(y)\Big) \geq 0$$

X and Y independent $\Leftrightarrow I(X{:}Y) = 0$

$I(X{:}Y)$ measures the statistical dependence between X and Y

Subadditivity of entropies

$$I(X{:}Y) = H(X) + H(Y) - H(X,Y) \geq 0$$

thus $\boxed{H(X,Y) \leq H(X) + H(Y)}$

Strong subadditivity of entropies

$$I(X{:}Y|Z) = \sum_z p(z) \underbrace{I(X{:}Y|Z=z)}_{\sum_{x,y} p(x,y|z) \log_2 \frac{p(x,y|z)}{p(x|z)p(y|z)}} \geq 0$$
$$= \sum_{x,y,z} p(x,y,z) \log_2 \frac{p(x,y,z)p(z)}{p(x,z)p(y,z)}$$
$$= H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)$$

thus $\boxed{H(X,Y,Z) + H(Z) \leq H(X,Z) + H(Y,Z)}$

Interpretation of strong subadditivity

$$\begin{array}{rcl} H(X,Y,Z) + H(Z) & \leq & H(X,Z) + H(Y,Z) \\ +H(X) & & +H(X) \end{array}$$

$$\begin{aligned} & H(X) + H(Z) - H(X,Z) \\ & \leq \quad H(X) + H(Y,Z) - H(X,Y,Z) \end{aligned}$$

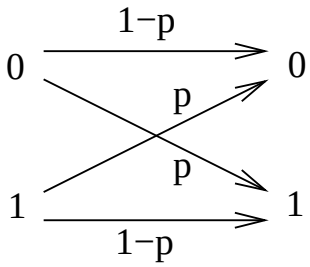
$$\Rightarrow I(X{:}Z) \quad \leq \quad I(X{:}Y,Z)$$

Discarding the variable Y can only decrease the mutual information with X

Shannon capacity

$C = \max_{p(x)} I(X{:}Y)$ for given channel $p(y|x)$

Binary symmetric channel



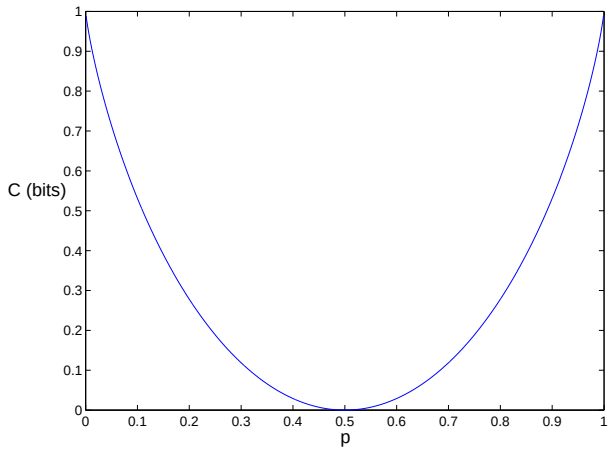
p=bit-flip probability

$$
\begin{aligned}
I(X{:}Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \textstyle\sum_x p(x) H(Y|X=x) \\
&= H(Y) - \textstyle\sum_x p(x) H_2[p] \\
&= H(Y) - H_2[p]
\end{aligned}
$$

with $H_2[p] = -(1-p)\log_2(1-p) - p\log_2 p$

$p(x) = \{1/2, 1/2\} \Rightarrow p(y) = \{1/2, 1/2\} \Rightarrow H(Y) = 1$

$C = 1 - H_2[p]$ bits per use of the channel



Quantum information theory

What if information is encoded into quantum states ?

Bits $b \in \{0,1\} \longrightarrow$ Quantum bits $|\psi\rangle \in \{|0\rangle, |1\rangle\}$

Quantum state in a 2-d Hilbert space, regardless the physical support (e.g., spin 1/2, photon polarization)

Unique properties (compared to classical info):

- Superposition principle:

$\alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$

Computational basis $\{|0\rangle, |1\rangle\}$ (convention)
Dual basis $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$

- Non-orthogonal states are not perfectly distinguishable

$\Rightarrow$ quantum data compression, accessible information, no-cloning principle

- Non-classical correlations (entanglement)

$$|\text{EPR}\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$$
$$= \frac{1}{\sqrt{2}}(|+\rangle|+\rangle + |-\rangle|-\rangle)$$

Distinguishing quantum states

Alice $\longrightarrow \begin{cases} |\psi_0\rangle & \text{prob. } (1-p) \\ |\psi_1\rangle & \text{prob. } p \end{cases} \longrightarrow$ Bob

Source entropy $H_2[p] = -p\log_2 p - (1-p)\log_2(1-p)$

If $\langle\psi_0|\psi_1\rangle \neq 0$, then states cannot be reliably distinguished, unlike classical bits !

Density matrix $\rho = (1-p)|\psi_0\rangle\langle\psi_0| + p|\psi_1\rangle\langle\psi_1|$

von Neumann entropy $S(\rho) = -\mathrm{Tr}(\rho\log_2\rho)$

Diagonalization: $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$

$$S(\rho) = -\sum_i \lambda_i \log_2 \lambda_i \equiv H[\lambda_i]$$

$\boxed{S(\rho) \leq H_2[p]}$ (equality iff $\langle\psi_0|\psi_1\rangle = 0$)

1) "Quantum redundancy" $\Rightarrow$ compression is possible even if $p = 1/2$.

2) Bob maximizes $I(X{:}Y)$ but cannot get all of $H_2[p] \Rightarrow$ limited accessible information

Quantum redundancy

Source $\{|\psi_i\rangle, p_i\}$
resulting mixture $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$.

Preparer P $\longrightarrow i \longrightarrow$ Source Q $\Rightarrow |\psi_i\rangle$

Assume that the preparer is in an entangled state with the source:

$$|PQ\rangle = \sum_i \sqrt{p_i}\ |i\rangle_P\ |\psi_i\rangle_Q$$

with $|i\rangle_P$ being orthonormal so that $\rho_Q = \sum_i\ p_i\ |\psi_i\rangle\langle\psi_i|$

- Before measurement of P:
  $S(P) = S(Q) \equiv S(\rho)$
- After measurement of P:
  if P measures i then Q projected onto $|\psi_i\rangle$
  $\rho'_P = \sum_i\ p_i\ |i\rangle\langle i| \Rightarrow S(P') = H[P_i]$
  $\rho'_Q = \rho_Q \equiv \rho$ (unchanged)

A measurement can only increase the entropy

$$S(P) \leq S(P') \Rightarrow S(\rho) \leq H[p_i]$$

Quantum source coding theorem

(quantum noiseless coding)

B. Schumacher, Phys. Rev. A **51** (1995) 2738

Source $\{|\psi_i\rangle, p_i\}$ with $|\psi_i\rangle \in \mathcal{H}$ and $i = 1, \ldots m$

$$\rightarrow \rho = \sum_{i=1}^{m} p_i \, \psi_i \quad \text{with } \psi_i = |\psi_i\rangle\langle\psi_i|$$

Information-theoretic meaning of $S(\rho)$?

$$\underbrace{|\psi\rangle|\psi\rangle \cdots |\psi\rangle}_{n \text{ states}} \rightarrow \boxed{\text{CODING}} \rightarrow W$$

fidelity = average of $\langle\psi|\langle\psi| \cdots \langle\psi|W|\psi\rangle|\psi\rangle \cdots |\psi\rangle$

For arbitrarily small ϵ and δ,

There exists a coding scheme applied to blocks of n signal states with sufficiently large n

that uses $S(\rho) + \delta$ qubits per signal state and gives a fidelity $F > 1 - \epsilon$

Converse: If only $S(\rho) - \delta$ qubits are available per signal state, then blocks of n signal states will be decoded with fidelity $F < \epsilon$.

Example of quantum coding

$$\boxed{\text{Alice}} \longrightarrow \begin{cases} |\psi_0\rangle = \sqrt{0.9}\,|e_0\rangle + \sqrt{0.1}\,|e_1\rangle & \text{prob. } 1/2 \\ |\psi_1\rangle = \sqrt{0.9}\,|e_0\rangle - \sqrt{0.1}\,|e_1\rangle & \text{prob. } 1/2 \end{cases}$$

$$\begin{aligned} \rho &= \frac{1}{2}\begin{pmatrix} 0.9 & \sqrt{0.09} \\ \sqrt{0.09} & 0.1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0.9 & -\sqrt{0.09} \\ -\sqrt{0.09} & 0.1 \end{pmatrix} \\ &= \begin{pmatrix} 0.9 & 0 \\ 0 & 0.1 \end{pmatrix} \qquad \text{diagonal in eigenbasis } \{e_0, e_1\} \end{aligned}$$

$$\begin{aligned} S(\rho) &= -0.9\,\log_2 0.9 - 0.1\,\log_2 0.1 \\ &= 0.47 \text{ qubits per signal state} \end{aligned}$$

Coding individual states (1 qubit per signal state)

$$\begin{aligned} |e_0\rangle &\to |0\rangle \\ |e_1\rangle &\to |1\rangle \end{aligned} \qquad \text{Rate} = 1 \qquad \text{Fidelity} = 1$$

Coding blocs of 3 signal states into 2 qubits

Project onto 4-dim subspace spanned by states with majority of $|e_0\rangle$ (contains most of the weight):

$$\begin{aligned} |e_0\rangle|e_0\rangle|e_0\rangle &\to |0\rangle|0\rangle \\ |e_1\rangle|e_0\rangle|e_0\rangle &\to |0\rangle|1\rangle \\ |e_0\rangle|e_1\rangle|e_0\rangle &\to |1\rangle|0\rangle \\ |e_0\rangle|e_0\rangle|e_1\rangle &\to |1\rangle|1\rangle \end{aligned} \qquad \text{otherwise replace by } |0\rangle|0\rangle$$

Rate $= 2/3 = 0.67$

Prob. of success $= (0.9)^3 + 3(0.9)^2(0.1) = 0.97$

Fidelity $= 0.97 + 0.03(0.99)^3 = 0.99$

Source $\{|\psi_i\rangle, p_i\}$ with $|\psi_i\rangle \in \mathcal{H}$ and $i = 1, \ldots m$

$$\rightarrow \rho = \sum_{i=1}^{m} p_i \, \psi_i \qquad \text{with } \psi_i = |\psi_i\rangle\langle\psi_i|$$

Individual coding scheme: $|\psi_i\rangle \rightarrow \boxed{\text{ENC}} \rightarrow W_i$

with W_i keeping "most of the weight" of $|\psi_i\rangle$

$$\rightarrow \text{fidelity } F = \sum_{i=1}^{m} p_i \langle\psi_i|W_i|\psi_i\rangle$$

Diag: $\rho = \sum_{i=1}^{m} \lambda_i \, e_i \qquad \text{with } e_i = |e_i\rangle\langle e_i|$

$\mathcal{L} \equiv$ span(l eigenvectors of ρ with highest eigenvalues)

$\equiv$ l-dim subspace of $\mathcal{H}$ on which ρ has largest weight

$\Lambda \equiv \sum_{j=1}^{l} e_i$ projector onto $\mathcal{L}$ $\qquad \boxed{l < m}$

W_i keeps the part of ψ_i lying in $\mathcal{L}$

Space $\mathcal{L}^{\perp}$ (orthogonal complement of $\mathcal{L}$ in $\mathcal{H}$) is discarded

If $|\psi_i\rangle = \alpha_i|l_i\rangle + \alpha_i^\perp|l_i^\perp\rangle$ with $\begin{cases} |l_i\rangle \in \mathcal{L} \\ |l_i^\perp\rangle \in \mathcal{L}^\perp \end{cases}$

then $W_i = \underbrace{|\alpha_i|^2}_{\mathsf{Tr}(\Lambda\,\psi_i)}\ |l_i\rangle\langle l_i| \ + \ \underbrace{|\alpha_i^\perp|^2}_{\mathsf{Tr}(\Lambda^\perp\,\psi_i)}\ |e_1\rangle\langle e_1|$

Take l "big enough" so that

$$\mathsf{Tr}(\Lambda\rho) = \sum_{j=1}^{l} \lambda_j > 1-\epsilon$$

$$\begin{aligned}
F &= \sum_i p_i \mathsf{Tr}(W_i\,\psi_i) \\
&= \sum_i p_i \Big(|\alpha_i|^2 \underbrace{\mathsf{Tr}(|l_i\rangle\langle l_i|\,\psi_i)}_{|\langle l_i|\psi_i\rangle|^2 = |\alpha_i|^2} + |\alpha_i^\perp|^2 \underbrace{\mathsf{Tr}(|e_1\rangle\langle e_1|\,\psi_i)}_{|\langle e_1|\psi_i\rangle|^2 \geq 0} \Big) \\
&\geq \sum_i p_i\,|\alpha_i|^4 \\
&\geq \sum_i p_i \left(2|\alpha_i|^2 - 1\right) \qquad \text{use } x^4 \geq 2x^2 - 1 \\
&= \sum_i p_i \left(2\,\mathsf{Tr}(\Lambda\,\psi_i) - 1\right) \\
&= 2\,\mathsf{Tr}(\Lambda\rho) - 1
\end{aligned}$$

$\mathsf{Tr}(\Lambda\rho) > 1-\epsilon \ \Rightarrow\ F > 2(1-\epsilon) - 1 = 1 - 2\epsilon$

Block coding

Sequence of n independent signal states

$$\psi_{i_1} \otimes \psi_{i_2} \otimes \ldots \psi_{i_n} \in \mathcal{H}^{\otimes n}$$

Diag: $\rho^{\otimes n}$ has m^n eigenstates $e_{i_1} \otimes e_{i_2} \otimes \ldots e_{i_n}$
with eigenvalues = prob(sequence $i_1, i_2, \ldots i_n$)

$$= \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_n}$$

$\mathcal{L}$ = typical subspace of $\mathcal{H}^{\otimes n}$

= span(l typical sequences of eigenstates)

nr. of typical eigenstates $l \simeq 2^{nH[\lambda_i]} = 2^{nS(\rho)}$

typical eigenvalues $\simeq 2^{-nH[\lambda_i]} = 2^{-nS(\rho)}$.

$\forall \epsilon > 0, \delta > 0; \exists n$ and $\mathcal{L}$:

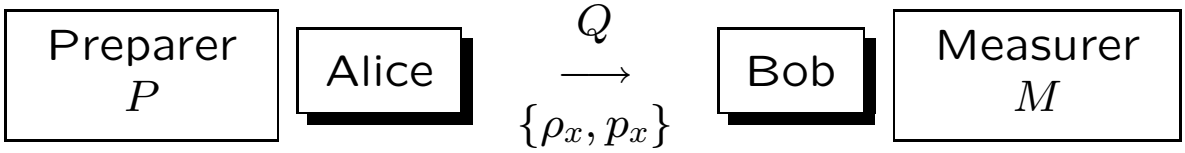
$\mathrm{Tr}(\Lambda) < 2^{n[S(\rho)+\delta]}$ and $\mathrm{Tr}(\Lambda \rho^{\otimes n}) > 1 - \epsilon$

There exists a coding scheme that uses at most $n[S(\rho) + \delta]$ qubits per sequence of n signal states, and gives a fidelity greater than $1 - \epsilon$.

Accessible information
(bound on)

A. S. Holevo, Probl. Inf. Transmission **9** (1973) 110

Communication of classical information over a quantum channel.



Alice prepares signal state ρ_x with probability p_x

Bob measures signal state, getting outcome y

For any measurement that Bob may do,

$$H(X{:}Y) \leq S(\rho) - \textstyle\sum_x p_x\, S(\rho_x)$$

with $\rho = \sum_x p_x \rho_x$

$$S(P{:}Q) = S(\rho_P) + S(\rho_Q) - S(\rho_{PQ}) = ?$$

$\rho_{PQ} = \sum_x p_x \, |x\rangle\langle x| \otimes \rho_x$ with $\{|x\rangle\}$ orthonormal

$\rho_P = \sum_x p_x \, |x\rangle\langle x| \quad \Rightarrow S(\rho_P) = H(X)$

$\rho_Q = \sum_x p_x \, \rho_x \equiv \rho \quad \Rightarrow S(\rho_Q) = S(\rho)$

$\Rightarrow S(\rho_{PQ}) = H(X) + \sum_x p_x S(\rho_x)$ block diagonal

thus $S(P{:}Q) = S(\rho) - \sum_x p_x \, S(\rho_x) =$ Holevo bound

Bob's generalized measurement (POVM)
characterized by $\{E_y\}$ with $E_y \geq 0$, $\sum_y E_y = 1$
$\longrightarrow$ prob. outcome $y = \mathrm{Tr}(E_y\,\rho)$

Before POVM

$\rho_{PQM} = \sum_x p_x\,|x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0|$

After POVM

$\rho'_{PQM} = \sum_{x,y} p_x\,|x\rangle\langle x| \otimes \sqrt{E_y}\,\rho_x\,\sqrt{E_y} \otimes |y\rangle\langle y|$

with $\{|y\rangle\}$ orthonormal

$\rho'_{PM} = \sum_{x,y} p_x \underbrace{\mathrm{Tr}(E_y \rho_x)}_{p_{y|x}}\,|x\rangle\langle x| \otimes |y\rangle\langle y|$

$=$ diagonal matrix with elements $p_{x,y} = p_x\,p_{y|x}$

$S(\rho'_P) = H(X)$
$S(\rho'_M) = H(Y)$
$S(\rho'_{PM}) = H(X,Y)$

thus $S(P'{:}M') = H(X{:}Y) =$ accessed information

remains to prove that $\boxed{S(P'{:}M') \leq S(P{:}Q)}$

$$\boxed{S(P{:}Q) = S(P{:}QM) \geq S(P'{:}Q'M') \geq S(P'{:}M')}$$

Including uncorrelated pure state conserves entropy

$$\rho_{PQM} = \rho_{PQ} \otimes |0\rangle\langle 0|$$
$$S(\rho_{PQM}) = S(\rho_{PQ}) + \underbrace{S(|0\rangle\langle 0|)}_{0} = S(\rho_{PQ})$$

Discarding a part cannot increase mutual information

$$S(P'{:}Q'M') = S(P'{:}M') + \underbrace{S(P'{:}Q'|M')}_{\geq 0}$$

strong subadditivity !!!

Quantum operation cannot increase mutual information

Operation can be simulated by adding an ancillary system A, acting unitarily on the joint system, then discarding the ancillary system A.

$$\begin{aligned} S(P{:}QM) &= S(P{:}QMA) && \text{including } A \\ &= S(P'{:}Q'M'A') && \text{unitary operation} \\ &\geq S(P'{:}Q'M') && \text{discarding } A \end{aligned}$$

Communication over noisy quantum channels (capacity for classical information)

Holevo, Schumacher, Westmoreland

$$\rho \longrightarrow \boxed{\text{channel } \mathcal{E}} \longrightarrow \rho'$$

Operator-sum representation of $\mathcal{E}$:

$\rho' = \mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger$

with $\sum_k E_k^\dagger E_k = I$ for trace preservation

If sending blocks

$\rho_{x_1} \otimes \rho_{x_2} \otimes \cdots \rho_{x_n} \longrightarrow \boxed{\text{channel } \mathcal{E}} \longrightarrow \rho'_{x_1} \otimes \rho'_{x_2} \otimes \cdots \rho'_{x_n}$

$$C_{\text{HSW}} = \max_{\{p_x, \rho_x\}} \left(S(\rho') - \sum_x p_x S(\rho'_x) \right)$$

$$\text{with } \begin{cases} \rho'_x = \mathcal{E}(\rho_x) \\ \rho' = \mathcal{E}(\sum_x p_x \rho_x) = \sum_x p_x \rho'_x \end{cases}$$

Note: Sufficient to maximize over ensembles of at most d^2 pure states

Note: Only product signal states
but joint (entangled) measurements needed

Quantum no-cloning principle

> Non-orthogonal quantum states cannot be copied exactly.

If $\langle\psi_0|\psi_1\rangle \neq 0$,
then $|\psi_i\rangle \rightarrow \boxed{\text{CLONER}} \rightarrow |\psi_i\rangle|\psi_i\rangle$ is impossible.

Indistinguishability $\Rightarrow$ No-cloning

Assume perfect cloning is possible:
Repeatedly clone the state $|\psi_i\rangle$ until we get $n \rightarrow \infty$ perfect copies $|\Psi_i\rangle = \underbrace{|\psi_i\rangle|\psi_i\rangle\cdots|\psi_i\rangle}_{n \text{ clones}}$

Since $\langle\Psi_0|\Psi_1\rangle \rightarrow 0$, they can be distinguished arbitrarily well, hence $|\psi_i\rangle$ can be distinguished arbitrarily well.

No-cloning $\Rightarrow$ Indistinguishability

Assume $|\psi_i\rangle$ are perfectly distinguishable:
Measure to identify whether it is $|\psi_0\rangle$ or $|\psi_1\rangle$.

Since the information (0 or 1) is classical, it can be cloned, hence the n clones $|\Psi_i\rangle$ can be prepared exactly.