QI Lecture 20

Limitations

Non-Distinguishability of Non-Orthogonal States

Assume that it's possible to make a measurement of non-orthogonal states that can accurately distinguish them. Assume we have two states, $|\psi_1\rangle$ and $|\psi_1\rangle$ such that $\langle\psi_1|\psi_2\rangle \neq 0$, as well as a POVM, $\{\Pi_1, \Pi_2\}$, such that

$$\langle\psi_1|\Pi_1|\psi_1\rangle = \langle\psi_2|\Pi_2|\psi_2\rangle = 1$$

Because of the completeness relation for POVMs, $\Pi_1 + \Pi_2 = I$, we have that

$$1 = \langle\psi_1|\psi_1\rangle = \langle\psi_1|\left(\Pi_1 + \Pi_2\right)|\psi_1\rangle = \langle\psi_1|\Pi_1|\psi_1\rangle + \langle\psi_1|\Pi_2|\psi_1\rangle$$

and therefore (making the same argument for $|\psi_2\rangle$)

$$\langle\psi_1|\Pi_2|\psi_1\rangle = \langle\psi_2|\Pi_1|\psi_2\rangle = 0$$

We can decompose $|\psi_2\rangle$ into components perpendicular and parallel to $|\psi_1\rangle$ in this way

$$|\psi_2\rangle = \alpha|\psi_1\rangle + \beta|\varphi\rangle$$

Since $\langle\psi_1|\varphi\rangle = 0$, we have that $|\alpha|^2 + |\beta|^2 = 1$ (by the Pythagorean theorem). Since $\langle\psi_1|\psi_2\rangle \neq 0$ we have that $\alpha \neq 0$ and therefore $|\beta| < 1$. But this is a contradiction, because

$$1 = \langle\psi_2|\Pi_2|\psi_2\rangle = |\beta|^2\langle\varphi|\Pi_2|\varphi\rangle \leq |\beta|^2\langle\varphi|\left(\Pi_1 + \Pi_2\right)|\varphi\rangle = |\beta|^2\langle\varphi|\varphi\rangle = |\beta|^2 < 1$$

A discerning logical mind will observe that $1 < 1$ is false. The only way around the contradiction is $|\beta| = 1$ and $\alpha = 0$. In other words; in order for this perfectly-distinguishing POVM to exist, the states need to be perpendicular.

The No-Cloning Theorem

We'd like to copy an arbitrary, unknown quantum state $|\psi\rangle$. When we copy a page of text we create a new page of text and the two have no connection to one another.

To copy a state, we need some unitary process that acts on an arbitrary state $|\psi\rangle$ and copies it onto some (possibly prepared) "blank" state,[1] $|s\rangle$.

$$U\left[|\psi\rangle_a|s\rangle_b\right] = |\psi\rangle_a|\psi\rangle_b$$

Clearly this procedure is only a true "quantum copier" if it works for more than one state.[2] Assume that it works for at least $|\psi\rangle$ and $|\phi\rangle$.

$$U\left[|\phi\rangle_a|s\rangle_b\right] = |\phi\rangle_a|\phi\rangle_b$$

Take the inner product of these two equations

$$\begin{aligned}
\left[\langle\phi|_a\langle s|_b U^\dagger\right]\left[U|\psi\rangle_a|s\rangle_b\right] &= \left[\langle\phi|_a\langle\phi|_b\right]\left[|\psi\rangle_a|\psi\rangle_b\right] \\
\langle\phi|_a\langle s|_b U^\dagger U|\psi\rangle_a|s\rangle_b &= \langle\phi|\psi\rangle_a\langle\phi|\psi\rangle_b \\
\langle\phi|_a\langle s|_b|\psi\rangle_a|s\rangle_b &= \left[\langle\phi|\psi\rangle_a\right]^2 \\
\langle\phi|\psi\rangle_a\langle s|s\rangle_b &= \left[\langle\phi|\psi\rangle_a\right]^2 \\
\langle\phi|\psi\rangle_a &= \left[\langle\phi|\psi\rangle_a\right]^2
\end{aligned}$$

We immediately have that $\langle\phi|\psi\rangle = 0, 1$.

So either $\langle\phi|\psi\rangle = 1$ and the states are the same (which contradicts our assumption) or $\langle\phi|\psi\rangle = 0$ and the states are orthogonal. Again we find that an arbitrary cloner is impossible. For example, a potential cloner cannot copy both $|0\rangle$ and $|+\rangle$ because they aren't orthogonal, but it may clone both $|0\rangle$ and $|1\rangle$, because they are.

Notice that, since the cloner's states are orthogonal and specific, what we're essentially describing here is "measure in the computational basis, and then make that state twice" which is what a classical computer (or a photocopier) does.

Remarkably, it is possible to *approximately* clone an unknown 2-level state[3] with a probability of up to $\frac{5}{6}$.

The No-Cloning theorem was originally discovered as part of an attempt to explain how Nick Herbert was wrong about a certain faster-than-light scheme. Herbert believed that the effects of measurements are carried by "wave-function collapse" through the universe. For example, if Alice and Bob share $|\Phi_+\rangle$ and Alice does a measurement resulting in $|0\rangle_a$, then the overall state is $|00\rangle$: Bob no longer has an entangled qubit, he has $|0\rangle_b$.

Herbert's idea was that an entangled photon pair can be created, one sent to Bob on Mars and one to Alice on Earth, and that by copying the state of the photon sent to Bob,

[1] There's no such thing as a "blank state". A blank sheet of paper isn't nothing. It doesn't say much, but it's something.

[2] Otherwise it's a "prepare that one state" machine.

[3] You can read all the details in Bužek and Hillery's "Quantum copying: Beyond the no-cloning theorem".

Figure 1: The Fundamental Fysiks Group. Nick Herbert (in the back) wrote a paper describing a method for super-luminal communication requiring the creation of many copies of the same entangled state. Finding his error led to the discovery of the "no-cloning theorem".

Alice could communicate with him. By copying the photon's state through lasing,[4] Bob could measure many photons and determine the state of the original.

In order to send one bit of information, Alice measures her qubit in either the $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$ bases. The photon going to Bob instantly (faster than light) collapses to the same state. If Alice sees $|0\rangle_a$, then Bob will receive and copy $|0\rangle_b$. With those copies in hand, Bob measures at a range of angles and determines that $|\langle\theta|0\rangle|^2 = \cos^2(\theta)$. If he had received $|1\rangle_b$, he would have found that $|\langle\theta|1\rangle|^2 = \cos^2\left(\theta - \frac{\pi}{2}\right)$. So although Bob knows that the 0/1 result is random, the choice of the $\{|0\rangle, |1\rangle\}$ basis was not. If he copies and measures the incoming photon and finds that $|\langle\theta|+\rangle|^2 = \cos^2\left(\theta - \frac{\pi}{4}\right)$ or $|\langle\theta|-\rangle|^2 = \cos^2\left(\theta + \frac{\pi}{4}\right)$, then

[4] In a laser, the presence of many identical nearby photons makes the probability of the emission of new photons in an identical state much higher. As a result, laser light is coherent, with all of the photons (nearly) at the same frequency and phase.

Alice must have chosen to measure in the $\{|+\rangle, |-\rangle\}$ basis.[5] This implies that Alice should be able to communicate with Bob instantaneously by either measuring or not measuring her state and Bob should be able to detect her signal by looking at the statistics of all of the photons he sees.

The No-Communication Theorem

Alice and Bob start in the usual situation, sharing some entangled state[6] between them,

$$\rho = \sum_{ijk\ell} C_{ijk\ell} |i\rangle_a |j\rangle_b \langle k|_a \langle \ell|_b \in A \otimes B$$

If Alice, using local operations, does *anything* to her piece of the state, will Bob be able to tell?

By "anything" of course we mean "a quantum operation", $\mathcal{E}[\rho] = \sum_m E_m \rho E_m^\dagger$ where $\sum_m E_m^\dagger E_m = I$. The operator sum representation covers essentially every allowable quantum process imaginable: unitary operations, undisclosed measurements, noise, abandoning states, etc. Because Alice is performing a local operation on only space A, all of her operations take the form $(E_m)_a \otimes I_b$.

Everything that Bob can discern from his half of the state is encapsulated in his reduced density matrix

$$\begin{aligned} \rho_b &= Tr_a[\rho] \\ &= \sum_n \langle n|_a \left(\sum_{ijk\ell} C_{ijk\ell} |i\rangle_a |j\rangle_b \langle k|_a \langle \ell|_b \right) |n\rangle_a \\ &= \sum_{nj\ell} C_{njn\ell} |j\rangle_b \langle \ell|_b \end{aligned}$$

After Alice does whatever she chooses, the state will be transformed into

$$\rho' = \mathcal{E}[\rho] = \sum_m \left(E_m \otimes I\right) \rho \left(E_m^\dagger \otimes I\right)$$

Once again, all that Bob can know about the state is contained in the partial trace over A,

[5] If you're beginning to feel the need to argue against this experiment or you've noticed some holes in it, keep in mind: it's not real and no version of it works in even the smallest degree.

[6] It sometimes helps to write the density matrix out completely with complete generality.

$$
\begin{aligned}
Tr_a\left[\rho'\right] &= Tr_a\left[\textstyle\sum_m \left(E_m \otimes I\right) \rho \left(E_m^\dagger \otimes I\right)\right] \\
&= \textstyle\sum_n \langle n|_a \left(\sum_m \left(E_m \otimes I\right) \left(\sum_{ijk\ell} C_{ijk\ell} |i\rangle_a |j\rangle_b \langle k|_a \langle \ell|_b\right) \left(E_m^\dagger \otimes I\right)\right) |n\rangle_a \\
&= \textstyle\sum_{ijk\ell mn} C_{ijk\ell} \langle n|_a E_m |i\rangle_a \langle k|_a E_m^\dagger |n\rangle_a |j\rangle_b \langle \ell|_b \\
&= \textstyle\sum_{ijk\ell mn} C_{ijk\ell} \langle k|_a E_m^\dagger |n\rangle_a \langle n|_a E_m |i\rangle_a |j\rangle_b \langle \ell|_b \\
&= \textstyle\sum_{ijk\ell m} C_{ijk\ell} \langle k|_a E_m^\dagger \left(\sum_n |n\rangle_a \langle n|_a\right) E_m |i\rangle_a |j\rangle_b \langle \ell|_b \\
&= \textstyle\sum_{ijk\ell m} C_{ijk\ell} \langle k|_a E_m^\dagger E_m |i\rangle_a |j\rangle_b \langle \ell|_b \\
&= \textstyle\sum_{ijk\ell} C_{ijk\ell} \langle k|_a \left(\sum_m E_m^\dagger E_m\right) |i\rangle_a |j\rangle_b \langle \ell|_b \\
&= \textstyle\sum_{ijk\ell} C_{ijk\ell} \langle k|i\rangle_a |j\rangle_b \langle \ell|_b \\
&= \textstyle\sum_{jk\ell} C_{kjk\ell} |j\rangle_b \langle \ell|_b \\
&= \rho_b
\end{aligned}
$$

In other words, as we've been using this entire class, nothing that Alice does to her half of an entangled state has any direct, detectable effect on Bob's half. Remember that, as with classical probability, if Alice lets Bob know about her state, such as the results of measurements done to it, then Bob's reduced density matrix can certainly change. That's the key idea behind teleportation and POVMs[7], but without that intervention of classical information, nothing about Bob's state ever changes. This is the difference between "local operators" and "local operators with classical communication" (LOCC).

This theorem is absolute. The no-cloning theorem permits a little leeway, but the no-communication theorem has no exceptions of any kind in any degree.

The Uncertainty Principle

The expectation of an operator is always defined with respect to some state,

$$E[M] = \langle \psi | M | \psi \rangle$$

Define the variance of an operator M to be

$$(\Delta M)^2 \equiv E\left[\left(M - E[M]\right)^2\right] = \langle \psi | \left(M - E[M]\right)^2 | \psi \rangle = \langle \psi | M^2 | \psi \rangle - \left(\langle \psi | M | \psi \rangle\right)^2$$

The commutator and anticommutator are

[7] Via Neumark's theorem.

$$[A,B] = AB - BA \qquad \{A,B\} = AB + BA$$

and we assume that A and B are hermitian ($A^\dagger = A$).
Define x and y such that $\langle\psi|AB|\psi\rangle = x + yi$. It follows that

$$x - yi = (\langle\psi|AB|\psi\rangle)^\dagger = \langle\psi|B^\dagger A^\dagger|\psi\rangle = \langle\psi|BA|\psi\rangle$$

Therefore

$$\langle\psi|[A,B]|\psi\rangle = \langle\psi|AB|\psi\rangle - \langle\psi|BA|\psi\rangle = (x+iy) - (x-iy) = 2yi$$

$$\langle\psi|\{A,B\}|\psi\rangle = \langle\psi|AB|\psi\rangle + \langle\psi|BA|\psi\rangle = (x+iy) + (x-iy) = 2x$$

and since $|\langle\psi|AB|\psi\rangle|^2 = x^2 + y^2$ we have that

$$|\langle\psi|[A,B]|\psi\rangle|^2 + |\langle\psi|\{A,B\}|\psi\rangle|^2 = 4|\langle\psi|AB|\psi\rangle|^2$$

The Cauchy Inequality says that

$$|\langle\phi|N^\dagger M|\psi\rangle|^2 \le \|N|\phi\rangle\|^2 \, \|M|\psi\rangle\|^2 = \langle\phi|N^\dagger N|\phi\rangle\langle\psi|M^\dagger M|\psi\rangle$$

and since A and B are hermitian we have that

$$|\langle\psi|AB|\psi\rangle|^2 \le \langle\psi|A^2|\psi\rangle\langle\psi|B^2|\psi\rangle$$

Which means that

$$|\langle\psi|[A,B]|\psi\rangle|^2 \le |\langle\psi|[A,B]|\psi\rangle|^2 + |\langle\psi|\{A,B\}|\psi\rangle|^2 = 4|\langle\psi|AB|\psi\rangle|^2 \le 4\langle\psi|A^2|\psi\rangle\langle\psi|B^2|\psi\rangle$$

and therefore

$$\boxed{|\langle\psi|[A,B]|\psi\rangle| \le 2\sqrt{\langle\psi|A^2|\psi\rangle}\sqrt{\langle\psi|B^2|\psi\rangle}}$$

Finally, suppose that C and D are observables and that $A = C - E[C]$ and $B = D - E[D]$. On the left we have

$$\begin{aligned}
& |\langle\psi|[A,B]|\psi\rangle| \\
= \; & |\langle\psi|\,[AB - BA]\,|\psi\rangle| \\
= \; & |\langle\psi|\,[(C - E[C])\,(D - E[D]) - (D - E[D])\,(C - E[C])]\,|\psi\rangle| \\
= \; & |\langle\psi|\,[CD - E[D]C - E[C]D + E[C]E[D] - DC + E[C]D + E[D]C - E[C]E[D]]\,|\psi\rangle| \\
= \; & |\langle\psi|\,[CD - DC]\,|\psi\rangle| \\
= \; & |\langle\psi|\,[C,D]\,|\psi\rangle|
\end{aligned}$$

and on the right we have

$$2\sqrt{\langle\psi|A^2|\psi\rangle}\sqrt{\langle\psi|B^2|\psi\rangle} = 2\sqrt{\langle\psi|\left(C - E[C]\right)^2|\psi\rangle}\sqrt{\langle\psi|\left(D - E[D]\right)^2|\psi\rangle} = 2\Delta C\Delta D$$

So if a pair of operators don't commute, then there is a lower limit to the product of their standard deviations.

$$\boxed{\frac{|\langle\psi|[C,D]|\psi\rangle|}{2} \leq \Delta C\Delta D}$$

This is the "Heisenberg Uncertainty Principle". It is <u>not</u> about the effect of one measurement on another, it's a statement about the results of many measurements, using C and D, on many identically prepared states, $|\psi\rangle$. After many measurements we can build up a statistical distribution of results from which we can calculate the standard deviations, ΔC and ΔD, and for a given state $|\psi\rangle$ we find that the Heisenberg Uncertainty Principle is always satisfied.

Example Consider our favorite operators, X and Z, and the states

$$|0\rangle \qquad |\circlearrowleft\rangle = \frac{|0\rangle + i|1\rangle}{\sqrt{2}}$$

Note that $[X,Z] = XZ - ZX = 2XZ = -2iY$.

First $|\psi\rangle = |0\rangle$. By choosing an eigenstate of Z we have ensured that the standard deviation, ΔZ, is zero.

$$\Delta X = \sqrt{\langle 0|X^2|0\rangle - \left[\langle 0|X|0\rangle\right]^2} = \sqrt{\langle 0|0\rangle - \left[\langle 0|1\rangle\right]^2} = 1$$

$$\Delta Z = \sqrt{\langle 0|Z^2|0\rangle - \left[\langle 0|Z|0\rangle\right]^2} = \sqrt{\langle 0|0\rangle - \left[\langle 0|0\rangle\right]^2} = 0$$

$$\frac{|\langle 0|[X,Z]|0\rangle|}{2} = \frac{|-2i\langle 0|Y|0\rangle|}{2} = |\langle 0|Y|0\rangle| = |i\langle 0|1\rangle| = 0$$

The honed mathematical mind will perceive that, indeed,

$$0 \leq 1 \cdot 0$$

Now $|\psi\rangle = |\circlearrowleft\rangle$. By choosing an eigenstate of Y we have maximized the value of $\frac{|\langle 0|[X,Z]|0\rangle|}{2}$.

$$\begin{aligned}
\Delta X &= \sqrt{\langle \circlearrowleft | X^2 | \circlearrowleft \rangle - [\langle \circlearrowleft | X | \circlearrowleft \rangle]^2} \\
&= \sqrt{\langle \circlearrowleft \mid \circlearrowleft \rangle - \left[\left(\tfrac{\langle 0| - i\langle 1|}{\sqrt{2}}\right)\left(\tfrac{|1\rangle + i|0\rangle}{\sqrt{2}}\right)\right]^2} \\
&= \sqrt{1 - \left[\tfrac{i-i}{2}\right]^2} \\
&= 1
\end{aligned}$$

$$\begin{aligned}
\Delta Z &= \sqrt{\langle \circlearrowleft | Z^2 | \circlearrowleft \rangle - [\langle \circlearrowleft | Z | \circlearrowleft \rangle]^2} \\
&= \sqrt{\langle \circlearrowleft \mid \circlearrowleft \rangle - \left[\left(\tfrac{\langle 0| - i\langle 1|}{\sqrt{2}}\right)\left(\tfrac{|0\rangle - i|1\rangle}{\sqrt{2}}\right)\right]^2} \\
&= \sqrt{1 - \left[\tfrac{1-1}{2}\right]^2} \\
&= 1
\end{aligned}$$

$$\frac{|\langle \circlearrowleft | [X, Z] | \circlearrowleft \rangle|}{2} = \frac{|-2i\langle \circlearrowleft | Y | \circlearrowleft \rangle|}{2} = |\langle \circlearrowleft \mid \circlearrowleft \rangle| = 1$$

Once again, the trained intellect will note that

$$1 \leq 1 \cdot 1$$

■

Holevo's Bound

In order to communicate with Bob, Alice produces a set of states, $\{\rho_x\}$, with probabilities $p(\rho_x) = p_x$, defining her random variable X. Bob uses a POVM, $\{\Pi_y\}$, to measure those states and recover Alice's message. This POVM gives rise to Bob's random variable, Y, since $p(y|x) = Tr[\Pi_y \rho_x]$. We quantify the amount of information communicated using the mutual information between Alice and Bob's random variables, $I[X;Y]$.

Holevo's bound says that

$$\boxed{I[X;Y] \leq \chi \equiv S\left[\sum_x p_x \rho_x\right] - \sum_x p_x S[\rho_x]}$$

The "Holevo χ quantity" is the average drop in Bob's entropy when Alice informs him which state she sent, or equivalently when Bob successfully figures it out for himself. χ is a hard bound on the accessible information available to Bob. Recall that mutual information is used to quantify communication, since a sent message and a received message

is information that is shared (and thus "mutual"). So χ acts like the classical channel capacity[8] when using quantum states for communication.

The derivation of Holevo's bound has deep roots, one theorem based on another based on another, so here we'll be a bit more philosophical. We've seen before[9] that

$$H[\{p_k\}] + \sum_k p_k S[\rho_k] = S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right] \leq H[\{p_k\}] + S\left[\sum_x p_x \rho_x\right]$$

When Alice is sending messages to Bob, the state of the full system can be represented as $\sum_k p_k |k\rangle\langle k| \otimes \rho_k$, because $|k\rangle\langle k| \otimes \rho_k$ represents the situation where Alice has chosen "k" and sent ρ_k to Bob in order to convey that fact. There's no ambiguity about which state was sent (just ask Alice), so the total entropy (left equality) is the classical Shannon entropy of the selection itself, $H[\{p_k\}]$, plus the average entropy of the state that Bob receives, $\sum_k p_k S[\rho_k]$.

The subadditivity of Von Neumann entropy says that the entropy of a system is less than or equal to the sum of its parts (right inequality). For Alice $S\left[\sum_k p_k |k\rangle\langle k|\right] = H\left[\{p_k\}\right]$, since ρ_a is clearly diagonalized. For Bob, without input from Alice, the entropy of ρ_b is $S\left[\sum_x p_x \rho_x\right]$ meaning all that Bob can say is that with probability p_x he has received ρ_x.

In very, *very* brief, here's the idea behind the proof of Holevo's Bound. The quantum mutual information, $S[A;B]$, like the classical mutual information,[10] $I[X;Y]$, is the sum of the entropies of the two subsystems independently, minus the entropy of the system as a whole

$$S[A;B] = S[A] + S[B] - S[A,B]$$

which in this case is

$$S[A]+S[B]-S[A,B] = H[\{p_k\}]+S\left[\sum_x p_x \rho_x\right]-S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right] = S\left[\sum_x p_x \rho_x\right]-\sum_x p_x S[\rho_x]$$

When Bob tries to measure his system and extract information from it he applies a quantum channel that can't increase the mutual information, but may decrease it.

[8] It isn't quite the same, since the classical channel capacity is maximized over the probability distribution on X and here that distribution, $\{p_x\}$, is given.

[9] Lecture 9

[10] History is resplendent with examples of someone somewhere writing something off the top of their head and then everyone else writing it the same way for the sake of consistency. The "I" in mutual classical information as opposed to the "S" for everything in quantum information is a prime example.

Example Alice wants to send information using $0_L = |0\rangle$ and $1_L = |+\rangle$. The entropy of her data is already maximized (she doesn't want to waste time sending data that isn't as dense as possible), so $p_0 = p_1 = \frac{1}{2}$.

What is the maximum amount of information that Bob can recover from Alice's signals?

The density matrices are

$$\rho_0 = |0\rangle\langle 0| \sim \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \rho_1 = |+\rangle\langle +| \sim \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

and since these are both pure states, $S[\rho_0] = S[\rho_1] = 0$. We have that the first term in χ is

$$S\left[\frac{1}{2}\rho_0 + \frac{1}{2}\rho_1\right] = S\left[\begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}\right]$$

and to calculate this we need to find the eigenvalues.

$$\begin{aligned} 0 &= \begin{vmatrix} \frac{3}{4} - \lambda & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} - \lambda \end{vmatrix} \\ 0 &= \left(\tfrac{3}{4} - \lambda\right)\left(\tfrac{1}{4} - \lambda\right) - \left(\tfrac{1}{4}\right)^2 \\ 0 &= \lambda^2 - \lambda + \tfrac{3}{16} - \tfrac{1}{16} \\ 0 &= \lambda^2 - \lambda + \tfrac{1}{8} \\ \lambda &= \tfrac{1}{2} \pm \tfrac{1}{2\sqrt{2}} \end{aligned}$$

Finally, we have that

$$\chi = -\left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)\log_2\left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right) - \left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right)\log_2\left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right) \approx 0.60$$

meaning that if Alice can only use $|0\rangle$ and $|+\rangle$ to communicate, then she can only send 0.6 bits on average per use of the channel. Here, since $\langle 0|+\rangle \neq 0$, the channel is essentially the symmetric channel (where there is a possibility that the bit will flip).

■

Qubit ⟶ Bit

If Alice's messages are selected from a 2^n dimensional space of quantum states, then

$$\chi = S\left[\sum_x p_x \rho_x\right] - \sum_x p_x S[\rho_x] \leq S\left[\sum_x p_x \rho_x\right] \leq \log_2\left(2^n\right) = n$$

In the best-case scenario, where $\{\rho_x\}$ is a set of 2^n mutually orthogonal pure states and $p_x = 2^{-n}$, the maximum amount of information that can be communicated is n bits. A prime example of a 2^n dimensional quantum state is n qubits.

This is a huge bottleneck for quantum computers. Internally they can compute with a superposition of truly colossal number of states, but must communicate the results using a number of bits that is the log of the number of internal states (at best).

The Threshold Theorem

At present the big stumbling block for the construction of functioning quantum computers is noise. Individually, every quantum operation works well the vast majority of the time. But if you string enough qubits together and do enough operations on them, then in short order you're effectively guaranteed to have a serious, calculation-destroying error.

As we saw with the Shor code, we can handle errors on a qubit by encoding it in a larger "logical qubit". As with the classical "majority voting" code ($0_L \equiv 000$, $1_L \equiv 111$), we find that through judicious encoding a (small) error probability p is reduced to $O\left(p^2\right)$. Through a concatenation of codes, we can further reduce this to $O\left(p^4\right)$, $O\left(p^8\right)$, etc. at the expense of having huge codes and the need for methods of computation that can be applied directly to the encoded states.

The Threshold Theorem states that it is possible to achieve arbitrarily low overall error probability, using a reasonably finite machine, assuming that the individual error probability is lower than some threshold probability, p_{th}. There are a lot of moving parts to this,[11] and this result is more the providence of engineers than mathematicians, so you're unlikely to find the Threshold Theorem written out succinctly. But, at least according to Nielsen and Chuang, a good estimate for the threshold probability is in the neighborhood of

$$\boxed{p_{th} \approx 10^{-5} - 10^{-6}}$$

We're still a couple of orders of magnitude away from that, but fortunately there's no (known) fundamental barrier and we're advancing fast.[12]

[11] The ability to supply "clean" ancilla states, parallelizable architectures, a not-insignificant classical computation cost, etc.

[12] I'd say we're definitely no more than a few years or centuries away from efficient, large scale quantum computers. Guaranteed. Probably.

Exercises

1) Holevo Good Time

If Alice is sending bits using

$$\rho_0 = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \qquad \rho_1 = \left(\frac{i|0\rangle + \sqrt{3}|1\rangle}{2}\right)\left(\frac{-i\langle 0| + \sqrt{3}\langle 1|}{2}\right)$$

with $p_0 = \frac{1}{3}$ and $p_1 = \frac{2}{3}$, then what is the maximum amount of information that can be transmitted, on average, to Bob?

2) But What About Super-Dense Coding?

Does super-dense coding violate the "1 bit per qubit" limit of Holevo's Bound? Why or why not?

3) But What About CNOT?

The CNOT gate seems to copy one qubit onto another, since if the target starts as $|0\rangle_b$, then

$$CNOT|0\rangle_a|0\rangle_b = |0\rangle_a|0\rangle_b \qquad CNOT|1\rangle_a|0\rangle_b = |1\rangle_a|1\rangle_b$$

Does CNOT clone $|\psi\rangle_a = \alpha|0\rangle_a + \beta|1\rangle_a$ and violate the No-Cloning Theorem? Why or why not?