Introduction to Quantum Information

Jeffrey Bub

Department of Philosophy
and
IPST
University of Maryland

RIT on Quantum Information and Computation, 2010

Outline

Web Resources

- Sam Lomonaco: *A Rosetta Stone for Quantum Mechanics with an Introduction to Quantum Computation*, http://arxiv.org/pdf/quant-ph/0007045.
- Todd Brun: *Lecture Notes on Quantum Information Processing*, http://almaak.usc.edu/ tbrun/Course/index.html
- Valerio Scarani: *Quantum Information: Primitive Notions and Quantum Correlations*, http://arxiv.org/pdf/0910.4222
- John Preskill: *Lecture Notes on Quantum Computation*, http://www.theory.caltech.edu/people/preskill/ph229/

Print Resources

- Sam Lomonaco: *A Rosetta Stone for Quantum Mechanics with an Introduction to Quantum Computation*, in *AMS Short Course Lecture Notes: Quantum Computation* (Providence: AMS, 2000).
- Michael A Nielsen and Isaac L. Chuang: *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press, 2000).
- Chris J. Isham: *Lectures on Quantum Theory: Mathematical and Structural Foundations* (London: Imperial College Press, 1995).

Print Resources

- Hoi-Kwong Lo, Sandu Popescu, Tom Spiller (eds.): *Introduction to Quantum Computation and Information* (World Scientific: 1998).
- L. Diosi: *A Short Course in Quantum Information Theory* (Springer, 2007).
- Michel Le Bellac: *A Short Introduction to Quantum Information and Quantum Computation* (Cambridge University Press, 2005).

Shannon Entropy

- Fundamental question considered by Shannon: how to quantify the minimal physical resources required to store messages produced by a source, so that they could be communicated via a channel without loss and reconstructed by a receiver.
- Shannon's source coding theorem (or noiseless channel coding theorem) answers this question.

Shannon Entropy

- Basic idea: consider a source that produces long sequences (messages) composed of symbols from a finite alphabet $a_1, a_2, \ldots, a_k$, where the individual symbols are produced with probabilities $p_1, p_2, \ldots, p_k$. A given sequence of symbols is represented as a sequence of values of independent and identically distributed (i.i.d.) discrete random variables $X_1, X_2, \ldots$
- A typical sequence of length n, for large n, will contain close to $p_i n$ symbols a_i, for $i = 1, \ldots, n$. So the probability of a sufficiently long typical sequence (assuming independence) will be:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

Shannon Entropy

- Basic idea: consider a source that produces long sequences (messages) composed of symbols from a finite alphabet $a_1, a_2, \ldots, a_k$, where the individual symbols are produced with probabilities $p_1, p_2, \ldots, p_k$. A given sequence of symbols is represented as a sequence of values of independent and identically distributed (i.i.d.) discrete random variables $X_1, X_2, \ldots$
- A typical sequence of length n, for large n, will contain close to $p_i n$ symbols a_i, for $i = 1, \ldots, n$. So the probability of a sufficiently long typical sequence (assuming independence) will be:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

Shannon Entropy

Taking the logarithm (conventionally, in information theory, to the base 2) yields:

$$\begin{aligned} \log p(x_1, \dots, x_n) &\approx \log p_1^{p_1 n} p_2^{p_2 n} \dots p_k^{p_k n} \\ &\approx n \sum_i p_i \log p_i \\ &= -nH(X) \end{aligned}$$

where $H(X) := -\sum_i p_i \log p_i$ is the Shannon entropy of the source.

Shannon Entropy

- If the probabilities p_i are all equal ($p_i = 1/k$ for all i), then $H(X) = \log k$, and if some $p_j = 1$ (and so $p_i = 0$ for $i \neq j$), then $H(X) = 0$ (taking $0 \log 0 = \lim_{x \to 0} x \log x = 0$). It can easily be shown that:

$$0 \leq H(X) \leq \log k.$$

- A source that produces one of two distinguishable symbols with equal probability, such as the toss of a fair coin, is said to have a Shannon entropy of 1 bit: ascertaining which symbol is produced is associated with an amount of information equal to 1 bit. If we already know which symbol will be produced (so the probabilities are 0 and 1), the entropy is 0: there is no uncertainty, and no information gain.

Shannon Entropy

- If the probabilities p_i are all equal ($p_i = 1/k$ for all i), then $H(X) = \log k$, and if some $p_j = 1$ (and so $p_i = 0$ for $i \neq j$), then $H(X) = 0$ (taking $0 \log 0 = \lim_{x \to 0} x \log x = 0$). It can easily be shown that:

$$0 \leq H(X) \leq \log k.$$

- A source that produces one of two distinguishable symbols with equal probability, such as the toss of a fair coin, is said to have a Shannon entropy of 1 bit: ascertaining which symbol is produced is associated with an amount of information equal to 1 bit. If we already know which symbol will be produced (so the probabilities are 0 and 1), the entropy is 0: there is no uncertainty, and no information gain.

Shannon Entropy

- If we encoded each of the k distinct symbols as a distinct binary number, i.e., as a distinct string of 0's and 1's, we would need strings composed of $\log k$ bits to represent each symbol ($2^{\log k} = k$).
- Shannon's analysis shows that messages produced by a stochastic source can be compressed, in the sense that (as $n \to \infty$ and the probability of an atypical n-length sequence tends to zero) n-length sequences can be encoded without loss of information using $nH(X)$ bits rather than the $n \log k$ bits required if we encoded each of the k symbols a_i as a distinct string of 0's and 1's: this *is* a compression, since $nH(X) < n \log k$ except for equiprobable distributions.

Shannon Entropy

- If we encoded each of the k distinct symbols as a distinct binary number, i.e., as a distinct string of 0's and 1's, we would need strings composed of $\log k$ bits to represent each symbol ($2^{\log k} = k$).
- Shannon's analysis shows that messages produced by a stochastic source can be compressed, in the sense that (as $n \to \infty$ and the probability of an atypical n-length sequence tends to zero) n-length sequences can be encoded without loss of information using $nH(X)$ bits rather than the $n \log k$ bits required if we encoded each of the k symbols a_i as a distinct string of 0's and 1's: this *is* a compression, since $nH(X) < n \log k$ except for equiprobable distributions.

Shannon Entropy

- Shannon's source coding theorem: the compression rate of $H(X)$ bits per symbol produced by a source of i.i.d. random variables is optimal.
- The Shannon entropy $H(X)$ is a measure of the minimal physical resources, in terms of the average number of bits per symbol, that are necessary and sufficient to reliably store the output of a source of messages. In this sense, it is a measure of the amount of information per symbol produced by an information source.
- The only relevant feature of a message with respect to reliable compression and decompression is the sequence of probabilities associated with the individual symbols: the nature of the physical systems embodying the representation of the message through their states is irrelevant to this notion of compression, as is the content or meaning of the message.

Shannon Entropy

- Shannon's source coding theorem: the compression rate of $H(X)$ bits per symbol produced by a source of i.i.d. random variables is optimal.
- The Shannon entropy $H(X)$ is a measure of the minimal physical resources, in terms of the average number of bits per symbol, that are necessary and sufficient to reliably store the output of a source of messages. In this sense, it is a measure of the amount of information per symbol produced by an information source.
- The only relevant feature of a message with respect to reliable compression and decompression is the sequence of probabilities associated with the individual symbols: the nature of the physical systems embodying the representation of the message through their states is irrelevant to this notion of compression, as is the content or meaning of the message.

Shannon Entropy

- Shannon's source coding theorem: the compression rate of $H(X)$ bits per symbol produced by a source of i.i.d. random variables is optimal.
- The Shannon entropy $H(X)$ is a measure of the minimal physical resources, in terms of the average number of bits per symbol, that are necessary and sufficient to reliably store the output of a source of messages. In this sense, it is a measure of the amount of information per symbol produced by an information source.
- The only relevant feature of a message with respect to reliable compression and decompression is the sequence of probabilities associated with the individual symbols: the nature of the physical systems embodying the representation of the message through their states is irrelevant to this notion of compression, as is the content or meaning of the message.

Shannon Entropy

- As a simple example of compression, consider an information source that produces sequences of symbols from a 4-symbol alphabet a_1, a_2, a_3, a_4 with probabilities $1/2$, $1/4$, $1/8$, $1/8$. Each symbol can be represented by a distinct 2-digit binary number:

$$\begin{aligned} a_1 &: \quad 00 \\ a_2 &: \quad 01 \\ a_3 &: \quad 10 \\ a_4 &: \quad 11 \end{aligned}$$

- So without compression we need two bits per symbol of storage space to store the output of the source.

Shannon Entropy

- The Shannon entropy of the source is:

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}$$

- Shannon's source coding theorem: there is a compression scheme that uses an average of 7/4 bits per symbol rather than two bits per symbol; such a compression scheme is optimal.

Shannon Entropy

- The Shannon entropy of the source is:

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}$$

- Shannon's source coding theorem: there is a compression scheme that uses an average of $7/4$ bits per symbol rather than two bits per symbol; such a compression scheme is optimal.

Shannon Entropy

The optimal scheme is provided by the following encoding:

$$\begin{aligned} a_1 &: \quad 0 \\ a_2 &: \quad 10 \\ a_3 &: \quad 110 \\ a_4 &: \quad 111 \end{aligned}$$

for which the *average* length of a compressed sequence is:

$$\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = \frac{7}{4}$$

bits per symbol.

Conditional Entropy

- So far, we've assumed a noiseless channel between the source and the receiver.
- An information channel maps inputs consisting of values of a random variable X onto outputs consisting of values of a random variable Y, and the map will generally not be 1-1 if the channel is noisy. So consider the conditional probabilities $p(y|x)$ of obtaining an output value y for a given input value x, for all x, y.

Conditional Entropy

- From the probabilities $p(x)$ we can calculate $p(y)$ as:

$$p(y) = \sum_x p(y|x)p(x)$$

and we can also calculate $p(x|y)$ by Bayes' rule from the probabilities $p(y|x)$ and $p(x)$, for all x, y, and hence the Shannon entropy of the conditional distribution $p(x|y)$, for all x and a fixed y, denoted by $H(X|Y = y)$.

- The quantity

$$H(X|Y) = \sum_y p(y)H(X|Y = y)$$

is known as the conditional entropy. It is the expected value of $H(X|Y = y)$ for all y.

Conditional Entropy

- From the probabilities $p(x)$ we can calculate $p(y)$ as:

$$p(y) = \sum_x p(y|x)p(x)$$

  and we can also calculate $p(x|y)$ by Bayes' rule from the probabilities $p(y|x)$ and $p(x)$, for all x, y, and hence the Shannon entropy of the conditional distribution $p(x|y)$, for all x and a fixed y, denoted by $H(X|Y = y)$.
- The quantity

$$H(X|Y) = \sum_y p(y)H(X|Y = y)$$

  is known as the conditional entropy. It is the expected value of $H(X|Y = y)$ for all y.

Conditional Entropy

- If we think of $H(X)$, the entropy of the distribution $\{p(x) : x \in \mathcal{X}\}$, as a measure of the uncertainty of the X-value, then $H(X|Y = y)$ is a measure of the uncertainty of the X-value, given the Y-value y, and $H(X|Y)$ is a measure of the average uncertainty of the X-value, given a Y-value.
- Putting it differently, the number of input sequences of length n that are consistent with a given output sequence (as $n \to \infty$) is $2^{nH(X|Y)}$, i.e., $H(X|Y)$ is the number of bits per symbol of additional information needed, on average, to identify an input X-sequence from a given Y-sequence.

Conditional Entropy

- If we think of $H(X)$, the entropy of the distribution $\{p(x) : x \in \mathcal{X}\}$, as a measure of the uncertainty of the X-value, then $H(X|Y = y)$ is a measure of the uncertainty of the X-value, given the Y-value y, and $H(X|Y)$ is a measure of the average uncertainty of the X-value, given a Y-value.
- Putting it differently, the number of input sequences of length n that are consistent with a given output sequence (as $n \to \infty$) is $2^{nH(X|Y)}$, i.e., $H(X|Y)$ is the number of bits per symbol of additional information needed, on average, to identify an input X-sequence from a given Y-sequence.

Conditional Entropy

- This follows because there are $2^{nH(X,Y)}$ typical sequences of pairs (x, y), where the joint entropy $H(X, Y)$ is calculated from the joint probability $p(x, y)$. So there are

$$\frac{2^{nH(X,Y)}}{2^{nH(Y)}} = 2^{n(H(X,Y)-H(Y))} = 2^{nH(X|Y)}$$

typical X-sequences associated with a given Y-sequence.

- Note that $H(X|Y) \neq H(Y|X)$.

Conditional Entropy

The equality

$$H(X,Y) - H(Y) = H(X|Y)$$

follows from the ‘chain rule’ equality

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) = H(Y,X)$$

derived from the logarithmic definitions of the quantities:

$$\begin{aligned} H(X,Y) &:= -\sum_{x,y} p(x,y)\log p(x,y) \\ &= -\sum_{x,y} p(x)p(y|x)\log\left(p(x)p(y|x)\right) \\ &= -\sum_{x,y} p(x)p(y|x)\log p(x) - \sum_{x,y} p(x)p(y|x)\log p(y|x) \\ &= H(X) + H(Y|X) \end{aligned}$$

Mutual Information

The mutual information $H(X:Y)$—sometimes $I(X:Y)$—of two random variables is a measure of how much information they have in common: the sum of the information content of the two random variables, as measured by the Shannon entropy (in which joint information is counted twice), minus their joint information.

$$H(X:Y) = H(X) + H(Y) - H(X,Y)$$

Mutual Information

- Note that $H(X:X) = H(X)$, as we would expect.
- Also, since $H(X,Y) = H(X) + H(Y|X)$, it follows that

$$H(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

i.e., the mutual information of two random variables represents the average information gain about one random variable obtained by measuring the other: the difference between the initial uncertainty of one of the random variables, and the average residual uncertainty of that random variable after ascertaining the value of the other random variable.

Entangled States

- Consider a quantum system Q which is part of a compound system QE (E for 'environment,' although E could be any quantum system of which Q is a subsystem). Pure states of QE are represented as rays or unit vectors in a tensor product Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^E$.
- A general pure state of QE is a state of the form:

$$|\Psi\rangle = \sum c_{ij}|q_i\rangle|e_j\rangle$$

where $|q_i\rangle \in \mathcal{H}^Q$ is a complete set of orthonormal states (a basis) in $\mathcal{H}^Q$ and $|e_j\rangle \in \mathcal{H}^E$ is a basis in $\mathcal{H}^E$. If the coefficients c_{ij} are such that $|\Psi\rangle$ cannot be expressed as a product state $|Q\rangle|E\rangle$, then $|\Psi\rangle$ is called an entangled state.

Entangled States

- Consider a quantum system Q which is part of a compound system QE (E for 'environment,' although E could be any quantum system of which Q is a subsystem). Pure states of QE are represented as rays or unit vectors in a tensor product Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^E$.
- A general pure state of QE is a state of the form:

$$|\Psi\rangle = \sum c_{ij}|q_i\rangle|e_j\rangle$$

where $|q_i\rangle \in \mathcal{H}^Q$ is a complete set of orthonormal states (a basis) in $\mathcal{H}^Q$ and $|e_j\rangle \in \mathcal{H}^E$ is a basis in $\mathcal{H}^E$. If the coefficients c_{ij} are such that $|\Psi\rangle$ cannot be expressed as a product state $|Q\rangle|E\rangle$, then $|\Psi\rangle$ is called an entangled state.

Entangled States

- For any state $|\Psi\rangle$ of QE, there exist orthonormal bases $|i\rangle \in \mathcal{H}^Q$, $|j\rangle \in \mathcal{H}^E$ such that $|\Psi\rangle$ can be expressed in a biorthogonal correlated form as:

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$$

  where the coefficients $\sqrt{p_i}$ are real and non-negative, and $\sum p_i = 1$.

- This representation is referred to as the Schmidt decomposition. The Schmidt decomposition is unique if and only if the p_i are all distinct.

Entangled States

- An example is the biorthogonal EPR state:

$$|\Psi\rangle = (|0\rangle|1\rangle - |1\rangle|0\rangle)/\sqrt{2};$$

  say, the singlet state of two spin-1/2 particles (the Schmidt form with positive coefficients is obtained by asborbing the relative phases in the definition of the basis vectors).

- In the singlet state, $|0\rangle$ and $|1\rangle$ can be taken as representing the two eigenstates of spin in the z-direction, but since the state is symmetric, $|\Psi\rangle$ retains the same form for spin in any direction.

Entangled States

- The EPR argument exploits the fact that spin measurements in the same direction on the two particles, which could be arbitrarily far apart, will yield outcomes that are perfectly anti-correlated for any spin direction.
-
- Bell's counterargument exploits the fact that when the spin is measured on one particle in a direction θ_1 to the z-axis, and on the other particle in a direction θ_2 to the z-axis, the probability of finding the same outcome for both particles (both 1 or both 0) is $\sin^2(\theta_1 - \theta_2)/2$. It follows that $3/4$ of the outcome pairs are the same when $\theta_1 - \theta_2 = 2\pi/3$.

Entangled States

- Suppose, for many EPR pairs, spin is measured in one of three directions $2\pi/3$ apart chosen randomly for each particle.
- It follows that, averaging over the nine possible pairs of measurement directions, half the outcome pairs will be the same ($\frac{1}{9}(3 \cdot 0 + 6 \cdot \frac{3}{4}) = \frac{1}{2}$). On the other hand, from Bell's inequality, derived under Einstein's realist assumptions of separability and locality, it can be shown that no more than $4/9$ of the outcome pairs can be the same.

Entangled States

- This means that the dynamical evolution of a quantum system can result in a state representing correlational information that no classical computer can simulate.
- For example, no classical computer can be programmed to perform the following task: for any pair of input angles, θ_1, θ_2, at different locations, output a pair of values (0 or 1) such that the values are perfectly correlated when $\theta_1 - \theta_2 = \pi$, perfectly anti-correlated when $\theta_1 = \theta_2$, and 75% correlated when $\theta_1 - \theta_2 = 2\pi/3$, where the response time between given the input and producing the output in each case is less than the time taken by light to travel between the two locations.

Entangled States

- This means that the dynamical evolution of a quantum system can result in a state representing correlational information that no classical computer can simulate.
- For example, no classical computer can be programmed to perform the following task: for any pair of input angles, θ_1, θ_2, at different locations, output a pair of values (0 or 1) such that the values are perfectly correlated when $\theta_1 - \theta_2 = \pi$, perfectly anti-correlated when $\theta_1 = \theta_2$, and 75% correlated when $\theta_1 - \theta_2 = 2\pi/3$, where the response time between given the input and producing the output in each case is less than the time taken by light to travel between the two locations.

Entangled States

The four states:

$$
\begin{aligned}
|1\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle) \\
|2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) \\
|3\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle) \\
|4\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)
\end{aligned}
$$

form an orthonormal basis, called the Bell basis, in the 2 x 2-dimensional Hilbert space.

Entangled States

Any Bell state can be transformed into any other Bell state by a local unitary transformation, X, Y, or Z, where X, Y, Z are the Pauli spin matrices:

$$X = \sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$Y = \sigma_y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$Z = \sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

For example:

$$X \otimes I \cdot |4\rangle = X \otimes I \cdot \frac{1}{\sqrt{2}}(|0\rangle\langle 1| - |1\rangle|0\rangle = -\frac{1}{\sqrt{2}}(|0\rangle\langle 0| - |1\rangle|1\rangle = -|3\rangle.$$

Entangled States

If QE is a closed system in an entangled pure state represented by

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$$

in the Schmidt decomposition, the expected value of any Q-observable A on $\mathcal{H}^Q$ can be computed as:

$$\begin{aligned}
\langle A\rangle &= \mathrm{Tr}(|\Psi\rangle\langle\Psi|A \otimes I) \\
&= \mathrm{Tr}_Q(\mathrm{Tr}_E(|\Psi\rangle\langle\Psi|A)) \\
&= \mathrm{Tr}_Q(\sum_i p_i|i\rangle\langle i|A) \\
&= \mathrm{Tr}_Q(\rho A)
\end{aligned}$$

Entangled States

So the expected value of the Q-observable A can be expressed as:

$$\langle A \rangle = \mathrm{Tr}_Q(\rho A)$$

where:

- $\mathrm{Tr}_Q() = \sum_q \langle q_i | \cdot | q_i \rangle$, for any orthonormal basis in $\mathcal{H}^Q$, is the partial trace over $\mathcal{H}^Q$,
- $\mathrm{Tr}_E()$ is the partial trace over $\mathcal{H}^E$, and
- $\rho = \sum_i p_i |i\rangle\langle i| \in \mathcal{H}^Q$ is the reduced density operator of the open system Q, a positive operator with unit trace.

Entangled States

- Since the density operator ρ yields the statistics of all Q-observables via the trace equation, ρ is taken as representing the quantum state of the system Q.
- If QE is an entangled pure state, then the open system Q is in a mixed state ρ, i.e., $\rho \neq \rho^2$; for pure states, ρ is a projection operator onto a ray and $\rho = \rho^2$.
- A mixed state represented by a density operator $\rho = \sum \rho_i |i\rangle\langle i|$ can be regarded as a mixture of pure states $|i\rangle$ prepared with prior probabilities p_i, but this representation is not unique—not even if the states combined in the mixture are orthogonal.

Entangled States

For example, the equal-weight mixture of orthonormal states $|0\rangle, |1\rangle$ in a 2-dimensional Hilbert space $\mathcal{H}_2$ has precisely the same statistical properties, and hence the same density operator $\rho = I/2$, as the equal weight mixture of any pair of orthonormal states, e.g.,

1. the states $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, or
2. the equal-weight mixture of nonorthogonal states $|0\rangle, \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle, \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle$ 120° degrees apart, or
3. the uniform continuous distribution over all possible states in $\mathcal{H}_2$.

Entangled States

- More generally, for any basis of orthonormal states $|e_i\rangle \in \mathcal{H}^E$, the entangled state $|\Psi\rangle$ can be expressed as:

$$|\Psi\rangle = \sum_{ij} c_{ij}|q_i\rangle|e_j\rangle = \sum_j \sqrt{w_j}|r_j\rangle|e_j\rangle$$

where the normalized states $|r_j\rangle = \sum_i \frac{c_{ij}}{\sqrt{w_j}}|q_i\rangle$ are relative states to the $|e_j\rangle$ ($\sqrt{w_j} = \sum_j |c_{ij}|^2$).

- Note that the states $|r_j\rangle$ are not in general orthogonal. Since the $|e_j\rangle$ are orthogonal, we can express the density operator representing the state of Q as:

$$\rho = \sum_i w_i|r_i\rangle\langle r_i|.$$

Entangled States

- More generally, for any basis of orthonormal states $|e_i\rangle \in \mathcal{H}^E$, the entangled state $|\Psi\rangle$ can be expressed as:

$$|\Psi\rangle = \sum_{ij} c_{ij}|q_i\rangle|e_j\rangle = \sum_j \sqrt{w_j}|r_j\rangle|e_j\rangle$$

  where the normalized states $|r_j\rangle = \sum_i \frac{c_{ij}}{\sqrt{w_j}}|q_i\rangle$ are relative states to the $|e_j\rangle$ ($\sqrt{w_j} = \sum_j |c_{ij}|^2$).
- Note that the states $|r_j\rangle$ are not in general orthogonal. Since the $|e_j\rangle$ are orthogonal, we can express the density operator representing the state of Q as:

$$\rho = \sum_i w_i|r_i\rangle\langle r_i|.$$

Entangled States

- In effect, a measurement of an E-observable with eigenstates $|e_i\rangle$ will leave the composite system QE in one of the states $|r_i\rangle|e_i\rangle$ with probability w_i, and a measurement of an E-observable with eigenstates $|i\rangle$ (the orthogonal states of the Schmidt decomposition) will leave the system QE in one of the states $|i\rangle|i\rangle$ with probability p_i.
- Since Q and E could be widely separated from each other in space, no measurement at E could affect the statistics of any Q-observable; or else measurements at E would allow superluminal signaling between Q and E.

Entangled States

- In effect, a measurement of an E-observable with eigenstates $|e_i\rangle$ will leave the composite system QE in one of the states $|r_i\rangle|e_i\rangle$ with probability w_i, and a measurement of an E-observable with eigenstates $|i\rangle$ (the orthogonal states of the Schmidt decomposition) will leave the system QE in one of the states $|i\rangle|i\rangle$ with probability p_i.
- Since Q and E could be widely separated from each other in space, no measurement at E could affect the statistics of any Q-observable; or else measurements at E would allow superluminal signaling between Q and E.

Entangled States

It follows that the mixed state ρ can be realized as a mixture of orthogonal states $|i\rangle$ (the eigenstates of ρ) with weights p_i, or as a mixture of non-orthogonal relative states $|r_j\rangle$ with weights w_j in infinitely many ways, depending on the choice of basis in $\mathcal{H}^E$:

$$\rho = \sum_i p_i |i\rangle\langle i| = \sum_j w_j |r_j\rangle\langle r_j|$$

and all these different mixtures with the same density operator ρ must be physically indistinguishable.

Entangled States

- Note that any mixed state density operator $\rho \in \mathcal{H}^Q$ can be 'purified' by adding a suitable ancilla system E, in the sense that ρ is the partial trace of a pure state $|\Psi\rangle \in \mathcal{H}^Q \otimes \mathcal{H}^E$ over $\mathcal{H}^E$.
- A purification of a mixed state is not unique, but depends on the choice of $|\Psi\rangle$ in $\mathcal{H}^E$.

Entangled States

- The Hughston-Jozsa-Wootters theorem (1993) shows that for *any* mixture of pure states $|r_i\rangle$ with weights w_i, where $\rho = \sum_j w_j |r_j\rangle\langle r_j|$, there is a purification of ρ and a suitable measurement on the system E that will leave Q in the mixture ρ.
- So an observer at E can remotely prepare Q in any mixture that corresponds to the density operator ρ (and of course all these different mixtures are physically indistinguishable).
- Similar results were proved earlier by Schrödinger (1935), Jaynes (1957), and Gisin (1989).

Measurement

- A standard von Neumann 'yes-no' measurement is associated with a projection operator; so a standard observable is represented in the spectral representation as a sum of projection operators, with coefficients representing the eigenvalues of the observable.
- Such a measurement is the quantum analogue of the measurement of a property of a system in classical physics. Classically, we think of a property of a system as being associated with a subset in the state space (phase space) of the system, and determining whether the system has the property amounts to determining whether the state of the system lies in the corresponding subset.

Measurement

- In quantum mechanics, the counterpart of a subset in phase space is a closed linear subspace in Hilbert space.
- Just as the different possible values of an observable (dynamical quantity) of a classical system correspond to the subsets in a mutually exclusive and collectively exhaustive set of subsets covering the classical state space, so the different values of a quantum observable correspond to the subspaces in a mutually exclusive (i.e., orthogonal) and collectively exhaustive set of subspaces spanning the quantum state space.

Measurement

- In quantum mechanics, and especially in the theory of quantum information (where any read-out of the quantum information encoded in a quantum state requires a quantum measurement), it is useful to consider a more general class of measurements than the projective measurements associated with the determination of the value of an observable.
- It is common to speak of generalized measurements and generalized observables. A generalized measurement is not a procedure that reveals whether or not a quantum system has some sort of generalized property. Rather, the point of the generalization is to exploit the difference between quantum and classical states for new possibilities in the representation and manipulation of information.

Measurement

- A quantum measurement can be characterized, completely generally, as a certain sort of interaction between two quantum systems, Q (the measured system) and M (the measuring system).
- We suppose that Q is initially in a state $|\psi\rangle$ and that M is initially in some standard state $|0\rangle$, where $|m\rangle$ is an orthonormal basis of 'pointer' eigenstates in $\mathcal{H}^M$.

Measurement

The interaction is defined by a unitary transformation U on the Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^M$ that yields the transition:

$$|\psi\rangle|0\rangle \xrightarrow{U} \sum_m M_m|\psi\rangle|m\rangle$$

where $\{M_m\}$ is a set of linear operators (the Kraus operators) defined on $\mathcal{H}^Q$ satisfying the completeness condition:

$$\sum_m M_m^\dagger M_m = I.$$

(The symbol $\dagger$ denotes the adjoint or Hermitian conjugate.)

Measurement

The completeness condition guarantees that this evolution is unitary, because it guarantees that U preserves inner products, i.e.

$$\begin{aligned} \langle\phi|\langle 0|U^\dagger U|\psi\rangle|0\rangle &= \sum_{m,m'}\langle m|\langle\phi|M_m^\dagger M_{m'}|\psi\rangle|m'\rangle \\ &= \sum_m \langle\phi|M_m^\dagger M_m|\psi\rangle \\ &= \langle\phi|\psi\rangle \end{aligned}$$

from which it follows that U, defined for any product state $|\psi\rangle|0\rangle$ (for any $|\psi\rangle \in \mathcal{H}^Q$) can be extended to a unitary operator on the Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^M$.

Measurement

Any set of linear operators $\{M_m\}$ defined on the Hilbert space of the system Q satisfying the completeness condition defines a measurement in this general sense, with the index m labeling the possible outcomes of the measurement, and any such set is referred to as a set of measurement operators.

Measurement

If we now perform a standard projective measurement on M to determine the value m of the pointer observable, defined by the projection operator

$$P_m = I_Q \otimes |m\rangle\langle m|$$

then the probability of obtaining the outcome m is:

$$\begin{aligned} p(m) &= \langle 0|\langle\psi|U^\dagger P_m U|\psi\rangle|0\rangle \\ &= \sum_{m'\,m''} \langle m'|\langle\psi|M^\dagger_{m'}(I_Q \otimes |m\rangle\langle m|)M_{m''}|\psi\rangle|m''\rangle \\ &= \sum_{m'\,m''} \langle\psi|M^\dagger_{m'}\langle m'|m\rangle\langle m|m''\rangle M_{m''}|\psi\rangle \\ &= \langle\psi|M^\dagger_m M_m|\psi\rangle \end{aligned}$$

and, more generally, if the initial state of Q is a mixed state ρ, then

$$p(m) = \mathrm{Tr}_Q(M_m \rho M^\dagger_m).$$

Measurement

- The final state of QM after the projective measurement on M yielding the outcome m is:

$$\frac{P_m U|\psi\rangle|0\rangle}{\sqrt{\langle\psi|U^\dagger P U|\psi\rangle}} = \frac{M_m|\psi\rangle|m\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}.$$

- So the final state of M is $|m\rangle$ and the final state of Q is:

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}};$$

and, more generally, if the initial state of Q is a mixed state ρ, then the final state of Q is:

$$\frac{M_m \rho M_m^\dagger}{\mathrm{Tr}_Q(M_m \rho M_m^\dagger)}.$$

Measurement

- This general notion of measurement covers the case of standard projective measurements. In this case $\{M_m\} = \{P_m\}$, where $\{P_m\}$ is the set of projection operators defined by the spectral measure of a standard quantum observable represented by a self-adjoint operator. It also covers the measurement of generalized observables associated with positive operator valued measures (POVMs).
- Let

$$E_m = M_m^\dagger M_m$$

then the set $\{E_m\}$ defines a set of positive operators ('effects') such that

$$\sum E_m = I$$

Measurement

- A POVM can be regarded as a generalization of a projection valued measure (PVM), in the sense that $\sum E_m = I$ defines a 'resolution of the identity' without requiring the PVM orthogonality condition:

$$P_m P_{m'} = \delta_{mm'} P_m.$$

- Note that for a POVM:

$$p(m) = \langle\psi|E_m|\psi\rangle.$$

Measurement

Given a set of positive operators $\{E_m\}$ such that $\sum E_m = I$, measurement operators M_m can be defined via

$$M_m = U\sqrt{E_m},$$

where U is a unitary operator, from which it follows that

$$\sum_m M_m^\dagger M_m = \sum E_m = I$$

Measurement

- As a special case we can take $U = 1$ and $M_m = \sqrt{E_m}$.
- Conversely, given a set of measurement operators $\{M_m\}$, there exist unitary operators U_m such that $M_m = U_m\sqrt{E_m}$, where $\{E_m\}$ is a POVM.

Measurement

- Except for the standard case of projective measurements, one might wonder why it might be useful to single out such unitary transformations, and why in the general case such a process should be called a measurement of Q.
- Suppose we know that a system with a 2-dimensional Hilbert space is in one of two nonorthogonal states:

$$\begin{aligned} |\psi_1\rangle &= |0\rangle \\ |\psi_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \end{aligned}$$

It is impossible to reliably distinguish these states by a quantum measurement, even in the above generalized sense. Here 'reliably' means that the state is identified correctly with zero probability of error.

Measurement

- Suppose there is such a measurement, defined by two measurement operators M_1, M_2 satisfying the completeness condition.
- Then we require

$$p(1) = \langle\psi_1|M_1^\dagger M_1|\psi_1\rangle = 1,$$

to represent reliability if the state is $|\psi_1\rangle$; and

$$p(2) = \langle\psi_2|M_2^\dagger M_2|\psi_2\rangle = 1$$

to represent reliability if the state is $|\psi_2\rangle$.

Measurement

- By the completeness condition we must have

$$\langle\psi_1|M_1^\dagger M_1 + M_2^\dagger M_2|\psi_1\rangle = 1$$

  from which it follows that $\langle\psi_1|M_2^\dagger M_2|\psi_1\rangle = 0$, i.e., $M_2|\psi_1\rangle = M_2|0\rangle = 0$.
- Hence

$$M_2|\psi_2\rangle = M_2\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}}M_2|1\rangle$$

  and so:

$$p(2) = \langle\psi_2|M_2^\dagger M_2|\psi_2\rangle = \frac{1}{2}\langle 1|M_2^\dagger M_2|1\rangle$$

Measurement

But by the completeness condition we also have

$$\langle 1|M_2^\dagger M_2|1\rangle \leq \langle 1|M_1^\dagger M_1 + M_2^\dagger M_2|1\rangle = \langle 1|1\rangle = 1$$

from which it follows that

$$p(2) \leq \frac{1}{2}$$

which contradicts $p(2) = 1$.

Measurement

However, it is possible to perform a measurement in the generalized sense, with *three* possible outcomes, that will allow us to correctly identify the state some of the time, i.e., for two of the possible outcomes, while nothing about the identity of the state can be inferred from the third outcome.

Measurement

Here's how: The three operators

$$\begin{aligned} E_1 &= \frac{\sqrt{2}}{1+\sqrt{2}} \frac{(|0\rangle - |1\rangle)(\langle 0| - \langle 1|)}{2} \\ E_2 &= \frac{\sqrt{2}}{1+\sqrt{2}} |1\rangle\langle 1| \\ E_3 &= I - E_1 - E_2 \end{aligned}$$

are all positive operators and $E_1 + E_2 + E_3 = I$, so they define a POVM.

Measurement

In fact, E_1, E_2, E_3 are each multiples of projection operators onto the states

$$\begin{aligned} |\phi_1\rangle &= |\psi_2\rangle^\perp \\ |\phi_2\rangle &= |\psi_1\rangle^\perp \\ |\phi_3\rangle &= \frac{(1+\sqrt{2})|0\rangle + |1\rangle}{\sqrt{2\sqrt{2}(1+\sqrt{2})}} \end{aligned}$$

with coefficients $\frac{\sqrt{2}}{1+\sqrt{2}}, \frac{\sqrt{2}}{1+\sqrt{2}}, \frac{1}{1+\sqrt{2}}$ respectively.

Measurement

The measurement involves a system M with three orthogonal pointer states $|1\rangle, |2\rangle, |3\rangle$. The appropriate unitary interaction U results in the transition, for an input state $|\psi\rangle$:

$$|\psi\rangle|0\rangle \xrightarrow{U} \sum_m M_m|\psi\rangle|m\rangle$$

where $M_m = \sqrt{E_m}$.

Measurement

- If the input state is $|\psi_1\rangle = |0\rangle$, we have the transition:

$$\begin{aligned} |\psi_1\rangle|0\rangle &\xrightarrow{U} \sqrt{E_1}|0\rangle|1\rangle + \sqrt{E_3}|0\rangle|3\rangle \\ &= \alpha|\phi_1\rangle|1\rangle + \beta|\phi_3\rangle|3\rangle \end{aligned}$$

(because $\sqrt{E_2}|\psi_1\rangle = \sqrt{E_2}|0\rangle = 0$).

- if the input state is $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, we have the transition:

$$\begin{aligned} |\psi_2\rangle|0\rangle &\xrightarrow{U} \sqrt{E_2}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|2\rangle + \sqrt{E_3}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|3\rangle \\ &= \gamma|\phi_2\rangle|2\rangle + \delta|\phi_3\rangle|3\rangle \end{aligned}$$

(because $\sqrt{E_1}|\psi_2\rangle = \sqrt{E_1}\frac{|0\rangle+|1\rangle}{\sqrt{2}} = 0$), where $\alpha, \beta, \gamma, \delta$ are real numerical coefficients.

Measurement

- We see that a projective measurement of the pointer of M that yields the outcome $m = 1$ indicates, with certainty, that the input state was $|\psi_1\rangle = |0\rangle$. In this case, the measurement leaves the system Q in the state $|\phi_1\rangle$.
- A measurement outcome $m = 2$ indicates, with certainty, that the input state was $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and in this case the measurement leaves the system Q in the state $|\phi_2\rangle$.
- If the outcome is $m = 3$, the input state could have been either $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and Q is left in the state $|\phi_3\rangle$.

Quantum Operations

- When a closed system QE initially in a product state $\rho \otimes \rho_E$ evolves under a unitary transformation, Q can be shown to evolve under a quantum operation, i.e., a completely positive linear map:

$$\mathcal{E} : \rho \to \rho'$$

$$\mathcal{E}(\rho) = \mathrm{Tr}_E(U\rho \otimes \rho_E U^\dagger)$$

- The map $\mathcal{E}$ is linear (or convex-linear) in the sense that $\mathcal{E}(\sum_i p_i \rho_i) = \sum_i p_i \mathcal{E}(p_i)$, positive in the sense that $\mathcal{E}$ maps positive operators to positive operators, and completely positive in the sense that $\mathcal{E} \otimes I$ is a positive map on the extension of $\mathcal{H}^Q$ to a Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^E$, associated with the addition of any ancilla system E to Q.

Quantum Operations

- Every quantum operation (i.e., completely positive linear map) on a Hilbert space $\mathcal{H}^Q$ has a (non-unique) representation as a unitary evolution on an extended Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^E$, i.e.,

$$\mathcal{E}(\rho) = \mathrm{Tr}_E(U(\rho \otimes \rho_E)U^\dagger)$$

  where ρ_E is an appropriately chosen initial state of an ancilla system E (which we can think of as the environment of Q).
- It turns out that it suffices to take ρ_E as a pure state, i.e., $|0\rangle\langle 0|$, since a mixed state of E can always be purified by enlarging the Hilbert space (i.e., adding a further ancilla system). So the evolution of a system Q described by a quantum operation can always be modeled as the unitary evolution of a system QE, for an initial pure state of E.

Quantum Operations

- Every quantum operation (i.e., completely positive linear map) on a Hilbert space $\mathcal{H}^Q$ has a (non-unique) representation as a unitary evolution on an extended Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^E$, i.e.,

$$\mathcal{E}(\rho) = \mathrm{Tr}_E(U(\rho \otimes \rho_E)U^\dagger)$$

  where ρ_E is an appropriately chosen initial state of an ancilla system E (which we can think of as the environment of Q).
- It turns out that it suffices to take ρ_E as a pure state, i.e., $|0\rangle\langle 0|$, since a mixed state of E can always be purified by enlarging the Hilbert space (i.e., adding a further ancilla system). So the evolution of a system Q described by a quantum operation can always be modeled as the unitary evolution of a system QE, for an initial pure state of E.

Quantum Operations

- Also, every quantum operation on a Hilbert space $\mathcal{H}^Q$ has a (non-unique) operator sum representation intrinsic to $\mathcal{H}^Q$:

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger$$

where $E_i = \langle i|U|0\rangle$ for some orthonormal basis $\{|i\rangle\}$ of E.

- If the operation is trace-preserving (or nonselective), then $\sum_i E_i^\dagger E_i = I$. For operations that are not trace-preserving (or selective), $\sum_i E_i^\dagger E_i \leq I$. This corresponds to the case where the outcome of a measurement on QE is taken into account (selected) in the transition $\mathcal{E} \to \mathcal{E}(\rho)$.

Quantum Operations

- If there is no interaction between Q and E, then $\mathcal{E}(\rho) = U_Q \rho U_Q^\dagger$, $U_Q U_Q^\dagger = I$, i.e., there is only one operator in the sum. In this case, $U = U_Q \otimes U_E$ and

$$\begin{aligned} \mathcal{E}(\rho) &= \mathrm{Tr}_E(U_Q \otimes U_E(\rho \otimes |0\rangle\langle 0|)U_Q^\dagger \otimes U_E^\dagger) \\ &= U_Q \rho U_Q^\dagger . \end{aligned}$$

- So unitary evolution is a special case of the operator sum representation of a quantum operation and, of course, another special case is the transition $\mathcal{E} \rightarrow \mathcal{E}(\rho)$ that occurs in a quantum measurement process, where $E_i = M_i$.

Quantum Operations

A trace-preserving operation corresponds to a non-selective measurement:

$$\mathcal{E}(\rho) = \sum_i M_i \rho M_i^\dagger$$

while an operation that is not trace-preserving corresponds to a selective measurement, where the state 'collapses' onto the corresponding measurement outcome:

$$M_i \rho M_i^\dagger / \mathrm{Tr}(M_i \rho M_i^\dagger)$$

Quantum Operations

- The operator sum representation applies to quantum operations between possibly different input and output Hilbert spaces, and characterizes the following general situation: a quantum system in an unknown initial state ρ is allowed to interact unitarily with other systems prepared in standard states, after which some part of the composite system is discarded, leaving the final system in a state ρ'. The transition $\rho \to \rho'$ is defined by a quantum operation.
- So a quantum operation represents, quite generally, the unitary evolution of a closed quantum system, the nonunitary evolution of an open quantum system in interaction with its environment, and evolutions that result from a combination of unitary interactions and selective or nonselective measurements.

Quantum Operations

- The creed of the Church of the Larger Hilbert Space is that every state can be made pure, every measurement can be made ideal, and every evolution can be made unitary—on a larger Hilbert space.
- The Creed originates with John Smolin. This formulation is due to Ben Schumacher. See his *Lecture Notes on Quantum Information Theory*.

Von Neumann Entropy

- Information in Shannon's sense is a quantifiable resource associated with the output of a (suitably idealized) stochastic source of symbolic states, where the physical nature of the systems embodying these states is irrelevant to the amount of classical information associated with the source.
- The quantity of information associated with a stochastic source is defined by its optimal compressibility, and this is given by the Shannon entropy.
- The fact that some feature of the output of a stochastic source can be optimally compressed is, ultimately, what justifies the attribution of a quantifiable resource to the source.

Von Neumann Entropy

- Information is represented physically in the states of physical systems. The essential difference between classical and quantum information arises because of the different distinguishability properties of classical and quantum states.
- Only sets of orthogonal quantum states are reliably distinguishable (i.e., with zero probability of error), as are sets of different classical states (which are represented by disjoint singleton subsets in a phase space, and so are orthogonal as subsets of phase space in a sense analogous to orthogonal subspaces of a Hilbert space).

Von Neumann Entropy

- Classical informationis that sort of information represented in a set of distinguishable states—states of classical systems, or orthogonal quantum states—and so can be regarded as a subcategory of quantum information, where the states may or may not be distinguishable.
- The idea behind quantum information is to extend Shannon's notion of compressibility to a stochastic source of quantum states, which may or may not be distinguishable. For this we need to define a suitable measure of information for probability distributions of quantum states—mixtures—as a generalization of the notion of Shannon entropy.

Von Neumann Entropy

- Consider a system QE in an entangled state $|\Psi\rangle$. Then the subsystem Q is in a mixed state ρ, which can always be expressed as:
$$\rho = \sum_i p_i |i\rangle\langle i|$$
where the p_i are the eigenvalues of ρ and the pure states $|i\rangle$ are orthonormal eigenstates of ρ.
- This is the spectral representation of ρ, and any density operator—a positive (hence Hermitian) operator—can be expressed in this way.

Von Neumann Entropy

The representation is unique if and only if the p_i are all distinct. If some of the p_i are equal, there is a unique representation of ρ as a sum of projection operators with the distinct values of the p_i as coefficients, but some of the projection operators will project onto multi-dimensional subspaces.

Von Neumann Entropy

- Since ρ has unit trace, $\sum p_i = 1$, and so the spectral representation of ρ represents a classical probability distribution of orthogonal, and hence distinguishable, pure states.
- If we measure a Q-observable with eigenstates $|i\rangle$, then the outcomes can be associated with the values of a random variable X, where $\Pr(X = i) = p_i$. Then

$$H(X) = -\sum p_i \log p_i$$

is the Shannon entropy of the probability distribution of measurement outcomes.

Von Neumann Entropy

Now,

$$-\mathrm{Tr}(\rho \log \rho) = -\sum p_i \log p_i$$

(because the eigenvalues of $\rho \log \rho$ are $p_i \log p_i$ and the trace of an operator is the sum of the eigenvalues), so a natural generalization of Shannon entropy for any mixture of quantum states with density operator ρ is the von Neumann entropy:

$$S := -\mathrm{Tr}(\rho \log \rho)$$

which coincides with the Shannon entropy for measurements in the eigenbasis of ρ.

Von Neumann Entropy

- For a completely mixed state $\rho = I/d$, where $\mathrm{dim}\mathcal{H}^Q = d$, the d eigenvalues of ρ are all equal to $1/d$ and $S = \log d$.
- $\log d$ is the maximum value of S in a d-dimensional Hilbert space.
- The von Neumann entropy S is zero, the minimum value, if and only if ρ is a pure state, where the eigenvalues of ρ are 1 and 0.
- So $0 \leq S \leq \log d$, where d is the dimension of $\mathcal{H}^Q$.

Von Neumann Entropy

- We can think of the Shannon entropy as a measure of the average amount of information gained by identifying the state produced by a known stochastic source. Alternatively, the Shannon entropy represents the optimal compressibility of the information produced by an information source.
- The von Neumann entropy does *not*, in general, represent the amount of information gained by identifying the quantum state produced by a stochastic source characterized as a mixed state, because nonorthogonal quantum states in a mixture cannot be reliably identified.

Von Neumann Entropy

- We can think of the Shannon entropy as a measure of the average amount of information gained by identifying the state produced by a known stochastic source. Alternatively, the Shannon entropy represents the optimal compressibility of the information produced by an information source.
- The von Neumann entropy does *not*, in general, represent the amount of information gained by identifying the quantum state produced by a stochastic source characterized as a mixed state, because nonorthogonal quantum states in a mixture cannot be reliably identified.

Von Neumann Entropy

The von Neumann entropy can be interpreted in terms of compressibility via Schumacher's source coding theorem for quantum information, a generalization of Shannon's source coding theorem for classical information.

Von Neumann Entropy

- For an elementary two-state quantum system with a 2-dimensional Hilbert space considered as representing the output of an elementary quantum information source, $S = 1$ for an equal weight distribution over two orthogonal states (i.e., for the density operator $\rho = I/2$), so Schumacher takes the basic unit of quantum information as the 'qubit.'
- By analogy with the term 'bit,' the term 'qubit' refers to the basic unit of quantum information in terms of the von Neumann entropy, and to an elementary two-state quantum system considered as representing the possible outputs of an elementary quantum information source.

Von Neumann Entropy

The difference between quantum information as measured by von Neumann entropy S and classical information as measured by Shannon entropy H can be brought out by considering the quantum notions of conditional entropy and mutual information, and in particular the peculiar feature of inaccessibility associated with quantum information.

Von Neumann Entropy

For a composite system AB, conditional von Neumann entropy and mutual information are defined in terms of the joint entropy $S(AB) = -\mathrm{Tr}(\rho^{AB} \log \rho^{AB})$ by analogy with the corresponding notions for Shannon entropy:

$$\begin{aligned} S(A|B) &= S(A,B) - S(B) \\ S(A\!:\!B) &= S(A) - S(A|B) \\ &= S(B) - S(B|A) \\ &= S(A) + S(B) - S(A,B) \end{aligned}$$

Von Neumann Entropy

The joint entropy satisfies the subadditivity inequality:

$$S(A, B) \leq S(A) + S(B)$$

with equality if and only if A and B are uncorrelated, i.e., $\rho^{AB} = \rho^A \otimes \rho^B$.

Von Neumann Entropy

- $S(A|B)$ can be negative, while the conditional Shannon entropy is always positive or zero.
- Consider the entangled state $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. Since $|\Psi\rangle$ is a pure state, $S(A, B) = 0$. But $S(A) = S(B) = 1$. So $S(A|B) = S(A, B) - S(A) = -1$.
- In fact, for a pure state $|\Psi\rangle$ of a composite system AB, $S(A|B) < 0$ if and only if $|\Psi\rangle$ is entangled.

Von Neumann Entropy

- $S(A|B)$ can be negative, while the conditional Shannon entropy is always positive or zero.
- Consider the entangled state $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. Since $|\Psi\rangle$ is a pure state, $S(A, B) = 0$. But $S(A) = S(B) = 1$. So $S(A|B) = S(A, B) - S(A) = -1$.
- In fact, for a pure state $|\Psi\rangle$ of a composite system AB, $S(A|B) < 0$ if and only if $|\Psi\rangle$ is entangled.

Von Neumann Entropy

- $S(A|B)$ can be negative, while the conditional Shannon entropy is always positive or zero.
- Consider the entangled state $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. Since $|\Psi\rangle$ is a pure state, $S(A, B) = 0$. But $S(A) = S(B) = 1$. So $S(A|B) = S(A, B) - S(A) = -1$.
- In fact, for a pure state $|\Psi\rangle$ of a composite system AB, $S(A|B) < 0$ if and only if $|\Psi\rangle$ is entangled.

Von Neumann Entropy

- For a composite system AB in a product state $\rho \otimes \sigma$, it follows from the definition of joint entropy that:

$$S(A, B) = S(\rho \otimes \sigma) = S(\rho) + S(\sigma) = S(A) + S(B).$$

Von Neumann Entropy

If AB is in a pure state $|\Psi\rangle$, it follows from the Schmidt decomposition theorem that $|\Psi\rangle$ can be expressed as

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle\langle i|$$

from which it follows that

$$\begin{array}{rcccl} \rho_A & = & \mathrm{Tr}_B(|\Psi\rangle\langle\Psi|) & = & \sum_i p_i|i\rangle\langle i| \\ \rho_B & = & \mathrm{Tr}_A(|\Psi\rangle\langle\Psi|) & = & \sum_i p_i|i\rangle\langle i|; \end{array}$$

and so:

$$S(A) = S(B) = -\sum_i p_i \log p_i .$$

Von Neumann Entropy

- For a mixed state prepared as a mixture of states ρ_i with weights p_i, it can be shown that

$$S(\sum_i p_i \rho_i) \leq H(p_i) + \sum_i p_i S(\rho_i)$$

with equality if and only if the states ρ_i have support on orthogonal subspaces.

- The entropy $H(p_i)$ is referred to as the entropy of preparation of the mixture ρ.

Von Neumann Entropy

- For a mixed state prepared as a mixture of states ρ_i with weights p_i, it can be shown that

$$S(\sum_i p_i \rho_i) \leq H(p_i) + \sum_i p_i S(\rho_i)$$

  with equality if and only if the states ρ_i have support on orthogonal subspaces.
- The entropy $H(p_i)$ is referred to as the entropy of preparation of the mixture ρ.

Von Neumann Entropy

- If the states ρ_i are pure states, then $S(\rho) \leq H(p_i)$
- For example, suppose $\mathcal{H}^Q$ is 2-dimensional and $p_1 = p_2 = 1/2$, then $H(p_i) = 1$.
- So if we had a classical information source producing the symbols 1 and 2 with equal probabilities, no compression of the information would be possible. However, if the symbols 1 and 2 are encoded as nonorthogonal quantum states $|r_1\rangle$ and $|r_2\rangle$, then $S(\rho) < 1$.

Von Neumann Entropy

According to Schumacher's source coding theorem, since $S(\rho) < 1$, quantum compression is possible, i.e., we can transmit long sequences of qubits reliably using $S < 1$ qubits per quantum state produced by the source.

Von Neumann Entropy

The von Neumann entropy of a mixture of states ρ_i with weights p_i, $\sum p_i \rho_i$, is a concave function of the states in the distribution, i.e.,

$$S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i).$$

Von Neumann Entropy

To see this, consider a composite system AB in the state

$$\rho^{AB} = \sum p_i \rho_i \otimes |i\rangle\langle i|.$$

We have

$$S(A) = S(\sum_i p_i \rho_i)$$

$$S(B) = S(\sum_i p_i |i\rangle\langle i|) = H(p_i)$$

and

$$S(A,B) = H(p_i) + \sum_i p_i S(\rho_i)$$

By subadditivity, $S(A) + S(B) \geq S(A,B)$, so:

$$S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i).$$

Von Neumann Entropy

- It turns out that projective measurements always increase entropy, i.e., if $\rho' = \sum_i P_i \rho P_i$, then $S(\rho') \geq S(\rho)$, but generalized measurements can *decrease* entropy.
- Consider, for example, the generalized measurement on a qubit in the initial state ρ defined by the measurement operators $M_1 = |0\rangle\langle 0|$ and $M_2 = |0\rangle\langle 1|$. (Note that these operators *do* define a generalized measurement because $M_1^\dagger M_1 + M_2^\dagger M_2 = |0\rangle\langle 0| + |1\rangle\langle 1| = I$.)

Von Neumann Entropy

- It turns out that projective measurements always increase entropy, i.e., if $\rho' = \sum_i P_i \rho P_i$, then $S(\rho') \geq S(\rho)$, but generalized measurements can *decrease* entropy.
- Consider, for example, the generalized measurement on a qubit in the initial state ρ defined by the measurement operators $M_1 = |0\rangle\langle 0|$ and $M_2 = |0\rangle\langle 1|$. (Note that these operators *do* define a generalized measurement because $M_1^\dagger M_1 + M_2^\dagger M_2 = |0\rangle\langle 0| + |1\rangle\langle 1| = I$.)

Von Neumann Entropy

After the measurement

$$\begin{aligned} \rho' &= |0\rangle\langle 0|\rho|0\rangle\langle 0| + |0\rangle\langle 1|\rho|1\rangle\langle 0| \\ &= \mathrm{Tr}(\rho)|0\rangle\langle 0| \\ &= |0\rangle\langle 0|. \end{aligned} \tag{1}$$

So $S(\rho') = 0 \leq S(\rho)$.

Accessible Information

- The ability to exploit quantum states to perform new sorts of information-processing tasks arises because quantum states have different distinguishability properties than classical states. Of course, it is not the mere lack of distinguishability of quantum states that is relevant here, but the different sort of distinguishability enjoyed by quantum states.
- This indistinguishability is reflected in the limited accessibility of quantum information.

Accessible Information

- Consider a classical information source in Shannon's sense, with Shannon entropy $H(X)$. Suppose the source produces symbols represented as the values x (in an alphabet $\mathcal{X}$) of a random variable X, with probabilities p_x, and that the symbols are encoded as quantum states ρ_x, $x \in X$.
- The mutual information $H(X : Y)$ is a measure of how much information one gains, on average, about the value of the random variable X on the basis of the outcome Y of a measurement on a given quantum state.

Accessible Information

The accessible information is defined as:

$$\text{Sup } H(X : Y)$$

over all possible measurements.

Accessible Information

- The Holevo bound on mutual information provides an important upper bound to accessible information:

$$H(X:Y) \leq S(\rho) - \sum_x p_x S(\rho_x)$$

where $\rho = \sum_x p_x \rho_x$ and the measurement outcome Y is obtained from a measurement defined by a POVM $\{E_y\}$.

- Since $S(\rho) - \sum_x p_x S(\rho_x) \leq H(X)$, with equality if and only if the states ρ_x have orthogonal support, we have:

$$H(X:Y) \leq H(X)$$

Accessible Information

- Note that X can be distinguished from Y if and only if $H(X:Y) = H(X)$.
- If the states ρ_x are orthogonal pure states, then in principle there exists a measurement that will distinguish the states, and for such a measurement $H(X:Y) = H(X)$.
- In this case, the accessible information is the same as the entropy of preparation of the quantum states, $H(X)$.
- But if the states are nonorthogonal, then $H(X:Y) < H(X)$ and there is no measurement, even in the generalized sense, that will enable the reliable identification of X.

Accessible Information

- If the values of X are encoded as the pure states of a qubit, then $H(X:Y) \leq S(\rho)$ and $S(\rho) \leq 1$. It follows that at most 1 bit of information can be extracted from a qubit by measurement.
- If X has k equiprobable values, $H(X) = \log k$. Alice could encode these k values into a qubit by preparing it in an equal-weight mixture of k nonorthogonal pure states, but Bob could only extract at most 1 bit of information about the value of X.

Accessible Information

For an n-state quantum system associated with an n-dimensional Hilbert space, $S(\rho) \leq \log n$. So even though Alice could encode any amount of information into such an n-state quantum system (by preparing the state as a mixture of nonorthogonal states), the most information that Bob could extract from the state by measurement is $\log n$, which is the same as the maximum amount of information that could be encoded into and extracted from an n-state classical system.

Accessible Information

- It might seem, then, that the inaccessibility of quantum information as quantified by the Holevo bound would thwart any attempt to exploit quantum information to perform nonclassical information-processing tasks.
- Surprisingly, the inaccessibility of quantum information can actually be exploited in information-processing tasks that transcend the scope of classical information.

Deriving the Holevo Bound

- To derive the Holevo bound (see Nielsen and Chuang, Theorem 12.1), suppose Alice encodes the distinguishable symbols of a classical information source with entropy $H(X)$ as quantum states ρ_x (not necessarily orthogonal).
- That is, Alice has a quantum system P, the preparation device, with an orthonormal pointer basis $|x\rangle$ corresponding to the values of the random variable X, which are produced by the source with probabilities p_x.

Deriving the Holevo Bound

- The preparation interaction correlates the pointer states $|x\rangle$ with the states ρ_x of a quantum system Q, so that the final state of P and Q after the preparation interaction is:

$$\rho^{PQ} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x.$$

- Alice sends the system Q to Bob, who attempts to determine the value of the random variable X by measuring the state of Q.

Deriving the Holevo Bound

- The initial state of P, Q, and Bob's measuring instrument M is:
$$\rho^{PQM} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0|$$
where $|0\rangle\langle 0|$ is the initial ready state of M.
- Bob's measurement can be described by a quantum operation $\mathcal{E}$ on the Hilbert space $\mathcal{H}^Q \otimes \mathcal{H}^M$ that stores a value of y, associated with a POVM $\{E_y\}$ on $\mathcal{H}^Q$, in the pointer state $|y\rangle$ of M, i.e., $\mathcal{E}$ is defined for any state $\sigma \in \mathcal{H}^Q$ and initial ready state $|0\rangle \in \mathcal{H}^M$ by:
$$\sigma \otimes |0\rangle\langle 0| \xrightarrow{\mathcal{E}} \sum_y \sqrt{E_y}\sigma\sqrt{E_y} \otimes |y\rangle\langle y|.$$

Deriving the Holevo Bound

- From the definition of quantum mutual information:

$$S(P:Q) = S(P:Q,M)$$

because M is initially uncorrelated with PQ and

$$S(P':Q',M') \leq S(P:Q,M)$$

because it can be shown that quantum operations never increase mutual information (primes here indicate states after the application of $\mathcal{E}$).

Deriving the Holevo Bound

The notation $S(P:Q,M)$ refers to the mutual information between the system P and the composite system consisting of the system Q and the measuring device M, in the initial state $\rho^{PQ} = \sum_x p_x|x\rangle\langle x| \otimes \rho_x$. That is, the comma notation refers to the joint system:
$S(P:Q,M) = S(P) - S(P|Q,M) = S(P) + S(Q,M) - S(P,Q,M)$

Deriving the Holevo Bound

Finally:

$$S(P' : Q', M')$$

because discarding systems never increases mutual information, and so:

$$S(P' : M') \leq S(P : Q)$$

which (following some algebraic manipulation) is the statement of the Holevo bound, i.e., $(S(P' : M') \leq S(P : Q)$ reduces to $H(X : Y) \leq S(\rho) - \sum_x p_x S(\rho_x)$.

Deriving the Holevo Bound

- To see this, note that

$$\rho^{PQ} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$$

- So $S(P) = H(p_{x)}), S(Q) = S(\sum_x p_x \rho_x) = S(\rho)$ and:

$$S(P, Q) = H(p_{x)} + \sum_x p_x S(\rho_x)$$

since the states $|x\rangle\langle x| \otimes \rho_x$ have support on orthogonal subspaces in $\mathcal{H}^P \otimes \mathcal{H}^Q$.

- It follows that

$$\begin{aligned} S(P:Q) &= S(P) + S(Q) - S(P, Q) \\ &= S(\rho) - \sum_x p_x S(\rho_x) \end{aligned}$$

which is the right hand side of the Holevo bound.

Deriving the Holevo Bound

For the left hand side:

$$\begin{aligned}
\rho^{P'M'} &= \mathrm{Tr}_{Q'}(\rho^{P'Q'M'}) \\
&= \mathrm{Tr}_{Q'}(\sum_{xy} p_x |x\rangle\langle x| \otimes \sqrt{E_y}\rho_x\sqrt{E_y} \otimes |y\rangle\langle y|) \\
&= \sum_{xy} p_x \mathrm{Tr}(E_y \rho_x E_y)|x\rangle\langle x| \otimes |y\rangle\langle y| \\
&= \sum_{xy} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y|
\end{aligned}$$

since $p(x,y) = p_x p(y \mid x) = p_x \mathrm{Tr}(\rho_x E_y) = p_x \mathrm{Tr}(\sqrt{E_y}\rho_x\sqrt{E_y})$, and so $S(P' : M') = H(X : Y)$.

Deriving the Holevo Bound

- The Holevo bound limits the representation of classical bits by qubits. Putting it another way, the Holevo bound characterizes the resource cost of encoding classical bits as qubits: one qubit is necessary and sufficient.
- Can we represent qubits by bits? If so, what is the cost of a qubit in terms of bits?

Deriving the Holevo Bound

- This question is answered by a result by Barnum, Hayden, Jozsa, and Winter (2001): A quantum source of nonorthogonal signal states can be compressed with arbitarily high fidelity to α qubits per signal plus any number of classical bits per signal if and only if α is at least as large as the von Neumann entropy S of the source.
- This means that a generic quantum source cannot be separated into a classical and quantum part: quantum information cannot be traded for any amount of classical information.