Quantum Information Theory

Nilanjana Datta (nd255@cam)

December 22, 2008

Course notes are available online at http://cam.qubit.org/lectures/qitheory.php.

There are three lent courses which follow on from this one. The notes online may not cover everything, they are only a guide. There is no one book that covers the course, but the website gives a few recommendations. The lectures will be taught very interactively; if the audience does not ask questions they will themselves be asked.

We shall first cover classical information theory, for around 4 lectures. This was mostly done by Shannon in 1948, and later lead to quantum information theory. We shall only focus on two important topic, and may not give many proofs; if you'd like to see a proof of the result, email the lecturer.

For the rest of this course we shall be covering quantum information theory, which was only really done in the 1990s (although some important results had been found by clever Russians isolated from the rest of the mathematical world). It's an interdisciplinary field, involving elements of mathematics, computer science and physics.

Classical information theory is the mathematical theory of information processing tasks - e.g. compression, storage, and transmission. Quantum information theory is the study of the same tasks approached using QM systems such as photons and electrons. This introduces new effects such as entanglement.

What is information? We just don't know. Shannon answered this by relating information to uncertainty; we shall use an extended example of watching cars of four possible colours (red/blue/white/green) emerging from a tunnel. When we see one come out, we receive information and this reduces our uncertainty.

An information source is something that produces messages - e.g. this flow of cars, and our messages here are the colours of cars. We model this by a sequence of i.i.d. RVs $M_1, M_2, \ldots$, taking values $m \in \mathcal{M}$; here $\mathcal{M} = \{r, b, w, g\}$ or $\{1, 2, 3, 4\}$. This means we have a common probability mask function, $P(M_k = m) = p(m) \forall k$, and e.g. $P(M_1 = 2, M_2 = 4) = p(2)p(4)$.

How can we measure uncertainty (which equivalently gives a measure of information)? A measure of uncertainty in getting outcome m, $u(m)$, is given by $-\log p(m)$, the "surprisal" or self-information of m (logs being always base 2 in this course). This has some desirable properties: if $p(m) \simeq 0, u(m)$ is large - we are very surprised to see an outcome of m. If $p(m) = 1$, $u(m) = 0$. Finally, information uncertainty is addative for independent events: $u(2,1) = -\log p(2,1) = -\log(p(2)p(1)) = -\log p(2) - \log p(1) = u(2) + u(1)$.

The average surprisal is entropy: for M a random variable with p.m.f. $p(m)$, the entropy $H(M) = -\sum_{m \in \mathcal{M}} p(m) \log p(m)$ (setting $0 \log 0 = 0$ e.g. by a

continuity argument). We have $H(M) = H(M_1) = H(M_2) = \ldots$; this is called the Shannon entropy of the source.

We shall use the terms information, data, signals, and messages to mean the same thing.

Shannon asked two simple questions: 1) What is the limit to which information can be compressed reliably? 2) What is the maximum rate at which information can be transmitted reliably? The answers to these are given by Shannon's noiseless channel coding theorem and Shannon's noisy channel coding theorem, respetctively.

For the first question, the simplest example of an information source is an IID source. This is characterised by a probability distribution $\{p(u)\}, u \in J$. We consider the case where J is a finite set, called an alphabet; on each use of the source, the letter u is emitted with probability $p(u)$, and this emission of each letter is independent. The message is a sequence of letters $u_1 u_2 \ldots u_n$; we model this by a set of IID RVs $U_1, \ldots, U_n$. $P(U_k = u) = p(u) \forall k$; $P(U_1 = u_1, \ldots, U_n = u_n) = p(u_1) \ldots p(u_n)$. Shannon's answer is that the data compression limit is $H(U) = -\sum_{u \in J} p(u) \log p(u)$; $H(U) = H(U_1) = \ldots$. Note that H is a functional, and completely independent of the values u taken by U; it only depends on the probabilities $p(u)$. So we sometimes instead write $H(\{p(u)\})$.

Example: $J = \{0, 1\}$; we write $d = |J|$. Say U takes values $u = 0$ with probability $\mathfrak{p}$, 1 with probability $1 = \mathfrak{p}$. Then $H(U) = -\mathfrak{p} \log \mathfrak{p} - (1 - \mathfrak{p}) \log(1 - \mathfrak{p})$. This is the Binary Entropy $H(\mathfrak{p})$. Shannon entropy is measured in bits.

Why is compression possible at all? Redundancy - some letters occur more often than others. There are two possible approaches: variable length coding and fixed length coding. In variable length coding, we assign fewer bits to more frequent letters and more bits to less frequent ones - perhaps we would code e as 0 and z as 1010. In fixed length coding, we consider the set of all binary sequences of length m, and the set of all possible sequences in J^n; we map typical (i.e. more frequent) sequences in a 1:1 fashion, and map all the atypical sequences to a single point in the first set. Of course this is not perfectly decodable for finite n and m, only in the limit as $n \to \infty$.

For variable length coding, suppose we have horses labelled $1, \ldots, 8$, and want to send a sequence of these (each labels the winner of a race, e.g. 8311227) Without compression we need 3 bits [per result] to store this data, since 2^n different messages can be stored in n bits: $c(1) = 000, c(2) = 001, \ldots, c(8) = 111$. Now suppose the probabilities $p(i)$ of the ith horse winning were e.g. $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}$. Then consider $c(1) = 0, c(2) = 10, c(3) = 110, c(4) = 1110, c(5) = 111100, c(6) = 111101, c(7) = 111110, c(8) = 111111$. Let l_i be the length of the ith codeword, so e.g. $l_8 = 6$; then we have the average length of codeword is $\sum_{i=1}^{8} p_i l_i = 2$ bits; thus we see compression is possible. Suppose we were sent results of e.g. 10011001110; then the winners are $2, 1, 3, 1, 4$; we can decode this unambiguously because the code is prefix-free, no codeword is a prefix of another.

Fixed Length Data Compression

Assume the information source is IID; messages are sequences $(u_1 \ldots u_n)$; the source emits $u \in J$ with probability $p(u)$. We model the message as always by a sequence of IID RVs $U_1, U_2, \ldots$.

For a compression-decompression scheme of rate R, we have $C^{(n)} : \mathbf{u}^{(n)} = (u_1 \dots u_n) \mapsto$ a codeword $(x_1 \dots x_{\lceil nR \rceil}) = \mathbf{x}$. We have the $u_i \in J, x_i \in \{0, 1\}$, and $|J| = d$. The size of $\{\mathbf{u}^{(n)}\}$ is $d^n = 2^{\log d^n} = 2^{n \log d}$, so $n \log d$ bits are needed to store this. $x_1 \dots x_{\lceil nR \rceil}$ uses $\lceil nR \rceil$ bits, so if $\lceil nR \rceil < n \log d$ then we have compression. For decompression we have $\mathcal{D}^{(n)} : \mathbf{x} \mapsto \mathbf{u}'^{(n)} = (u'_1 \dots u'_n) \in J^n$; an error occurs if $\mathbf{u}'^{(n)} \neq \mathbf{u}^{(n)}$. The average probability of an error is $p_{\text{av}}^{(n)} = \sum_{\mathbf{u}^{(n)} \in J^n} p(\mathbf{u}^{(n)}) P(\mathcal{D}^{(n)}(\mathcal{C}^{(n)}(\mathbf{u}^{(n)})) \neq \mathbf{u}^{(n)})$. We say this compression-decompression scheme is reliable if $p_{\text{av}}^{(n)} \to 0$ as $n \to \infty$.

The notion of typical sequences is important for the quantum case; our result will be that $R = H(U)$.

Definition: For $\{U_i\}$ IID as usual, for any $\epsilon > 0$, the typical set $T_\epsilon^{(n)}$ is the set of sequences $(u_1 \dots u_n) \in J^n$ whose probability $p(u_1 \dots u_n)$ has $2^{-n(H(U)+\epsilon)} \leq p(u_1 \dots u_n) \leq 2^{-n(H(u)-\epsilon)}$. So the typical sequences are approximately equiprobable, with probability $\simeq 2^{-nH(u)}$.

$|T_\epsilon^{(n)}|$ is the total number of typical sequences, $P\{T_\epsilon^{(n)}\}$ is the probability of a typical sequence.

Now, a justification for this notion of "typical": consider a sequence $\mathbf{u}^{(n)} = (u_1 \dots u_n)$; this is typical if $\frac{\#\text{ times } u \text{ appears in } \mathbf{u}^{(n)}}{n} \simeq p(u)$. So the sequence of length n is typical if the number of copies of u in $u_1 \dots u_n$ is approximately $np(u)$. The probability of such a sequence is $\prod_{u \in J} p(u)^{np(u)} = \prod_{u \in J} 2^{\log p(u)^{np(u)}} = \prod_{u \in J} 2^{np(u) \log p(u)} = 2^{2 \sum p(u) \log p(u)} = 2^{-nH(u)}$.

Typical Sequence Theorem: Fix $\epsilon > 0$. Then $\forall \delta > 0 \exists n_0(\delta) > 0$ such that $\forall n \geq n_0(\delta)$: 1) If $(u_1 \dots u_n) \in T_\epsilon^{(n)}$, $H(U) - \epsilon \leq -\frac{1}{2} \log P(u_1 \dots u_n) \leq H(U) + \epsilon$ (this is just the log of the above). 2) $P\{T_\epsilon^{(n)}\} \geq 1 - \delta$ (without proof, by some law of large numbers) 3) $|T_\epsilon^{(n)}| \leq 2^{n(H(u)+\epsilon)}$ (proof below) 4) $|T_\epsilon^{(n)}| \geq (1 - \delta) 2^{n(H(U)-\epsilon)}$ - exercise.

For the proof of 3, $p(u_1 \dots u_n) \geq 2^{-n(H(U)+\epsilon)}$. So we have $|T_\epsilon^{(n)}| \times 2^{-n(H(U)+\epsilon)} \leq \sum_{(u_1 \dots u_n) \in T_\epsilon^{(n)}} p(u_1 \dots u_n)$, but this is a probability so ≤ 1, so $|T_\epsilon^{(n)}| \leq 2^{n(H(u)+\epsilon)}$.

(Physical) consequences of the typical sequence theorem: we write $J^n = T_\epsilon^{(n)} \sqcup \Pi_\epsilon^{(n)}$. We have $P\{\Pi_\epsilon^{(n)}\} < \delta$, i.e. atypical sequences rarely occur. Recall typical sequences are almost equiprobable, with probability $\simeq 2^{-nH(U)}$.

Shannon's Noiseless Channel Coding Theorem: under our usual assumptions, if $R > H(U)$ there is a reliable compression scheme of rate R, and if $R < H(U)$ then there is no reliable compression scheme of rate R. (Note that in a compression-decompression scheme of rate R we assign unique codewords to $2^{\lceil nR \rceil}$ messages).

Typical Sequence Theorem

Fix $\epsilon > 0$; for any $\delta > 0$ and n large enough, $|T_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)}$, $P(\{T_\epsilon^{(n)}\}) > 1 - \delta, P\{\Pi_\epsilon^{(n)}\} \leq \delta$. All typical sequences are equiprobable, $p(\mathbf{u}^{(n)}) \simeq 2^{-nH(U)}$.

Shannon's theorem: the optimal rate $R = H(U)$. 1) If $R > H(U)$ then there is a reliable compression scheme of rate R, 2) if $R < H(U)$ then there is no such scheme.

Recall that in a compression scheme of rate R, we assign unique codewords to 2^{nR} messages.

Proof: 1) Given $R > H(U)$, choose $\epsilon > 0$ such that $R > H(U) + \epsilon$. Choose n large enough that TST holds; $|T_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)} < 2^{nR}$.

For compression, first order the elements of $T_\epsilon^{(n)}$. Then, we can represent each typical sequence by its index/label/number, using at most $\lceil nR \rceil$ bits. When we get output from the source, examine it; if it lies in $T_\epsilon^{(n)}$, store its index (requiring at most $\lceil nR \rceil$ bits); otherwise assign a fixed codeword e.g. $\lceil nR \rceil$ zeroes (yes, this may also be the codeword for some typical sequence; this is fine).

2) If you assign codewords to $< 2^{nH(U)}$ messages (i.e. $R < H(U)$) then the compression scheme is never reliable: consider any set of codewords S_n with $|S_n| \leq 2^{nR}$. We want $P\{S_n\} < \delta \forall \delta > 0$, then we're done. We know $P(T_\epsilon^{(n)}) \geq 1 - \delta$.

Lemma: For $S^n \subset T^n$, $|S_n| \leq 2^{nR}$, $R < H(U)$ fixedso this is $< 2^{nH(U)}$, then $P\{S_n\} = \sum_{\mathbf{u}^{(n)} \in S_n} p(\mathbf{u}^{(n)}) < \delta$. This will do us: it implies that with a high probability the source will emit (typical) sequences which do not lie in S_n.

Proof: $\sum_{\mathbf{u}^{(n)} \in S_n} p(\mathbf{u}^{(n)}) = \sum_{\mathbf{u}^{(n)} \in S_n \cap T_\epsilon^{(n)}} + \sum_{\mathbf{u}^{(n)} \in S_n, \notin T_\epsilon^{(n)}}$. The lecond term is $\leq \widehat{\delta}$ (since $|\Pi_\epsilon^{(n)}| \leq \widehat{\delta}$); each typical sequence has $p(\mathbf{u}^{(n)}) \simeq 2^{-nH(U)}$ so the first term is $\lesssim 2^{-nH(U)}|S_n| \leq 2^{-nH(U)}2^{nR} = 2^{-n(H(U)-R)} \to 0$ as $n \to \infty$ (No, this isn't currently rigorous. Yes, it can be made so. No, we're not going to. Email the lecturer if you really care). So the first term can be made $\leq \delta'$ and the sum is $\leq \delta' + \hat{\delta} = \delta$, so $P\{S_n\} \leq \delta$.

Entropy of pairs of RVs

Say we have X, Y taking values in J_X, J_Y according to $X \sim p(x), Y \sim p(Y)$. Then we have the joint probability $P(X = x, Y = y) = p(x, y)$ and conditional probability $P(Y = y \mid X = x) = p(y|x)$. So we can define joint entropy $H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y)$ and conditional entropy $H(Y \mid X) = \sum_x p(x) H(Y \mid X = x) = -\sum_x p(x) \sum_y p(y \mid x) \log p(y \mid x) = -\sum_{x,y} p(x) p(x \mid x) \log p(y \mid x) = -\sum_{x,y} p(x, y) \log p(y \mid x)$. An exercise is to prove the chain rule that $H(X, Y) = H(Y \mid X) + H(X)$.

Relative entropy is a measure of "distance" between two probability distributions (although not a valid metric): if $p = \{p(x)\}, q = \{q(x)\}$ then $D(p||q) = \sum_{x \in J} p(x) \log \frac{p(x)}{q(x)}$ (where we define $0 \log \frac{0}{q} = 0, p \log \frac{p}{0} = \infty$. We will prove that $D(p||q) \geq 0$.

Set $M = \{x : p(x) > 0\}$. Then $D(p||q) = \sum_{x \in M} p(x) \log \frac{p(x)}{q(x)}$, so $-D(p||q) = \sum_{x \in M} p(x) \log \frac{q(x)}{p(x)}$. We can write this as $\mathbb{E}(\log X)$, where X is a random variable taking values $\frac{q(x)}{p(x)}$ with probability $p(x)$; then apply Jensen's inequality; log is concave, so $\mathbb{E}(\log X) \leq \log(\mathbb{E}(X)) = \log(\sum_{x \in M} p(x) \frac{q(x)}{p(x)}) = \log(\sum_{x \in M} q(x)) \leq \log(\sum_{x \in J} q(x)) = \log(1) = 0$, so $-D(p||q) \leq 0$ and $D(p||q) \geq 0$ as required. This result goes by various names, but is most commonly known as Klein's inequality.

Recall $H(X) \geq 0$: $H(X) = -\sum p(x) \log p(x)$ and $0 \leq p(x) \leq 1$. From this we can prove $H(X) \leq \log |J|$: consider $p = \{p(x)\}, q = \{\frac{1}{|J|}\}$. Then $D(p||q) \geq 0$: $\sum_x p(x) \log(|J| p(x)) \geq 0$, so $\sum p(x) \log |J| + \sum p(x) \log p(x) \geq 0$, i.e. $\log |J| - H(X) \geq 0$, so $H(X) \leq \log |J|$.

Mutual Information

$H(X : Y)$ is defined as $\sum p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$. This is a measure of how much information one random variable gives us about the other; if $p(x,y) = p(x)p(y)$ (independent variables) it becomes 0. As an exercise, prove this is $= H(X) + H(Y) - H(X,Y)$, and also $= H(X) - H(X \mid Y)$ or $= H(Y) - H(Y \mid X)$ (consider $p = \{p(x,y)\}, q = \{p(x)p(y)\}$).

How is information-theoretic entropy related to the entropy seen in thermodynamics or statistical mechanics? von Neumann suggested the name to Shannon, since the functions are actually the same: recall we will look for the function $H(U) = -\sum p(u) \log p(u)$.

Statistical mechanics relates macrostates e.g. temperature or pressure of a whole gas to microstates e.g. momentum or velocity of individual particles.

Definition: If there are Ω microstates corresponding to a particular macrostate then the dynamic entropy $S = k \log_e \Omega$ where k is the Boltzmann constant. Suppose the microstate r occurs with probability p_r. Consider a set of v replicas or copies of the [micro] system, called an "ensemble". Let v_r denote the number of replicas in the rth microstate, $v_r \simeq vp_r$. Then $\Omega = \binom{v}{v_1 v_2 \dots v_k}$ [the multinomial coefficient] $= \frac{v!}{v_1! v_2! \dots v_k!}$ (we have $\sum_{r=1}^{k} v_r = v$. By Stirling's this is $\simeq \frac{v^v}{v_1^{v_1} \dots v_k^{v_k}}$. The entropy of the ensemble is $S_v = k \log \Omega$ by definition; this is $k(v \log_e v - \sum_r v_r \log v_r)$ [taking log to be always base e just in this paragraph] $= k(-\sum_r (\log v_r - \log v)) = k(-\sum_r v_r \log \frac{v_r}{v}) = -kv \sum_r p_r \log p_r$. The entropy of a compound system is the sum of its parts, $S_v = vS$ where $S = \frac{S_v}{v} = -\sum_r p_r \log p_r$.

Typical Sequence Theorem

$T_\epsilon^{(n)}$: $\forall \delta > 0 \forall n$ large enough, $P\{T_\epsilon^{(n)}\} \geq 1 - \delta$, $|T_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)}$, and $|T_\epsilon^{(n)}| \geq (1 - \delta) 2^{n(H(U)-\epsilon)}$. The second of these follows from the Asymptotic Equipartition Property: for $U_1, \dots U_n$ IID RVs and our usual scenario, $\lim_{n\to\infty} \mathbb{P}(2^{-n(H(U)+\epsilon)} \leq p(U_1, \dots, U_n) \leq 2^{-n(H(U)-\epsilon)}) = 1$ (It is usually written in a different form, but clearly has this meaning). Here $\mathbb{P}(U_1, U_2, \dots U_n)$ is a random variable taking values $p(u_1, \dots, u_n)$.

This means that for any $\delta > 0$, for n large enough, $\mathbb{P}(2^{-n(H(U)+\epsilon)} \leq p(U_1, \dots, U_n) \leq 2^{-n(H(U)-\epsilon)}) \geq 1 - \delta$, i.e. the set of sequences $(u_1, \dots, u_n)$ for which $2^{-n(H(U)+\epsilon)} \leq p(u_1, \dots, u_n) \leq 2^{-n(H(U)-\epsilon)}$ has a probability $\geq 1 - \delta$. But this set is the typical set, so $P\{T_\epsilon^{(n)}\} \geq 1 - \delta$ for n large enough.

Shannon's second question: What is the maximum rate at which information can be transmitted through a channel? A channel takes input $X^{(N)}$, but is noisy; its output is $Y^{(N)}$. We will consider discrete channels: the input alphabet is discrete, $X^{(N)} = (X_1, \dots, X_N)$ taking values $\mathbf{x}^{(N)} = (x_1, \dots, x_n)$ for $x_k \in J_X$; we usually consider $J_X = \{0, 1\}$. Similarly the output $Y^{(N)} = (Y_1, \dots, Y_N)$ taking values $\mathbf{y}^{(N)} = y_1 \dots y_n \in J_Y$. We may have $J_Y \neq J_X$, e.g. $J_y = \{0, 1, \star\}$ where $\star$ denotes useless junk output.

We model our channel by an operation $\mathcal{N}$ categorized by a set of conditional probabilities $\{p(\mathbf{y}^{(N)} | \mathbf{x}^{(N)})\}$. If $\mathbf{y}^{(N)} \neq \mathbf{x}^{(N)}$, we have an error - cf noise. Shannon proved that it is possible to choose a non-confusable set of input sequences, such that there is only one highly likely input corresponding to a given output. We assume the channel is memoryless: $p(\mathbf{y}^{(N)} | \mathbf{x}^{(N)}) = \prod_{i=1}^{N} p(y_i | x_i)$.

Classical Communication System

Suppose our sender A wants to send messages $M \in \mathcal{M}$ to a receiver B through a noisy channel. So we encode M as some $\mathbf{x}^{(N)}$, and send that through the channel, characterised by $p(\mathbf{y}^{(N)}|\mathbf{x}^{(N)})$; $\mathbf{y}^{(N)}$ is then decoded as M'.

The simplest possible channel is a memoryless binary symmetric channel; given 0 as input, it outputs 1 with probability p, 0 with probability $1-p$, and vice versa when given 1 as input.

The idea is to add redundancy when encoding, e.g. repitition: encode $M = 0$ as $\mathbf{x}^{(N)} = 000$, then if $\mathcal{N}$ maps this to 010, Bob decodes it [correctly] by "majority vote" as 0. This method of course fails when ≥ 2 bits are flipped.

Note the duality between data compression and transmission - in one we use redundancy to reduce the size of the message, in the other we add redundancy.

Capacity of the Channel

For a memoryless channel, $C = \max_{p(x)} I(X : Y)$, where the maximum is taken over all possible probability distributions $\{p(x)\}$.

We will give an intuitive way of looking at Shannon's noisy coding theorem, rather than an actual proof. We have $|\mathcal{M}|$ = the number of messages, $\log|\mathcal{M}|$ is the number of bits of message. For encoding, $C^N : \mathcal{M} \to J_X^N = \{0,1\}^N$; noise gives us $\mathbf{x}^{(N)} \to \mathbf{y}^{(N)} \in J_Y^{(N)}$, and we shall take $J_Y^{(N)} = J_X^{(N)}$ for simplicity. Then decoding is $\mathcal{D}^N : J_Y^{(N)} \to \mathcal{M}$. The maximum probability of an error is $p_{\max}(C^N, \mathcal{D}^N) = \max_{M \in \mathcal{M}} P(M' \neq M)$ (where of course $M' = \mathcal{D}^N(C^N(M))$). The rate $R = \frac{\log|\mathcal{M}|}{N}$ - it is the number of bits of message transmitted per bit of codeword. A rate R is achievable if there exist $(C^N, \mathcal{D}^N)$ [of rate R] such that $p_{\max}(C^N, \mathcal{D}^N) \to 0$ as $N \to \infty$. The capacity C is defined (the expression at the start of this section is the result we shall prove) to be $\sup R$, where the supremum is taken over all achievable rates.

Shannon's Noisy Channel Coding Theorem

For a memoryless channel characterised by $p(x \mid y)$, $C = \max_{p(x)} I(X : Y)$ [where the maximum is taken over all possible probability distributions for x] $= H(X : Y) = H(X) + H(Y) - H(X, Y) = H(Y) - H(Y \mid X)$.

For each (typical) input sequence $\mathbf{x}^{(N)}$ there will be $2^{NH(Y|X)}$ typical output sequences: we saw that for $U_1, \ldots, U_n$ with $U \sim p(u)$, there are $2^{nH(U)}$ typical sequences; here the probability are $p(y \mid x)$ rather than $p(u)$, so this becomes $2^{NH(Y|X)}$. Recall $H(Y \mid X) = \sum p(x, y) \log p(x \mid y)$.

We want to avoid having the sets of typical outputs for different (typical) inputs overlap. $|T_Y^N| \simeq 2^{NH(Y)}$, so for unique inference we divide T_Y^N into disjoint subsets of size $2^{NH(Y|X)}$. The number of such disjoint sets is $\frac{|T_Y^N|}{2^{NH(Y|X)}} = 2^{N(H(Y)-H(Y|X))} \simeq 2^{NI(X:Y)}$, and this is the maximum number of input sequences i.e. messages we may send. So $R = \frac{\log|\mathcal{M}|}{N} = \log 2^{NI(X:Y)} \therefore C = \max_{p(x)} I(X : Y)$.

Example: memoryless binary symmetric channel (MBSC) such that $p(y \mid x) = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$. This is symmetric, by which we simply mean that the rows are permutations of each other. $I(X : Y) = H(Y) - H(Y \mid X) = H(Y) -$

$\sum p(x)H(Y \mid X = x)$. For a symmetric channel matrix $H(Y \mid X = x) = -\sum_y p(y \mid x) \log p(y \mid x)$ [is independent of x]; in this case it is $-((1-p)\log(1-p) + p \log p) = h(p)$. So $C = \max_{p(x)}(H(Y) - H(Y \mid X)) = \max_{p(x)}(H(Y) - h(p)) = \max_{p(x)} H(Y) - h(p)$. So to maximise $I(X : Y)$ we only need to maximise $H(Y)$; we have $H(Y) \leq \log |J_Y|$, with equality when Y is evenly distributed. If we can achieve this then we have $C = 1 - h(p)$, and in this case this is indeed achieved by $p(x = 0) = \frac{1}{2} = p(x = 1)$.

Quantum Mechanics

We need to consider open systems - these are not isolated, interactions with the world are unavoidable. These disturb the state and so distort the information - decoherence. So we need to understand the physics of open systems. First we will review the postulates for closed systems, then see how they change for open systems.

Postulate 1: the state of a system is a ray in a Hilbert space $\mathcal{H}$. We use the Dirac bra-ket notation: if $|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{C}^2$, then $\langle\psi| = (a^\star b^\star)$, the complex conjugate of the transpose. States are rays rather than vectors since global phase is undetectable - ψ and $e^{i\alpha}\psi$ represent the same physical state. (Of course relative phase factors are important - e.g. the states $\psi = a\psi_1 + e^{i\phi}b\psi_2$ are different for different ϕ). In classical information theory, the fundamental unit is a bit, taking values 0 or 1. In quantum information theory, we use qubits: $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. This is very different; where a classical bit is only 1 or 0, a qubit may be in any state $|\psi\rangle = a|0\rangle + b|1\rangle$ with $a, b \in \mathbb{C}, |a|^2 + |b|^2 = 1$; this is called the superposition principle. (The condition on a, b comes because we want the state to be normalized $\langle\psi|\psi\rangle = 1$; since for this state $\langle\psi| = a^\star\langle 0| + b^\star\langle 1|$, this means $(a^\star b^\star)\begin{pmatrix} a \\ b \end{pmatrix} = 1$ whence the condition). Notice $0 \leq |a|^2, |b|^2 \leq 1$, so we can see $|a|^2, |b|^2$ as probabilities; we shall see more of this later.

Notice that $|\psi\rangle$ is not, and shold not be confused with, a statistical mixture of states (e.g. if we had state $|\psi_1\rangle$ with probability p_1, $|\psi_2\rangle$ with probability p_2).

Postulate 2: time evolution is unitary, governed by the Hamiltonian and the Schrodinger equation.

Postulate 3: measurements correspond to orthogonal projections. We can measure only a self-adjoint operator $A = A^\dagger$; the outcome $A|\phi_j\rangle = a_j|\phi_j\rangle$ is some eigenvalue $\in a_1, \ldots, a_n$. What is the probability of the otcome a_j? It is $\langle\psi|P_j|\psi\rangle$, where P_j is the projection operator onto the eigenspace of A corresponding to a_j; we assume a_j non-degenerate and this is $P_j = |\phi_j\rangle\langle\phi_j|$.

Suppose our outcome is a_j. If before the measurement the system was in a state $|\psi\rangle$, due to the measurement $|\psi\rangle \to \frac{P_j|\psi\rangle}{\sqrt{\langle\psi|P_j|\psi\rangle}}$. Measurement causes a "collapse" - thus it really is a projection. And so it's "repeatable" - if we measure [immediately] again, we get the same result. This postulate is also known as "projective measurement" or "von Neumann measurement".

Consider $|\psi\rangle = a|0\rangle + b|1\rangle$. $|0\rangle, |1\rangle$ are two reliably distinguishable states, i.e. we can devise a measurement which can distinguish between them. E.g. suppose $|0\rangle$ represents horizontally polarized light, $|1\rangle$ vertically polarized light;

then $|\psi\rangle = \frac{|0\rangle+i|1\rangle}{\sqrt{2}}$ (i.e. $a = \frac{1}{\sqrt{2}}, b = \frac{i}{\sqrt{2}}$) is the state for right [I think] circularly polarized light. The only reliably distinguishable states are those which are orthogonal - here $\langle 0|1\rangle = 0$ so $|0\rangle, |1\rangle$ are distinguishable, but $\langle 0|\psi\rangle \neq 0$ so $|0\rangle, |\psi\rangle$ are not.

We aim to distinguish between $|0\rangle, |1\rangle$ without destroying either such state. We measure $A = |1\rangle\langle 1|$; we have $A = A^\dagger, A^2 = |1\rangle\langle 1|1\rangle\langle 1| = |1\rangle\langle 1| = A$, so this is a projection operator. Then for an eigenstate $P|e\rangle = \lambda e\rangle \Rightarrow P^2|e\rangle = \lambda P|e\rangle \therefore P|e\rangle = \lambda^2|e\rangle \Rightarrow \lambda = \lambda^2$ so $\lambda = 1$ or 0.

Measuring A gives outcome 0 or 1; there will be projection operators corresponding to these outcomes. What is $P(0)$? It is $|0\rangle\langle 0|, P(1) = |1\rangle\langle 1|$.

Digression: if we have an operator $Q = |\psi\rangle\langle\psi|$ with eigenvalues $0, 1$, $P(1) = Q, P(0) = I - Q$.

Mesaurement

Say the state before measurement is $|\psi\rangle$.

1. $|\psi\rangle = |0\rangle, A = |1\rangle\langle 1|$. Probability of the result 1 is $\langle\psi|P(1)|\psi\rangle = \langle 0|1\rangle\langle 1|0\rangle = 0$. Originally $|\psi\rangle = |0\rangle$; this becomes $\frac{P_0|\psi\rangle}{\sqrt{\langle\psi|P_0|\psi\rangle}} = \frac{|0\rangle\langle 0|0\rangle}{1} = |0\rangle$.

2. $|\psi\rangle = |1\rangle$. $p(1) = 1, p(0) = 0$; $|\psi\rangle = |1\rangle$ becomes $|1\rangle$ (exercise).

3. $|\psi\rangle = a|0\rangle + b|1\rangle$ (exercise). $p(1) = |b|^2, p(0) = |a|^2$. Note that our state vectors are always normalized.

Open Systems

It's useful to treat an open system as a composite system consisting of the principal system S which we are interested in, surrounded by the environment system E. The overall state space is then $\mathcal{H}_S \otimes \mathcal{H}_E$.

Density Matrix Formalism

For our open system $\mathcal{S}$, consider a statistical mixture of states $|\psi_1\rangle, |\psi_2\rangle, \ldots$ with probabilities $p_1, p_2, \ldots$. The state of the subsystem is given by the ensemble of state vectors $\{p_i, |\psi_i\rangle\}_{i=1}^k$. $p_i \geq 0, \sum_{i=1}^k p_i = 1$.

The density operator or matrix is $\rho = \sum_{i=1}^k p_i|\psi_i\rangle\langle\psi_i|$. But note these $|\psi_i\rangle$ need not be mutually orthogonal; we don't require $\langle\psi_j|\psi_i\rangle = \delta_{ij}$.

Properties: $\rho \geq 0$ (i.e. it is a positive semidefinite operator); this in turn implies $\rho = \rho^\dagger$. Also $\mathrm{tr}\rho = 1$ (Both these are exercises).

Positive semidefiniteness means $\forall|\phi\rangle \in \mathcal{H}, \langle\phi|\rho|\phi\rangle \geq 0$: $\langle\phi|\rho|\phi\rangle = \sum_i p_i\langle\phi|\psi_i\rangle\langle\psi_i|\phi\rangle = \sum p_i|\langle\phi|\psi_i\rangle|^2 \geq 0$.

For any self-adjoint operator ρ, we can write this as $\rho = \sum \lambda_j|e_j\rangle\langle e_j|$.

Lemma: An operator ρ is the density matrix for an ensemble of states iff 1) $\mathrm{tr}\rho = 1$ and 2) $\rho \geq 0$. If we assume ρ is such a matrix, corresponding to $\{p_i, |\psi_i\rangle\}_{i=1}^k$ (note we may not assume $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, for 1) $\rho = \sum_{i=1}^k p_i|\psi_i\rangle\langle\psi_i|$, so $\mathrm{tr}\rho = \sum p_i\mathrm{tr}|\psi_i\rangle\langle\psi_i|$; by taking a basis which contains $|\psi_i\rangle$ separately while calculating each trace, $\mathrm{tr}|\psi_i\rangle\langle\psi_i| = 1$ in each case, so $\mathrm{tr}\rho = \sum_i p_i = 1$. For any $|\phi\rangle \in \mathcal{H}$, $\langle\phi|\rho|\phi\rangle = \sum_i p_i\langle\phi|\psi_i\rangle\langle\psi_i|\phi\rangle = \sum p_i|\langle\phi|\psi_i\rangle|^2 \geq 0$. For the converse, assume ρ is an operator with $\mathrm{tr}\rho = 1, \rho \geq 0$. Then ρ corresponds to an ensemble of states: $\rho \geq 0 \Rightarrow \rho^\dagger = \rho \Rightarrow$ we have a spectral decomposition $\rho = \sum_{j=1}^d \lambda_j|e_j\rangle\langle e_j|$ where

$d = \dim\mathcal{H}, \langle e_j|e_i\rangle = \delta_{ij}, \lambda_j \geq 0 \forall j = 1,\ldots,d.$ $\mathrm{tr}\rho = 1 \Rightarrow \sum_{j=1}^{d}\lambda_j = 1 \therefore \{\lambda_j\}_{j=1}^{d}$ is a probability distribution. So associate ρ with the ensemble $\{\lambda_j, |e_j\rangle\}$. (Note that the stricter conditions (e.g. orthogonality) mean it may occasionally be worth performing this decomposition even for an operator we already know is a density matrix)

Pure and Mixed States

A pure state would be e.g. $|\psi_2\rangle : p_2 = 1, p_i = 0 \forall i \neq 2$. Then $\rho = |\psi_2\rangle\langle\psi_2| \Rightarrow \rho^2 = \rho$, the crucial property. So $\mathrm{tr}\rho^2 = \mathrm{tr}\rho = 1$; this is a way of telling whether a given ρ is a pure state. For a mixed state $\rho = \sum_j p_j|\psi\rangle\langle\psi_j|$, the reader may check $\mathrm{tr}\rho^2 < 1$. Note that we shall use "pure state" to talk about both states $|\psi\rangle$ and operators $\rho = |\psi\rangle\langle\psi|$.

For an operator A, the expectation value in the state ρ is $\langle A\rangle \equiv \langle A\rangle_\rho := \mathrm{tr}(A\rho)$. This is linear: $\langle aA + bB\rangle = a\mathrm{tr}\rho A + b\mathrm{tr}\rho B = a\langle A\rangle + b\langle B\rangle$. $\langle A\rangle \geq 0 \forall A \geq 0$; also it is normalized: $\langle I\rangle = \mathrm{tr}(\rho I) = \mathrm{tr}\rho = 1$.

Applications of the Density Matrix Formalism

This formalism allows us to describe the properties of subsystems of composite systems. Given a composite system AB, the state space is $\mathcal{A}\otimes\mathcal{H}_B$. Suppose the system is in the pure state $|\psi_{AB}\rangle$; $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$. What is the state of A? It is $\rho_A = \mathrm{tr}_B\rho_{AB}$, where tr_B is the partial trace, the trace taken only over $\mathcal{H}_B$. This ρ_A is called the reduced density matrix of A.

Say we have $\{|i_A\rangle\}$ an orthonormal basis of $\mathcal{H}_A$, $\{|\alpha_B\rangle\}$ an ONB of $\mathcal{H}_B$. Thes $\{|i_A\rangle\otimes|i_B\rangle\}$ is an ONB of $\mathcal{H}_A\otimes\mathcal{H}_B$. If we write $|\psi_{AB}\rangle = \sum_{i\alpha} a_{i\alpha}|i_A\rangle\otimes|\alpha_B\rangle$, then $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| = \sum_{i,\alpha,i',\alpha'} a_{i\alpha}a^\star_{i'\alpha'}(|i_A\rangle\otimes|\alpha_B\rangle)(\langle i'_A|\otimes\langle\alpha'_B|) = \sum a_{i\alpha}a^\star_{i'\alpha'}|i_A\rangle\langle i'_A|\otimes|\alpha_B\rangle\langle\alpha'_B|$. Then $\rho_A = \mathrm{tr}_B\rho_{AB}$ is defined as $\sum_{i,\alpha,i',\alpha'} a_{i\alpha}a^\star_{i'\alpha'}|i_A\rangle\langle i'_A|\mathrm{tr}(|\alpha_B\rangle\langle\alpha'_B|)$. $\mathrm{tr}(|\alpha\rangle\langle\alpha'|) = \sum_\beta\langle\beta|\alpha\rangle\langle\alpha'|\beta\rangle = \delta_{\alpha\beta}\delta_{\alpha'\beta} = \delta_{\alpha\alpha'}$, so this is $\sum_{ii'\alpha} a_{i\alpha}a^\star_{i'\alpha}|i\rangle\langle i'|$.

We have $\mathrm{tr}\rho_A = 1$, and $\rho_A \geq 0$ (this is easy to prove with the Schmidt decomposition - see later), $\rho_A^\dagger = \rho_A$. For example, $\mathrm{tr}\rho_A = \sum_{ii'\alpha} a_{i\alpha}a^\star_{i'\alpha}\delta_{ii'} = \sum_{i\alpha}|a_{i\alpha}|^2 = 1$ since $\langle\psi|\psi\rangle = 1$, $\rho_A^\dagger = \sum a^\star_{i\alpha}a_{i'\alpha}|i'\rangle\langle i| = \rho_A$.

Suppose an operator $M_{AB} = M_A\otimes I_B$; the system is in the state $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$. Then $\langle M_{AB}\rangle_{\rho_{AB}} = \mathrm{tr}(M_{AB}\rho_{AB}) = \mathrm{tr}(M_A\rho_A)$; this is an exercise (use the expansion of $|\psi_{AB}\rangle$).

Clearly $\rho_B := \mathrm{tr}_A\rho_{AB}$. For a general (mixed) $\rho_{AB} = \sum\lambda_i|\psi_i^{AB}\rangle\langle\psi_i^{AB}| \in \mathcal{H}_A\otimes\mathcal{H}_B$, $|\psi_i^{AB}\rangle = \sum_{\alpha j} a_{\alpha j}|\alpha\rangle_A\otimes|j\rangle_B$.

Examples: 1) AB = two qubits, $\rho_{AB} = \rho_1\otimes\rho_2$. The reduced density matrix of A is $\rho_A = \mathrm{tr}_B\rho_{AB} = \rho_1$, $\rho_B = \mathrm{tr}_A\rho_{AB} = \rho_2$.s 2) AB = two qubit, pure state $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ where $|\psi_{AB}\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ (where $|00\rangle = |0\rangle_A\otimes|0\rangle_B$ etc.) $\rho_{AB} = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) = \frac{1}{2}(|0\rangle\langle 0|_A\otimes|0\rangle\langle 0|_B + |1\rangle\langle 0|_A\otimes|1\rangle\langle 0|_B + \ldots)$, $\rho_A = \mathrm{tr}_B\rho_{AB} = \frac{1}{2}(|0\rangle\langle 0|_A + |1\rangle\langle 1|_A) = \frac{1}{2}I$. This is entanglement - we knew precisely the state $|\psi\rangle$, but this tells us nothing about the state of qubit A. $|\psi\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ is an entangled state.

We're considering bipartite state, i.e. the state of two systems A and B. Such a state will either be separable or entangled; for a pure state, $|\psi_{AB}\rangle$ Is separable if it can be written $|\psi_A\rangle\otimes|\psi_B\rangle$, entangled if it cannot be. A mixed state is separable

if it can be written $\rho_{AB} = \sum_i p_i |\alpha_i\rangle\langle\alpha_i|_A \otimes |\beta_i\rangle\langle\beta_i|_B$ (i.e. if it can be expressed with all of its constituent states separable)

Define $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}, |1\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}, |\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), |\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$. These are the Bell states or EPR states; they form an ONB of $\mathcal{H}_A \otimes \mathcal{H}_B$ (but note they are all entangled).

Consider $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ where $\psi_{AB} = \frac{|00\rangle+|11\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$; recall we can write $\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix}$. We label states by i in A-space and α in B-space. $\rho_{AB} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}(1001) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$. The elements of ρ_{AB} are $(\rho_{AB})_{i\alpha,j\beta}$. $\rho_A = \frac{1}{2}A$: $(\rho_A)_{i,j} = \sum_\alpha \rho^{AB}_{i\alpha,j\alpha}$. We know $\rho_{AB} \geq 0 \therefore (\rho_{AB})_{i\alpha,i\alpha}$ is always ≥ 0 (i.e. the diagonal elements are always ≥ 0) so $(\rho_A)_{i,i} = \sum_\alpha (\rho_{AB})_{i\alpha,i\alpha} \geq 0$, so $\rho_A \geq 0$.

Time Evolution for Open Systems

In a closed system, we know the time evolution is unitary - given by the Schodinger equation. $\psi(t) = U(t)\psi(0) = e^{-\frac{iHt}{\hbar}}\psi(0)$. In an open system, it need not be unitary. We use the Quantum Operation Formalism, which allows us to describe discrete state changes $\Phi : \rho \to \phi'$ without concerning ourselves with how the state behaves in the time between the initial and final state. Such a Φ is called a quantum operator; we call it a superoperator since it maps operators to operators.

Notation: $\mathcal{H}$ is our finite dimensional Hilbert space, $\mathcal{B}(\mathcal{H})$ is the algebra of all operators acting in $\mathcal{H}$. So $\rho \in \mathcal{B}(\mathcal{H})$, $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$, $\mathcal{H} \simeq \mathcal{R}$. The simplest possible Φ is a unitary transformation $\rho \mapsto \rho' = \Phi(\rho) = U\rho U^\dagger$ for some unitary operator U. The reader may check: $\rho' = U\rho U^\dagger$ is also a density matrix, i.e. $\rho' \geq 0, \mathrm{tr}\rho' = 1$. The change of state under any physical process would be given by such a Φ.

We ask: what are the conditions Φ must satisfy to represent an allowed physical process? We shall look from three equivalent perspectives: 1) physically motivated axioms, 2) system coupled to environment, and 3) operator-sum representation (Kraus Representation Theorem).

For the first, suppose we have $\Phi : \rho \to \rho'$ for $\rho, \rho' \in \mathcal{B}(\mathcal{H})$. We require a) linearity: $\Phi(a\rho_1 + b\rho_2) = a\Phi(\rho_1) + b\Phi(\rho_2)$. This is important for the probabilistic interpretation of mixed states: if $\rho = \sum p_i |\psi_i\rangle\langle\psi_i| = \sum p_i \rho_i$ where ρ_i is the pure state $|\psi_i\rangle\langle|$, linearity implies $\Phi(\rho) = \sum ip_i\Phi(\rho_i)$. b) Trace preserving: $\rho \to \rho' = \Phi(\rho)$ with $\mathrm{tr}\rho' = 1 = \mathrm{tr}\rho$. (Yes, this condition is necessary. Exercise: show that for a general linear map $\Phi(\rho)$ not necessarily trace-preserving, the normalized map $\tilde{\rho} = \frac{\Phi(\rho)}{\mathrm{tr}\Phi(\rho}$ will not generally be linear). c) Positivity: $\Phi(\rho) \geq 0$.

These three requirements are not enough, for a reason we shall see shortly. We require the stricter condition of Complete Positivity (CP): $(\Phi_A \otimes \mathrm{id}_B)(\rho_{AB})$

must be a legitimate state, i.e. ≥ 0, for all possible extensions B of the system. (Convention: we use I for the unit operator, id for the unit superoperator). Φ is a CP map on $\mathcal{B}(\mathcal{H}_A)$ if $(\Phi \otimes \mathrm{id}_B)$ is positive for all extensions of $\mathcal{H}_A$. We call the added auxilary system B an ancilla.

Suppose the system is in the state $\rho_A \otimes \sigma_B = \rho_{AB}$. Then $(\Phi \otimes \mathrm{id}_B)(\rho_A \otimes \sigma_B) = \rho'_{AB}$; the final state of A is $\rho'_A = \mathrm{tr}_B \rho'_{AB}$.

Example: there is a map which is positive but not CP: transposition. Let ρ be the DM of a qubit, $\rho = \begin{pmatrix} a & b \\ b^\star & 1-a \end{pmatrix}$. We have $\rho \geq 0$: all eigenvalues are ≥ 0. $\Phi = T$ takes the transpose: $\Phi(\rho) = \rho' = \begin{pmatrix} a & b^\star \\ b & 1-a \end{pmatrix}$. This leaves the characteristic equation of the matrix unchanged, so the eigenvalues are unchanged, and this $\Phi = T$ is a positive map. Consider two qubits: $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ where as before $|\psi_{AB}\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$. $(T = \Phi_A \otimes \mathrm{id}_B)(\rho_{AB}) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, and this matrix has eigenvalues $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}$. This will be a recurring theme in this course: new phenomena usually have their roots in entanglement.

So we call Φ a quantum operation if it is a linear CPT or CPTP map - completely positive trace preserving.

For the second, consider a system A with environment B. The combined system AB is closed: it has unitary time evolution. But $\rho_A \to \Phi(\rho_A)$ is not necessarily unitary.

Assume the initial state of A and B is non-interacting - it's $\rho \otimes \rho_{\mathrm{env}}$ (This would be unusual in nature but could reasonably be set up as the start of an experiment). The dynamics of AB are $\rho_{AB} = \rho \otimes \rho_{\mathrm{env}} \to U_{AB}(\rho \otimes \rho_{\mathrm{env}})U^\dagger = \rho'_{AB}$. So $\rho'_A = \mathrm{tr}_{\mathrm{env}}(U_{AB}(\rho \otimes \rho_{\mathrm{env}})U^\dagger) = \Phi(\rho_A)$. This is called the "Church of the higher Hilbert space" - to see how the operator Φ acts in the system, we add an ancilla taking the state "up" to $\rho \otimes \rho_{\mathrm{env}}$, then can "move it across" with unitary time evolution to $U(\rho \otimes \rho_{\mathrm{env}})U^\dagger$, then take the partial trace oven the environment system $\Phi(\rho) = \mathrm{tr}_{\mathrm{env}}(U(\rho \otimes \rho_{\mathrm{env}})U^\dagger)$. Any quantum operator can be implemented by these three steps.

These first two viewpoints are useful, but it is very hard to apply our criterion from the first one - how do we tell whether a given operator is CP? Thus the following is very useful:

For the third, the Kraus representation or operator-sum representation, a quantum operator Φ on a state ρ of a system A can be represented as $\Phi(\rho) = \sum_{k=1}^{M} A_k \rho A_k^\dagger$, where the A_k are a finite set of linear operators acting on $\mathcal{B}(\mathcal{H}_A)$ with $\sum_k A_k^\dagger A_k = I$. (This is actually an if and only if condition - any such expression represents a quantum operator). For any ρ we have A_k a Kraus operator or operational element; $\sum_k A_k^\dagger A_k = I$ is the completeness relation. $\mathrm{tr}\phi(\rho) = 1 = \sum_k \mathrm{tr}(A_k \rho A_k^\dagger)$. By cyclicity of trace, this is $\sum_k \mathrm{tr}(A^k \dagger A_k \rho) = \mathrm{tr}((\sum_k A_k^\dagger A_k)\rho) \forall \rho$, so $\sum_k A_k^\dagger A_k = 1$.

Theorem: A map Φ is CPT iff $\Phi(\rho) = \sum_i A_i \rho A_i^\dagger$ for some finite set of linear operators $\{A_i\}$ with $\sum_i A_i^\dagger A_i = I$. Any such Φ is clearly linear with $\mathrm{tr}\Phi(\rho) = 1$.

Kraus Representation Theorem

A map Φ is CPT iff it can be written in the form $\Phi(\rho) = \sum_i A_i \rho a_i^\dagger$, where $\{A_i\}$ are a finite set of linear operators with $\sum_i A_i^\dagger A_i = I$. If Φ can be expressed in this form, it is linear and CPT: $\Phi(a\rho_1 + b\rho_2) = a\Phi(\rho_1) + b\Phi(\rho_2)$, $(\Phi \otimes \mathrm{id}) \geq 0$, i.e. for $\rho \geq 0$, $\langle\phi|(\Phi \otimes \mathrm{id})\rho|\phi\rangle \geq 0 \forall |\phi\rangle \in \mathcal{H}$. The left hand side here is $\sum_i \langle\phi|(A_i \otimes I)\rho(A_i^\dagger \otimes I)|\phi\rangle$; set $(A_i^\dagger \otimes I)|\phi\rangle = |e_i\rangle$, then this is $\sum_i \langle e_i|\rho|e_i\rangle \geq 0$. Finally for trace preserving, $\Phi(\rho) = \sum_i A_i \rho A_i^\dagger$; $\mathrm{tr}\Phi(\rho) = \sum_i \mathrm{tr}(A_i \rho A_i^\dagger) = \sum_i \mathrm{tr}(A_i^\dagger A_i \rho) = \mathrm{tr}(\sum(A_i^\dagger A_i)\rho) = \mathrm{tr}\rho = 1$.

Any linear CPT map has a Kraus form; we shall use Schumacher's "relative-state method" to derive this, which is substantially easier than Kraus' original method. Characterise the action of Φ by $(\Phi_A \otimes \mathrm{id}_B)\rho$ where ρ is a special pure state, a maximally entangled state - e.g. for a qudit - a set of d qubits - $\rho = |\psi\rangle\langle\psi|$ where $\psi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i_A\rangle \otimes |i_B\rangle$, where $\{|i_A\rangle\}$ are an ONB for $\mathcal{H}_A$, $\{|i_B\rangle\}$ an ONB for $\mathcal{H}_B$ (we take the two spaces to have the same dimension; note that there is no special relation between two states $|i_A\rangle, |i_B\rangle$ with the same label). For a MES, the reduced state is a completely mixed state - $\rho_A = \mathrm{tr}_B \rho = \frac{I}{d} = \sum_{i=1}^d \frac{1}{d} |i_A\rangle\langle i_A|$.

To see why this suffices, we first consider the analagous thing for operators. What is the action of an operator $M_A|\phi_A\rangle$? It suffices to find $(M_A \otimes I)|\psi_{AB}\rangle$ where $|\psi_{AB}\rangle$ is again an MES; we will work with $|\tilde{\psi}_{AB}\rangle = \sqrt{d}|\psi_{AB}\rangle = \sum_{i=1}^d |i_A\rangle|i_B\rangle$ for convenience (of course $|i_A\rangle|i_B\rangle$ really means $|i_A\rangle \otimes |i_B\rangle$). We have $\langle\tilde{\psi}_{AB}|\tilde{\psi}_{AB}\rangle = d$. Any vector $|\phi_A\rangle \in \mathcal{H}_A$ can be obtained from the MES: write $|\phi_A\rangle = \sum_{i=1}^d a_i |i_A\rangle$. Then $|\phi_A\rangle = \langle\phi_B^\star|\tilde{\psi}_{AB}\rangle$ where $|\phi_B^\star\rangle$ is defined as $\sum a_i^\star |i_B\rangle$. ($\langle\phi_B^\star|\tilde{\psi}_{AB}\rangle$ is a "partial inner product"). To see this, then we have $\langle\phi_B^\star = \sum_{j=1}^d a_j \langle j_B|$, so our partial inner product is $(\sum_{j=1}^d a_j\langle j_B|)(\sum_i |i_A\rangle|i_B\rangle) = \sum_i a_i |i_A\rangle = |\phi_A\rangle$. This is the "relative state" to $|\phi_B^\star\rangle$; $|\phi_B^\star$ is the "index state" that yields $|\phi_A\rangle$ from the MES $|\tilde{\psi_{AB}}\rangle$. The reader may check the map $|\phi_A\rangle \to |\phi_B^\star\rangle$ is antilinear ($A(cf(x)) = c^\star A(f(x))$, $A(f_1(x) + f_2(x)) = A(f_1(x)) + A(f_2(x))$).

Consider $(M_A \otimes I_B)|\tilde{\psi}_{AB}\rangle = (M_A \otimes I_B)\sum_i |i_A\rangle|i_B\rangle = \sum_i M_a|i_A\rangle|i_B\rangle$. So $M_A|\phi_A\rangle$, where $|\phi_A\rangle = \sum a_i|i_A\rangle$, can be found by $\langle\phi_B^\star|(M_A \otimes I_B)|\tilde{\psi}_{AB}\rangle = \sum_j a_j\langle j_B| \sum_i M_A i_A\rangle|i_B\rangle = \sum_i a_i M_A|i_A\rangle = M_A \sum a_i|i_A\rangle = M_a|\phi_A\rangle$. So in summary, $|\phi_A\rangle = \langle\psi_B^\star|\tilde{\psi}_{AB}\rangle$ the "partial inner product", and the action of any operator M_A on any arbitrary possible state $|\phi_A\rangle \in \mathcal{H}_A$ is obtained as a relative state from $(M_A \otimes I_B)|\tilde{\psi}_{AB}\rangle$. So to find $M_A|\phi_A\rangle$ it suffices to find $(M_A \otimes I)|\tilde{\psi}_{AB}\rangle$ - then we just take the relative state and take the partial inner product $\langle\phi_B^\star|$.

Apply the relative state method to superoperators - this is covered in more detail on the second example sheet for this course. $(\Phi_A \otimes \mathrm{id}_B) \geq 0 \therefore (\Phi_A \otimes \mathrm{id}_B)\tilde{\rho}_{AB}$, where $\tilde{\rho}_{AB} = |\tilde{\psi}_{AB}\rangle\langle\tilde{\psi}_{AB}|$, $|\tilde{\psi}_{AB}\rangle = \sum_{i=1}^d |i_A\rangle \otimes |i_B\rangle$, $\mathrm{tr}\tilde{\rho}_{AB} = d$. Consider the two operations 1) obtaining $|\phi_A\rangle$ from $|\tilde{\psi}_{AB}\rangle$ by the relative state method - which only acts nontrivially on the B part, and 2) applying the quantum operator Φ_A or $\Phi_A \otimes \mathrm{id}_B$; since each only acts on one space, these commute.

We want to find $\Phi_A(|\phi_A\rangle\langle\phi_A|)$ (then it is easy to find $\Phi(\rho)$ for general ρ (e.g. by linearity)). We can either apply $\langle\phi_B^\star|$ to $|\tilde{\psi}_{AB}\rangle$ to get $|\phi_A\rangle$ and then apply Φ_A, or apply $(\phi_A \otimes \mathrm{id}_B)$ and then take $\langle\phi_B^\star|$.

Recall we are trying to determine $\rho'_A = \Phi_A(|\phi_A\rangle\langle\phi_A|)$. This is $= \Phi_A(\langle\phi_B^\star|\tilde{\psi}_{AB}\rangle\langle\tilde{\psi}_{AB}|\phi_B^\star\rangle) = \langle\phi_B^\star|(\Phi_A \otimes \mathrm{id}_B)|\tilde{\psi}_{AB}\rangle\langle\tilde{\psi}_{AB}|\phi_B^\star\rangle$, as is proven on the second example sheet.

So $\rho'_A = \langle\phi_B^\star|\tilde{\rho}'_{AB}|\phi_B^\star\rangle$ where $\tilde{\rho}'_{AB}$ is defined as $(\Phi_A \otimes \mathrm{id}_B)\tilde{\rho}_{AB}$, where $\tilde{\rho}_{AB} =$

$|\tilde{\psi}_{AB}\rangle\langle\tilde{\psi}_{AB}|$. $\tilde{\rho}'_{AB}$ is a (unnormalized) density matrix - so we may associate to it an ensemble of pure states $\{\lambda_k, |\tilde{\alpha}_k^{AB}\rangle\}$, $\tilde{\rho}'_{AB} = \sum_k \lambda_k |\tilde{\alpha}_k^{AB}\rangle\langle\tilde{\alpha}_k^{AB}|$. The λ_k are probabilities so $\lambda_k > 0, \sum_k \lambda_k \langle\tilde{\alpha}_k^{AB}$
$tilde\alpha_k^{AB}\rangle = d$ (as $\langle\tilde{\psi}|\tilde{\psi}\rangle = d$). Substituting, $\rho'_A = \Phi_A(|phi_A\rangle\langle\phi_A|) = \sum_k \lambda_k \langle\phi_B^\star|\tilde{\alpha}_k^{AB}\rangle\langle\tilde{\alpha}_k^{AB}|\phi_B^\star\rangle$. Set $A_k|\phi_A\rangle = \sqrt{k}\langle\phi_B^\star|\tilde{\alpha}_k^{AB}\rangle$, so this becomes $\sum_k A_k(|\phi_A\rangle\langle\phi_A|)A_k^\dagger$, and as Φ is trace preserving, $\sum_k A_k^\dagger A_k = I$.

Summary: Φ completely positive implies $\Phi \otimes \mathrm{id}_B \geq 0$. $\tilde{\rho}^|_{AB} prime = (\Phi \otimes \mathrm{id}_B)(|\tilde{\psi}_{AB}\rangle\langle\tilde{\psi}_{AB}|) \geq 0$ Associate to $\tilde{\rho}'_{AB}$ $\{\lambda_k, |\tilde{\alpha}_k^{AB}\rangle\}$, and to each $|\tilde{\alpha}_k^{AB}\rangle$ corresponds $A_k|\phi_A\rangle = \sqrt{\lambda_k}\langle\phi_B^\star|\tilde{\alpha}_k^{AB}\rangle$.

Recap: Any CPT map Φ can be written $\Phi(\rho) = \sum_k A_k \rho A_k^\dagger$ with $\sum_k A_k^\dagger A_k = 1$. (1) $|\psi_{AB}\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i_A\rangle|i_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, taking $\dim \mathcal{H}_A = d = \dim \mathcal{H}_B$. $|\tilde{\psi}_{AB}\rangle = \sqrt{d}|\psi_{AB}\rangle = \sum |i_A\rangle|i_B\rangle$; $\langle\tilde{\psi}|\tilde{\psi}\rangle = d$.

1) Any $|\phi_A\rangle \in \mathcal{H}_A$ can be written as $\sum a_i|i_A\rangle$ where the $\{|i_A\rangle\}$ are an ONB in $\mathcal{H}_A$. We write $|\phi_B^\star\rangle = \sum a_i^\star|i_B\rangle$ and we have that $|\phi_A\rangle = \langle\phi_B^\star|\tilde{\psi}_{AB}\rangle$, a "partial inner product".

2) For $M_A \in \mathcal{B}(\mathcal{H}_A)$, $M_A|\phi_A\rangle = \langle\phi_B^\star(M_A \otimes I_B|\tilde{\psi}_{AB}\rangle$.

3) $\forall\Phi$ acting on $\mathcal{B}(\mathcal{H}_A)$, for a pure state $|\phi_A\rangle\langle\phi_A|$, writing $\tilde{\rho}_{AB} = |\tilde{\psi}_{AB}\rangle\langle\tilde{\psi}_{AB}|$, $\Phi(|\phi_A\rangle\langle\phi_A|) = \langle\phi^\star|(\Phi \otimes \mathrm{id})\tilde{\rho}_{AB}|\phi_B^\star\rangle$, as $\tilde{\rho}_{AB} = \sum_{ij} |i_A\rangle|i_B\rangle\langle j_A|\langle j_B|$, $\langle\phi_B^\star| = \sum_k a_k\langle k|$, $\phi_B^\star\rangle = \sum_l a_l^\star|l_B\rangle$. So this $= \sum_{klij} a_k a_l^\star \langle k_B|(\Phi\otimes\mathrm{id})(|i\rangle\langle j|_A \otimes |i\rangle\langle j|_B)|l_B\rangle = \sum a_k a_l^\star \Phi(|i_A\rangle\langle j_B|)\langle k_B|i_B\rangle\langle j_B|l_B\rangle$; these last terms are simply $\delta_{ik}\delta_{jl}$ so this is $\sum a_i a_j^\star \Phi(|i_A\rangle\langle j_A|) = \Phi(\sum_i a_i|i_A\rangle)(\sum_j a_j^\star\langle j|) = \Phi(|\phi_A\rangle\langle\phi_A|)$.

4) $\Phi(|\phi_A\rangle\langle\phi_A|) = \langle\phi_B^\star\tilde{\rho}'_{AB}|\phi_B^\star\rangle$ where $\tilde{\rho}'_{AB} = (\Phi \otimes \mathrm{id})\tilde{\rho}_{AB} \geq 0$; $\mathrm{tr}\tilde{\rho}_{AB} = d$. Associate this with a set $\{\lambda_k, \tilde{A}_k^{AB}\rangle\}$ where $\tilde{\rho}'_{AB} = \sum_k \lambda_k |\tilde{a}_k^{|B}\rangle\langle\tilde{\alpha}_k^{AB}|$, $\langle\tilde{\alpha}_k|\tilde{\alpha}_k\rangle = d$, $\lambda_k > 0$, $\sum_k \lambda_k = 1$.

5) $\Phi(|\phi_A\rangle\langle\phi_A|) = \sum \lambda_k\langle\phi_B^\star|\tilde{\alpha}_k^{AB}\rangle\langle\tilde{\alpha}_k^{AB}|\phi_B^\star\rangle$. We want to have $\Phi(|\phi_A\rangle\langle\phi_A|) = \sum A_k(|\phi_A\rangle\langle\phi_A|)A_k^\dagger$. So set $A_k : |\phi_A\rangle \mapsto \sqrt{\lambda_k}\langle\phi_B^\star|\tilde{\alpha}_k^{AB}\rangle$, then we have this result. The maximum number of Kraus operators is d^2 (as this is the dimension of $\mathcal{H}_{AB}$). We will see later that the Kraus representation is not unique - it is possible to have $\sum A_k \rho A_k^\dagger = \sum v_j \rho v_j^\dagger$.

Schmidt Decomposition

Consider $\mathcal{H}_A \otimes \mathcal{H}_B$ as always. For any pure state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, $|\psi_{AB}\rangle = \sum_i \lambda_i|i_A\rangle|i_B\rangle$, where $\dim \mathcal{H}_A = d_A$, $\dim \mathcal{H}_B = d_B$, $\{|i_A\rangle\}$ is a set of ON states in $\mathcal{H}_A$, $\{|i_B\rangle\}$ a set of ON states in $\mathcal{H}_B$, $\lambda_i \geq 0$ real and $\sum_i \lambda_i^2 = 1$: consider $\{|r_A\rangle\}$ an ONB in $\mathcal{H}_A$, $\{|\alpha_B\rangle\}$ an ONB in$\mathcal{H}_B$. We have $|\psi_{AB}\rangle = \sum_{r=1}^{d_A}\sum_{\alpha=1}^{d_B} a_{r\alpha}|r_A\rangle \otimes |\alpha_B\rangle$, since the $|r_A\rangle \otimes |\alpha_B\rangle$ form a basis for $H_A \otimes \mathcal{H}_B$. We can vie the $a_{r\alpha}$ as elements of a matrix A, $d_A \times d_B$. We have the Singular Value Decomposition $A = UDV$ for V, U unitary, D diagonal with non-negative entries. (At this point the lecturer wimped out and set $d_A = d_B = d$). Call elements of U u_{ri} and elements of V $v_{\beta\alpha}$. Then $a_{r\alpha} = \sum_i \sum_\beta u_{ri} d_{i\beta} v_{\beta\alpha}$. Substitute $d_{i\beta} = d_{ii}\delta_{i\beta}$, then by algebra $|\psi_{AB}\rangle = \sum_i \sum_\beta \delta_{ib} d_{ii} \sum_r u_{ri}|r_A\rangle \otimes \sum_\alpha u_{\beta\alpha}|\alpha_B\rangle = \sum_i d_{ii} \sum u_{ri}|r_A\rangle \otimes \sum v_{i\alpha}|\alpha_B\rangle$; call this respectively $\sum_i \lambda_i|i_A\rangle \otimes |i_B\rangle$, then we have the result. $\lambda_i \geq 0$ and the reader may check $|i_A\rangle = \sum_r u_{ri}|r_A\rangle$ and $|i_B|rangle$ are each orthonormal by unitarity of U, V; we have $\sum_i \lambda_i^2 = 1$.

$\rho_A = \mathrm{tr}_B|\psi_{AB}\rangle\langle\psi_{AB}| = \sum \lambda_i^2|i_A\rangle\langle i_A\rangle$, $\rho_B = \sum_i \lambda_i^2|i_B\rangle\langle i_B|$. So $\mathrm{tr}\rho_A = \sum_i \lambda_i^2 = 1 =$

$\mathrm{tr}\rho_B$. So ρ_A, ρ_B have identical non-zero eigenvalues. When writing $|\psi_{AB}\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle$, the λ_i are called the Schmidt coefficients. The Schmidt number is the number of nonzero Schmidt coefficients.

Bipartite pure state: $|\psi_{AB}\rangle$ is a product state if and only if $N_S = 1$. If $N_s > 1$ then $|\psi_{AB}\rangle$ is entangled.

Purification: We may alwas assume the environment is in a pure state. For $\mathcal{H}_A$ with mixed states ρ, we can always associate pure states $|\psi_{AR}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R$; $\mathcal{H}_R$ is the reference state space, physically meaningless. We want $\rho = \mathrm{tr}_R ||psi_{AR}\rangle\langle\psi_{AR}|$. Perform the spectral decomposition $\rho_A = \sum_{i=1}^{d} p_i |i_A\rangle\langle i_A|$; set $\dim \mathcal{H}_R = d$ and take $\{|i_R\rangle\}_{i=1}^{d}$¿ Then set $|\psi_{AR}\rangle := \sum_i \sqrt{p_i}|i_A\rangle|i_R\rangle$; compare this with the Schmidt decomposition.

No Cloning Theorem

The popular version of this theorem is that there cannot exist a "quantum copier": if $|\psi\rangle, |\phi\rangle$ have $\langle\phi|\psi\rangle \neq 0$, then we cannot copy the unknown state $|\omega\rangle$= one of $|\phi\rangle, |\psi\rangle$.

The actual statement is simpler: an unknown quantum state cannot be "copied" or "cloned" by a unitary transformation. Assume we have a quantum copier: there is some standard "blank" state $|s\rangle$ and U with $U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle, U(|\phi\rangle \otimes |s\rangle) = |\phi\rangle \otimes |\phi\rangle$. But then take inner products— $\langle\phi|\psi\rangle\langle s|s\rangle = \langle\phi|\psi\rangle\langle\phi|\psi\rangle$, i.e. $x = \langle\phi|\psi\rangle$ satisfies $x = x^2$ so must = 1 or 0, i.e. $|\phi\rangle = |\psi\rangle$ or $\langle\phi|\psi\rangle = 0, |\phi\rangle \perp |\psi\rangle$.

Generalized Measurement Postulate

In projective measurement, measurement can be characterised completely by $\{P_j\}$; P_j is a projection operator onto the eigenspace of A corresponding to the eigenvalue a_i. We meeasure A with $A = A^\dagger$; the system state is $|\psi\rangle$ and outcomes are eigenvalues $A|\phi_j\rangle = a_j|\phi_j\rangle$. The questions a theory of measurement must answer are: what is the probability of outcome a_j, $p(a_j)$, and what is the post-measurement state?

In other words, for projective measurement we have a spectral decomposition $A = \sum_j a_j P_j$. $p(a_j) = \langle\psi|P_j|\psi\rangle$, and if the outcome is a_j, $|\psi\rangle \mapsto |\psi'\rangle = \frac{P_j|\psi\rangle}{\sqrt{P_j|\psi\rangle}}$.

What about for general (impure) states ρ? The reader may check $p(a_j) = \mathrm{tr}(\rho P_j)$, and $\rho \mapsto \rho' = \frac{P_j \rho P_j}{\mathrm{tr}(P_j \rho)}$ if the outcome is a_j.

We express the projective measurement postulate for projectios $\{P_j\}$ with $A = \sum a_j P_j, A = A^\dagger$: 1) $P_j \geq 0$ because $p(a_j) = \mathrm{tr}(\rho P_j) \geq 0$ 2) $\sum_j P_j = I$ as $\sum_j p(a_j) = \mathrm{tr}(\rho \sum_j P_j) = \mathrm{tr}\rho = 1$ 3) $P_j P_{j'} = \delta_{jj'} P_j$; the projections are orthogonal to each other.

Generalized measurement will boil down to removing the third of these postulates - the first two postulates are obvious requirements for any sensible notion of measurement, but the third is merely an outcome of the spectral decomposition - which in turn only worked because $A = A^\dagger$.

Motivation for why we need a generalized measurement postulate:

1) Say we are measuring a system X in state ρ - but the measuring system itself, Y, will interact with X. And while the measurement will be projective

on the combined system XY, it need not be projective on X - compare this with time evolution, which is unitary on XY but not always on X.

2) Say we have $|\psi\rangle$ a state such that $\sigma \cdot \hat{n}|\psi\rangle = |\psi\rangle$ where $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ and $\hat{n}$ is some unit vector in $\mathbb{R}^3$. Ei.g. for $\widehat{n} = (0, 0, 1)$ this means $\sigma_z|\psi\rangle = |\psi\rangle$, so $|\psi\rangle = |0\rangle = \begin{pmatrix}1\\0\end{pmatrix}$. Can we do a measurement to determine an unknown $\widehat{n}$? $\widehat{n}$ does not correspond to any self-adjoint operator, since the only operators of a spin-$\frac{1}{2}$ operator that we can measure are linear combinations of $\sigma_x, \sigma_y, \sigma_z$ and I since together these span the space of 2×2 Hermitian matrices.

Generalized Measurement Postulate

(A generalized measurement is) characterised by a set of operators $\{M_a\}$, where each M_a correspond to an outcome labelled by a. The probability of outcome a is $p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle$ if the system is initally in a state $|\psi\rangle$ or $p(a) = \mathrm{tr}(M_a^\dagger M_a \rho)$ if the system is initially in a state ρ; $1 = \sum_a p(a) \Rightarrow$ we have the completeness relation $\sum_a M_a^\dagger M_a = I$. The post-measurement state if the outcome is A is $|\psi\rangle \mapsto |\psi'\rangle = \frac{M_a|\psi\rangle}{\sqrt{\langle\psi|M_a^\dagger M_a|\psi\rangle}}$ or $\rho \mapsto \rho' = \frac{M_a \rho M_a^\dagger}{\mathrm{tr}(M_a^\dagger M_a \rho)}$.

Claim: projective measurement is a special case of this: if $M_a = P_a$ a projection then the generalized measurement postulate reduces to the projective measurement postulate: we will have $M_a^\dagger = M_a, M_a^2 = M_a, M_a M_b = \delta_{ab} M_a$. $p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle = \langle\psi|P_a^2|\psi\rangle = \langle\psi|P_a|\psi\rangle$, and $|\psi\rangle \mapsto |\psi'\rangle = \frac{P_a|\psi\rangle}{\sqrt{\langle\psi|P_a|\psi\rangle}}$.

How can a generalized measurement be implemented? Suppose an experimentalist prepares the system A in a state $|\psi\rangle$; the ancilla (comprising the measuring device, immediate environment, etc.) is B; the preparation destroys correlations between A and B, so the initial state is a product state $|\psi\rangle \otimes |e_0\rangle$ (we may take the environment to be in a pure state by purification). Then measurement is a unitary evolution of AB, so will introduce interactions between A and B.

Define an operator U: $U(|\psi\rangle \otimes |e_0\rangle) = \sum_a M_a|\psi\rangle \otimes |e_a\rangle$. $\{|e_a\rangle\}$ are mutually orthonormal states of $\mathcal{H}_B$. To check that such a U is unitary, set $|\Psi\rangle = U(|\psi\rangle \otimes |e_0\rangle) = \sum M_a|\psi\rangle \otimes |e_a\rangle$, $|\Phi\rangle = U(|\phi\rangle \otimes |e_0\rangle) = \sum_{a'} M_{a'}|\phi\rangle \otimes |e_{a'}\rangle$. The reader should verify that $\langle\Phi|\Psi\rangle = \langle\phi|\psi\rangle$ (using that $\langle e_a|e_{a'}\rangle = \delta_{aa'}, \sum_a M_a^\dagger M_a = I$). This means $(\langle\phi|\otimes\langle e_0|U^\dagger)(U(|\phi\rangle\otimes|e_0\rangle)) = \langle\phi|\psi\rangle$, so the operator U preserves the scalar product between vectors in $\mathcal{H}_A \otimes \mathcal{H}_B$ of the form $(|\psi\rangle \otimes |e_0\rangle)$. $\{|\psi\rangle \otimes |e_0\rangle\}$ is a subspace $\mathcal{H}_S \subset \mathcal{H}_A \otimes \mathcal{H}_B$; $U : \mathcal{H}_S \to \mathcal{H}_{AB}$ since $U(|\psi\rangle \otimes |e_0\rangle) = \sum M_a|\psi\rangle \otimes |e_a\rangle$ does not generally lie in $\mathcal{H}_S$. U preserves inner products between vectors in $\mathcal{H}_S$, and (without proof) any such operator can be extended (non-uniquely) to a unitary operator $U' : \mathcal{H}_{AB} \to \mathcal{H}_{AB}$; for convenience we shall abuse notation and refer to this unitary extension as U also.

1) Initial state is $|\psi\rangle \otimes |e_0\rangle$. 2) Time evolution $U(|\psi\rangle \otimes |e_0\rangle) = \sum_a M_a|\psi\rangle \otimes |e_0\rangle$ 3) Projective measurement on $|\Psi\rangle = U(|\psi\rangle \otimes |e_0\rangle)$: $\{P_a\}$ where $P_a = I_A \otimes |e_a\rangle\langle e_a|$. $p(a) = \langle\Psi|P_a|\Psi\rangle = \langle\psi|M_a^\dagger M_a|\psi\rangle$ is the probability of outcome a, and the post measurement state if the outcome is a is $|\Psi_{AB}\rangle = U(|\psi\rangle \otimes |e_0\rangle) \mapsto |\Psi'\rangle = \frac{P_a|\Psi\rangle}{\sqrt{P_a|\Psi\rangle}}$. $P_a|\Psi\rangle = (I_a \otimes |e_a\rangle\langle e_a|)(U|\psi\rangle \otimes e_0)) = asum(I_a \otimes |e_a\rangle\langle e_a|)(M_{a'}|\psi\rangle \otimes e_{a'}) = M_a|\psi\rangle \otimes |e_a\rangle$.

Entanglement

We use a simplification known as POVM - positive operator valued measure. We want to find the probability of an outcome $p(a)$ but are sometimes not interested in the post-measurement state. We saw $p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle$. Define $E_a := M_a^\dagger M_a$. Properties: 0a) self-adjoint, $E_a = E_a^\dagger$ 1) $E_a \geq 0$: $\forall|\phi\rangle \in \mathcal{H}$, $\langle\phi|E_a|\phi\rangle = ||M_a|\phi\rangle||^2 \geq 0$. 2) Completeness: $\sum_a E_a = \sum_a M_a^\dagger M_a = I$.

The $\{E_a\}$ form a positive semidefinite partition of unity. $p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle = \langle\psi|E_a|\psi\rangle = \text{tr}(\rho E_a)$ if the initial state is ρ. 1) $E_a \geq 0 \Rightarrow p(a) \geq 0$, because $p(a) = \text{tr}(\rho E_a)$ 2) $\sum_a E_a = I \Rightarrow \sum_a p(a) = 1$. The E_a are called POVM elements and $\{E_a\}$ is a POVM.

Special case: initial state $|\psi\rangle$, on the ath outcome $|\psi\rangle \mapsto |\psi'\rangle \propto M_a|\psi\rangle$. From $E_a = M_a^\dagger M_a$, we could define $M_a = \sqrt{E_a}$. But this is not unique, so the post-measurement state is not uniquely defined.

Definition (POVM): A POVM is defined by any positive semidefinite partition of unity, $E_a \geq 0$, $\sum_a E_a = I$. If a measurement is described by $\{E_a\}$, $p(a) = \text{tr}(\rho E_a)$, $\sum_a p(a) = 1$.

Pure POVM: If $E_a = |\phi_a\rangle\langle\phi_a|\forall a$, $\{E_a\}$ is called a pure POVM; see the second example sheet.

Suppose an experimentalist prepares a qubit in state $|\psi\rangle$ with $\sigma \cdot \hat{n}|\psi\rangle = |\psi\rangle$, $\sigma = (\sigma_x, \sigma_y, \sigma_z)$, $\hat{n} \in \mathbb{R}^3$, and we are given that $\widehat{n}$ lies in some set $\{\widehat{n}_a\}$ such that $\sum_a \lambda_a \widehat{n}_a = 0$ for some $0 < \lambda_a < 1, \sum \lambda_a = 1$. What would be the POVM elements that would characterise the generalized measurement that would help us to determine $\widehat{n}$?

Define E_a (corresponding to $\widehat{n}_a$) by $\lambda_a(1 + \widehat{n}_a \cdot \sigma)$. The reader may check $E_a = 2\lambda_a P_{\widehat{n}_a}$ where $P_{\widehat{n}_a} = |\uparrow_{\widehat{n}_a}\rangle\langle\uparrow_{\widehat{n}_a}|$, projection onto up spin along the direction $\widehat{n}_a$. We have $E_a \geq 0$ as $P \geq 0, \lambda_a > 0$, and $\sum_a E_a = \sum_a \lambda_a 1 + \sum_a \lambda \widehat{n}_a \cdot \sigma = 1 + 0 = 1$. So $\{E_a\}$ forms a valid POVM.

Case 1: $\widehat{n} \in \{\widehat{n}_1, \widehat{n}_2\}$. $\sum_a \lambda_a \widehat{n}_a = 0 \Rightarrow \lambda_1 = \lambda_2 = \frac{1}{2}, \widehat{n}_1 = -\widehat{n}_2$. In this case $E_1 = 2\lambda_1 P_{\widehat{n}_1} = P_{\widehat{n}_1}, E_2 = P_{\widehat{n}_2} = I - P_{\widehat{n}_1} = P_{-\widehat{n}_1}$. E_1, E_2 are projection operators which project onto orthogonal spaces, so this is just a projective measurement. $p(1)$, the probability of outcome $\widehat{n}_1$, is, if the initial state was $|\psi\rangle$, $\langle\psi|E_1|\psi\rangle = \frac{1}{2}\langle\psi|I + \widehat{n}_1 \cdot \sigma|\psi\rangle$. If indeed $\widehat{n} = \widehat{n}_1$, then $\sigma \cdot \widehat{n}_1|\psi\rangle = |\psi\rangle$ and the reader may check $p(\widehat{n}_1) = 1$; if $\widehat{n} = \widehat{n}_2$ then the reader may check $p(\widehat{n}_1) = 0$.

Case 2: $\widehat{n} \in \{\widehat{n}_1, \widehat{n}_2, \widehat{n}_3\}$, and consider the symmetric case $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3}$. $E_a = \frac{2}{3}P_{\widehat{n}_a}$ for $a = 1, 2, 3$, and these are not mutually orthogonal. The probabilities $p(1), p(2), p(3)$ are computed on the second example sheet.

Entanglement and its applications

Consider a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$. A pure state $|\psi_{AB}\rangle$ is entangled if its Schmidt number is > 1 ($|\psi_{AB}\rangle = \sum_{i=1}^{n_s} \lambda_i |i_a\rangle|i_B\rangle$). Notice $\rho_A = \sum_{i=1}^{n_s} \lambda_i^2 |i_A\rangle\langle i_A|$ so N_S is the number of non-zero eigenvalues of ρ_A.

Entanglement has no classical analogue. It cannot be created or increased by local actions or classical communications (LOCC). E.g. for $|\psi_{AB}\rangle$, local unitary (LU) operations $(U_A \otimes U_B)$, $(U_A \otimes U_B)|\psi_{AB}\rangle = |\tilde{\phi}\rangle$¡ the entanglement of $|\tilde{\phi}\rangle$ is never $>$ the entanglement of $|\psi_{AB}\rangle$; specifically the Schmidt number n_ϕ is never $> n_\psi$. The same is true for classical communications and general local operations.

The Bell states are MES $|\psi\rangle$ for which $\rho_A = \frac{I}{2}$ - the reduced state is a

completely mixed state. For a pair of qubits, $|\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), |\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$ in $\mathcal{H}_A \otimes \mathcal{H}_B$. The four Bell states can be characterised by two classical bits: the parity bit, whether the spins of the two qubits are parallel or not, say 0 for parallel (a ϕ state) and 1 for antiparallel (a ψ state). The other bit is the phase bit, 0 for a + state and 1 for a - state. So one can encode two classical bits in the state of a 2-qubit system.

If we have both qubits, we can obviously recover the state by projective measurements - use $P_{01} = |\phi^-\rangle\langle\phi^-|$ and similar. But if we are only allowed to work locally on each qubit, there is no way to recover the information.

Recall: we can encode two classical bits in a Bell state, characterised by the parity bit (Φ or Ψ) and phase bit (+ or -). Then we can make a Bell measurement - a projective measurement on the Bell basis - to retrieve both bits. But, if the two qubits are in different locations A and B, we cannot recover the information of the two classical bits by LOCC.

(I) Effect of local unitary (LU) operations: Alice and Bob manipulate the information encoded in the shared state (e.g. $|\phi^+\rangle$), but neither of them can access the information by local measurements. E.g. Alice applies σ_z to her qubit; $\sigma_z|0\rangle = |0\rangle, \sigma_z|1\rangle = -|1\rangle$, so $(\sigma_z^A \otimes I_B)|\phi_{AB}^+\rangle = (\sigma_z^A \otimes I_B)(\frac{|00\rangle+|11\rangle}{\sqrt{2}}) = \frac{|00\rangle-|11\rangle}{\sqrt{2}} = |\phi_{AB}^-\rangle$ - the phase bit has flipped, and in fact this will always occur - σ_z^A takes $|\phi_{AB}^+\rangle \leftrightarrow |\phi_{AB}^-\rangle, |\psi_{AB}^+\rangle \leftrightarrow |\psi_{AB}^-\rangle$.

E.g. if Alice applies σ_x, $|\phi_{AB}^+\rangle \rightarrow |\psi_{AB}^+\rangle$, and the reader should calculate $\sigma_x^A \otimes I_B|\phi_{AB}^-\rangle$.

So the effect of local unitary operations is to send one Bell state to another; we still have $\rho_A = \frac{I_A}{2}$.

Effect of allowing CC

Note that the Bell states are simultaneous eigenstates of two commuting operators: $X_{AB} = \sigma_x^A \otimes \sigma_x^B, Z_{AB} = \sigma_z^A \otimes \sigma_z^B$; the reader may check $[X_{AB}, Z_{AB}] = 0$ even though e.g. $[\sigma_x^A \otimes I_B, Z_{AB}] \neq 0$. ($[\sigma_x^A \otimes I_B, X_{AB}] = 0$, trivially).

For $\alpha = \Phi, \Psi$, $X_{AB}|\alpha_{AB}^+\rangle = |\alpha_{AB}^+\rangle, X_{AB}|\alpha_{AB}^-\rangle = -|\alpha_{AB}^-\rangle$. So the eigenvalue of X_{AB} is the phase bit. The eigenvalue of Z_{AB} is the parity bit: $Z_{AB}|\Phi_{AB}^{\pm}\rangle = |\Phi_{AB}^{\pm}\rangle, Z_{AB}|\Psi_{AB}^{\pm}\rangle = -|\Psi_{AB}^{\pm}\rangle$ (the reader should check this).

So with LU and CC allowed, say Alice and Bob both decide to measure σ_x^A, σ_x^B. The final state remains an eigenstate of X_{AB} since σ_x^A, σ_x^B commute with X_{AB}. From their results Alice and Bob can figure out the phase bit. But, σ_x^A, σ_x^B do not commute with Z_{AB}.

Now, allowing arbitrary LO, including POVM. The result is that the Bell satte a (MES) can be converted into a different entangled state, not necessarily a MES. $|\phi_{AB}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. (1) Alice performs a measurement on qubit A which has two outcomes: measurement operators $M_1 = \begin{pmatrix} \cos\theta & 0 \\ 0 & \sin\theta \end{pmatrix}, M_2 = \begin{pmatrix} \sin\theta & 0 \\ 0 & \cos\theta \end{pmatrix}$. Recall: if the outcome is 1, Alice's qubit $|\phi\rangle \mapsto (|\phi'\rangle) \propto M_1|\phi\rangle$. $M_1|0\rangle = \cos\theta|0\rangle, M_1|1\rangle = \sin\theta|1\rangle, M_2|0\rangle = \sin\theta|0\rangle, M_2|1\rangle = \cos\theta|1\rangle$. If the outcome of Alice's measurement is 1, the final state is $\propto (M_1|0\rangle|0\rangle) + (M_1|1\rangle)|1\rangle$. So if the outcome is 1 then $|\Phi_{AB}^+\rangle \rightarrow \cos\theta|00\rangle + \sin\theta|11\rangle$; if the outcome is 2

$|\Phi^+_{AB}\rangle \to \sin\theta|00\rangle + \cos\theta|11\rangle$.

The protocol is: if this outcome is 1, Alice doesn't do anything. If the outcome is 2, Alice acts on her qubit with $\sigma_x^{(A)}$ - so the resulting state if her outcome was 2 is $\sin\theta|10\rangle + \cos\theta|01\rangle$. Alice tells Bob whether her outcome was 1 or 2; if it was 1, Bob does nothing to his qubit, while if Alice's was 2 then Bob acts by $\sigma_x^{(B)}$. So the final shared state is $\cos\theta|00\rangle + \sin\theta|11\rangle$, whichever result Alice obtained - so $|\Phi^+_{AB}\rangle \mapsto |\chi_{AB}\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$. We claim $\rho^A_\chi \neq \frac{I}{2} \neq \rho^B_\chi$.

What about general states (in AB) $|\psi\rangle, |\phi\rangle$? Is it possible to have $|\psi\rangle \to |\phi\rangle$ by LOCC. Here a distinct area of pure mathematics can be applied: Majorization. This is about ordering of real n-dimensional vectors. Given $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, form $\mathbf{x}^\downarrow = (x_1^\downarrow, x_2^\downarrow, \ldots, x_n^\downarrow)$ where $\downarrow$ is a permutation such that $x_1^\downarrow \geq x_2^\downarrow \geq \cdots \geq x_n^\downarrow$, similarly for $\mathbf{y}^\downarrow$. We say $\mathbf{x}$ is majorized by $\mathbf{y}$, $\mathbf{x} \prec \mathbf{y}$, if $\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow \forall k = 1, \ldots, n-1$ and $\sum_{i=1}^n x_i^\downarrow \sum_{i=1}^n y_i^\downarrow$.

Let ρ^A_ψ, ρ^A_ϕ be the respective reduced density matrices, $\lambda_\psi = (\nu_1, \ldots, \nu_n), \lambda_\phi = (\mu_1, \ldots, \mu_n)$ be vectors whose elements are eigenvalues of ρ^A_ψ and ρ^A_ϕ. Take $\nu_1 \geq \cdots \geq \nu_n, \mu_1 \geq \cdots \geq \mu_n$.

Theorem: $|\psi\rangle$ can be mapped to $|\phi\rangle$ by LOCC iff $\lambda_\psi \prec \lambda_\phi$, i.e. $\sum_{i=1}^k \nu_i \leq \sum_{i=1}^k \mu_i$ $(\forall k < n)$ (we have $\sum_{i=1}^n \nu_i = 1 = \sum_{i=1}^n \mu_i$). Consequence: entanglement cannot be increased by LOCC.

Recap: If Alice and Bob share a bell state $|\psi\rangle \in \{|\Phi^\pm\rangle, |\Psi^\pm\rangle\}$, 1) By a LU operation, we can have $|\psi\rangle \mapsto$ another Bell state. 2) By LOCC a) where the LO is projective measurement, e.g. A and B decide (using CC) to measure σ_x or σ_z on each qubit, A and B can determine either the phase bit or the parity bit, but not both - see the second example sheet. b) where the LO is a generalized measurement (POVM), and we have CC, we can have $|\psi\rangle_{\text{MES}} \overset{\text{LOCC}}{\mapsto} |\phi\rangle$ not necessarily a MES. So we ask: given $|\psi\rangle, |\phi\rangle$, can one convert $|\psi\rangle$ to $|\phi\rangle$ by LOCC?

Theorem (Nielsen): we may have $|\psi\rangle \mapsto |\phi\rangle$ iff $\lambda_\psi \prec \lambda_\phi$, for $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Recall: $\rho^A_\psi = \text{tr}_B|\psi\rangle\langle\psi|$, similarly ρ^A_ϕ. $\lambda_\psi = (\nu_1, \ldots, \nu_n), \lambda_\phi = (\mu_1, \ldots, \mu_n)$ where $n = \dim \mathcal{H}_A = \dim \mathcal{H}_B$, ν_i are the eigenvalues of ρ^A_ψ including zero, μ_i similarly for ρ^A_ϕ, and we choose to construct $\lambda_\psi, \lambda_\phi$ such that $\nu_1 \geq \nu_2 \geq \ldots, \mu_1 \geq \mu_2 \geq \ldots$ $\lambda_\psi \prec \lambda_\phi$ means $\sum_{i=1}^k \nu_i \leq \sum_{i=1}^k \mu_i \forall k = 1, \ldots, -1$ and we have equality for $k = n$; the proof of this theorem is very clear but too long for this course; it is given in Nielsen and Schrong [sp?].

This theorem implies the following lemma: entanglement (of a pure state) cannot be increased by LOCC. Suppose n_ψ is the Schmidt number of $|\psi\rangle$, similarly n_ϕ, and suppose $|\psi\rangle \overset{\text{LOCC}}{\mapsto} |\phi\rangle$, then we cannot have $n_\phi > n_\psi$ (n_ψ characterises the entanglement of $|\psi\rangle$; $|\psi\rangle$ is an entangled state iff $n_\psi > 1$). We prove by contradiction: assume we have some such LOCC and $n_\phi > n_\psi$. This implies $\exists m \leq n$ such that $\mu_m \neq 0, \nu_m = 0$; (then $\nu_w = 0$ for $w > m$). So $\sum_{i=1}^{m-1} \mu_i \neq 1, \sum_{i=1}^{n-1} \nu_i = 1$ so $\lambda_\psi \not\prec \lambda_\phi$, a contradiction.

Recall: a mixed state ρ is separable if we can express it as $\rho = \sum p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i|$, entangled otherwise.

Applications of entanglement

Superdense coding: suppose Alice wants to send 2 classical bits to Bob; no CC is possible between them, but they have a quantum (qubit) channel which A can use to send a single qubit to B. Can she send two classical bits at once by transmitting only one qubit? The answer is yes, if A and B initially share a Bell state (MES). How? Suppose A and B initially share $|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Alice's possible messages are 00, 01, 10, 11; she has one qubit, her half of $|\Phi^+_{AB}\rangle$. She acts on it according to: if the message is 00, the operation is $\sigma_0 = \mathrm{id}$ giving the final shared state $|\Phi^+_{AB}\rangle$; for 01 she acts with σ_z to give $|\Phi^-_{AB}\rangle$, for 10 σ_x giving $|\Psi^+_{AB}\rangle$¡ and for 11 $i\sigma_y$ giving $|\Psi^-_{AB}\rangle$. Then she sends her qubit to B; Bob now has both qubits A, B, so can perform a Bell measurement on the two to unambiguously identify the state, and hence infer Alice's message.

Also notice that if Eve intercepts Alice's qubit, it will always be in the state $\rho^A_\psi = \frac{I}{2}$; thus there is some eavesdropping resistance.

Quantum teleportation: suppose Alice wants to send an unknown (to her) quantum pure state of a qubit to B, but there is no quantum channel between them; only CC is possible. Again, this is possible if Alice and Bob initally share a Bell state. Let Alice and Bob share $|\Phi^+_{AB}\rangle$ initially; the unknown ntate is $|\psi\rangle_C = |\psi\rangle = a|0\rangle_C + b|1\rangle_C \in \mathcal{H}_C$. The protocol is that A unites the C-qubit $|\psi\rangle_C$ with her member of $|\Phi^+_{AB}\rangle$ - we take the tensor product and consider the tripartite state shared between A and B, which is initally $|\phi\rangle_C \otimes |\Phi^+_{AB}\rangle = (a|0\rangle_C + b|1\rangle_C) \otimes \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}) = \frac{1}{2}(|\Phi^+\rangle_{CA} \otimes |\psi\rangle_B + \frac{1}{2}|\Phi^s\rangle_{CA} \otimes \sigma_z|\psi\rangle_B + \frac{1}{2}|\Psi^+\rangle_{CA} \otimes \sigma_z|\psi\rangle_B + \frac{1}{2}|\Psi^-\rangle_{CA} \otimes (-i\sigma_y)|\psi\rangle_B$, where $|\psi\rangle_B = a|0\rangle_B + b|1\rangle_B$. Alice has both C and A qubits so can perform a Bell Measurement and determine unambiguously the state of CA. She characterizes the outcome by two classical bits - phase and parity - and if the outcome is $|\Phi^-\rangle$, sends 01 to Bob classically. Then Bob acts on his qubit $\sigma_z|\psi\rangle_B$ with σ_z to get $\sigma_z^2|\psi\rangle_B = |\psi\rangle_B$; similarly, Bob will act with other operators if Alice obtained other outcomes from her Bell measurement. Bob's final state will therefore be $|\psi\rangle_B$, whichever outcome Alice obtained. Thus, the state $|\psi\rangle$ has been "teleported" to B without disrupting it.

Note that there is no violation of the no-cloning theorem, because the state of qubit C is destroyed by this operation.

Quantum Entropy

The von Neumann entropy is the quantum analogue of the Shannon entropy.

A quantum information source can be characterised by a set of states $\{|\psi_k\rangle\} \subset \mathcal{H}$ and a set of corresponding probabilities $\{p_k\}$; therefore it can be completely characterised by $(\rho, \mathcal{H})$ where $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$.

Definition: for a density matrix ρ, the von Neumann entropy is $S(\rho) = -\mathrm{tr}(\rho \log \rho)$. (As usual log is base 2 and $0 \log 0$ is defined to be 0). If we choose an orthonormal basis $\{|\psi_i\rangle\}$ which diagonalises ρ we obtain the spectral decomposition $\rho = \sum_i = 1^n \lambda_i |\psi_i\rangle\langle\psi_i|$, with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. $\{|\psi_i\rangle\}$ here are orthonormal eigenvectors corresponding to the eigenvalues $\{\lambda_i\}$. In this case the von Neumann entropy reduces to the Shannon entropy, $S(\rho) = -\sum_i \lambda_i \log \lambda_i = H(\lambda)$, where $\lambda = \{\lambda_i\}$ is the set of eigenvalues of ρ. This makes sense, since these eigenvalues obey the same rules as a probability distribution.

We shall see later that von Neumann information quantifies the incompressible information content (data compression limit) of a memoryless/IID quantum information source, just as Shannon entropy does for a classical information source.

Definition: quantum relative entropy: for density operators ρ_1, ρ_2 (acting on the same Hilbert space), the relative entropy is $S(\rho_1\|\rho_2) = \mathrm{tr}\rho_1(\log\rho_1 - \log\rho_2)$. (This is analagous to the classical relative entropy $D(\mathbf{p}\|\mathbf{q}) = \sum_i p_i \log\frac{p_i}{q_i}$). Note that this is well defined only if $\mathrm{supp}\rho_1 \subset \mathrm{supp}\rho_2$, where $\mathrm{supp}\rho$ is the support of ρ, the subspace spanned by eigenvectors of ρ with nonzero eigenvalues; otherwise $S(\rho_1\|\rho_2) = \infty$.

Two important properties of this are: 1) non-negativity: $S(\rho_1\|\rho_2) \geq 0$ with equality iff $\rho_1 = \rho_2$, 2) joint convexity: for $p_1, p_2 \geq 0$ and $p_1 + p_2 = 1$, $S(p_1\rho_1 + p_2\sigma_1\|p_1\rho_2 + p_2\sigma_2) \leq p_1 S(\rho_1\|\rho_2) + p_2 S(\sigma_1\|\sigma_2)$. This inequality implies $S(\rho_1\|\rho_2)$ is convex in each of its arguments. These will be proven on the third example sheet.

Properties of von Neumann entropy $S(\rho)$

$S(\rho) \geq 0$ with equality iff ρ is a pure state density matrix. If we have a pure state $\rho = |\psi_i\rangle\langle\psi_i|$ then $\lambda_j = \delta_{ij}$ so $S(\rho) = -1\log 1 = 0$.

$S(\rho)$ is invariant under unitary transformations $\rho \to U^\dagger\rho U$; this is obvious because $S(\rho)$ depends only on the eigenvalues of ρ, and the spectrum of an operator is invariant under such a transformation.

If $\dim\mathcal{H} = d$ then $S(\rho) \leq \log d$ with equality iff the system is in a completely mixed state (entropy is maximised when the state is chosen randomly): let $\rho_1 = \rho, \rho_2 = \frac{I}{d}$, then $S(\rho_1\|\rho_2) = \mathrm{tr}\rho(\log\rho - \log\frac{I}{d}) = \mathrm{tr}(\rho\log\rho) - \log\frac{1}{d}\mathrm{tr}\rho I = -S(\rho) + \log d \geq 0$.

The quantum joint entropy is $S(A,B) = -\mathrm{tr}(\rho_{AB}\log\rho_{AB})$, the quantum conditional entropy $S(A|B) = S(A,B) - S(B)$ and the quantum mutual information of two subsystems A, B of a composite system AB is $S(A:B) = S(A) + S(B) - S(A,B) = S(A) - S(A|B) = S(B) - S(B|A)$, all by analogy with their classical counterparts. However, note that some results from the classical case no longer hold: $H(X) \leq H(X,Y)$ and $H(Y) \leq H(X,Y)$ (the entropy in a single variable is always $\leq$ that in a pair of which it is a member), so $H(Y|X) \geq 0$: the conditional entropy is always nonnegative. However, this is not the case for quantum entropy:

Let AB be a system of two qubits in the Bell state $|\Phi^+_{AB}\rangle$. So the density matrix of AB is $|\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|$; those for A and B are completely mixed states $\rho_A = \frac{I_A}{2}, \rho_B = \frac{I_B}{2}$. So we have $S(A,B) = S(\rho_{AB}) = 0$ since ρ_{AB} is a pure state, but $S(A) = S(\rho_A) = 1 \not\leq S(A,B)$. So the quantum conditional entropy $S(B|A) = S(A,B) - S(A) = -1 < 0$.

Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Note that for T an operator of the form $T = T_A \otimes I_B$, $\mathrm{tr}(\rho_{AB}T) = \mathrm{tr}_A(\rho_A T_A)$; in fact this can be taken as the definition of the reduced density matrix.

Concavity: von Neumann entropy is a concave function of its inputs: given probabilities $p_i \geq 0$ with $\sum p_i = 1$ and corresponding density operators ρ_i, $S(\sum_i p_i\rho_i) \geq \sum_i p_i S(\rho_i)$ (proven on the third example sheet, using concavity of $-x\log x$ and spectral decompositions).

Additivity: For ρ_A, ρ_B acting in spaces $\mathcal{H}_A, \mathcal{H}_B$, the entropy of the density

matrix $\rho_A \otimes \rho_B$ acting in $\mathcal{H}_A \otimes \mathcal{H}_B$ is $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$, with the obvious (inductive) generalization for $\rho_1 \otimes \cdots \otimes \rho_n$: let ρ_A, ρ_B have spectral decompositions $\rho_A = \sum_k p_k |\phi_k\rangle\langle\phi_k|, \rho_B = \sum_j q_j |\psi_j\rangle\langle\psi_j|$, then $|\phi_k\rangle \otimes |\psi_j\rangle$ is an eigenvector of $\rho_A \otimes \rho_B$ corresponding to the eigenvalue $p_k q_j$. So $S(\rho_A \otimes \rho_B) = -\sum_{j,k} p_k q_j \log(p_k q_j) = -(\sum_j q_j)\sum_k p_k \log p_k - (\sum_k p_k)\sum_j q_j \log q_j = S(\rho_A) + S(\rho_B)$ (because $\sum_j q_j = 1 = \sum_k p_k$). This is as we should expect: for two independent subsystems, the information of the total system they form is the sum of the information of the two.

Subadditivity: For a system AB composed of two subsystems A, B (no longer assumed independent) and a general (non-product) state ρ_{AB}, $S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$. To see this, set $\rho_1 = \rho_{AB}, \rho_2 = \rho_A \otimes \rho_B$, then $S(\rho_1 || \rho_2) = \mathrm{tr}\rho_{AB} \log \rho_{AB} - \mathrm{tr}\rho_{AB} \log(\rho_A \otimes \rho_B) = -S(\rho_{AB}) - \mathrm{tr}(\rho_{AB} \log((\rho_A \otimes I_B)(I_A \otimes \rho_B))) = -S(\rho_{AB}) - \mathrm{tr}(\rho_{AB} \log(\rho_A \otimes I_B)) - \mathrm{tr}(\rho_{AB} \log(I_A \otimes \rho_B))$. Since $\mathrm{tr}(\rho_{AB}(T_A \otimes I_B)) = \mathrm{tr}_A(\rho_A T_A)$ we have $\mathrm{tr}(\rho_{AB}(\log(\rho_A \otimes I_B))) = \mathrm{tr}(\rho_A \log \rho_A)$, and the above is $= -S(\rho_{AB}) - \mathrm{tr}(\rho_A \log \rho_A) - \mathrm{tr}(\rho_B \log \rho_B) = -S(\rho_{AB}) + S(\rho_A) + S(\rho_B) \geq 0$. So we see that entropy is additive for independent systems, but otherwise the entropy of a composite bipartite system is less than the sum of the entropies of its constituent subsystems. This is analagous to the classical property $H(X, Y) \leq H(X) + H(Y)$.

If a composite bipartite system is in a pure state ρ_{AB} then the von Neumann entropies of its subsystems (whose states are given by the reduced density matrices ρ_A, ρ_B) are equal, $S(\rho_A) = S(\rho_B)$. This follows directly from the Schmidt decomposition: the nonzero eigenvalues of the density matrices ρ_A, ρ_B are the same, and the entropy is determined completely by these eigenvalues.

Triangle inequality (Araki-Lieb inequality): for a bipartite system AB is state ρ_{AB}, $S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|$: we introduce a reference system R which purifies AB, i.e. the composite system ABR is in a pure state. By subadditivity $S(A, R) \leq S(A) + S(R)$; since ABR is in a pure state $S(A, B) = S(R)$ and $S(A, R) = S(B)$. Substituting and rearranging we have $S(A, B) \geq S(B) - S(A)$, and by symmetry also $S(A, B) \geq S(A) - S(B)$, so we have the result.

Let $\rho = \sum_i p_i \rho_i$ where ρ_i are density matrices which have support on orthogonal subspaces. Then $S(\sum_i p_i \rho_i) = H(\mathbf{p}) + \sum_i p_i S(\rho_i)$, where $\mathbf{p} = \{p_i\}$ and $H(\mathbf{p})$ is the corresponding Shannon entropy (see example sheet 3).

Strong subadditivity: for any state ρ_{ABC} of a tripartite system, $S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$. This is one of the most important and beautiful results in Quantum Information Theory, proven by Lieb and Ruskai, but beyond the scope of this course.

Consequences: 1. Conditioning reduces entropy: $S(A|BC) \leq S(A|B)$ 2. Discarding quantum systems never increases mutual information: $S(A : B) \leq S(A : B, C)$ 3. Quantum operations never increase mutual information: let Φ be a CPT map acting on the subsystem B alone, let $A'B'$ be the composite system AB after the action of Φ. Then $S(A' : B') \leq S(A : B)$: note $A' = A$. Go to a larger Hilbert space (Stinespring Dilation Theorem); Φ_B corresponds to some U_{BC} unitary for some ancilla C. We wlog take C initially in a pure state $|0\rangle\langle 0|_C$, so ρ_{ABC} is initally $\rho_A \otimes |0\rangle\langle 0|_C$. $\rho_{A'B'C'} = (I_A \otimes U_{BC})\rho_{ABC}(I_A \otimes U_{BC}^\dagger)$ and $\rho_{A'B'} = \mathrm{tr}_{C'}\rho_{A'B'C'}$. 1) We claim $S(A : B) = S(A : BC)$; the RHS here is $S(\rho_A) + S(\rho_{BC}) - S(\rho_{ABC})$; $\rho_{ABC} = \rho_{AB} \otimes \rho_C$ so $S(\rho_{ABC}) = S(\rho_{AB}) + S(\rho_C) = S(\rho_{AB})$ (as state of C is $|0\rangle\langle 0|_C$). Similarly $S(BC) = S(B)$ so this is just the idenity that $S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = S(A : B)$. 2) $S(A : BC) = S(A' : B'C')$ as $A' = A$ and $BC \to B'C'$ was unitary. 3) $S(A' : B'C') \geq S(A' : B')$ by the above.

Quantum Data Compression

A quantum information source is defined by $\{p_k, |\psi_k\rangle\}$ probabilities and corresponding signals, so can be characterised by $\{\rho, \mathcal{H}\}$ where $\rho = \sum p_k |\psi_k\rangle\langle\psi_k|$; note we do not generally have $\langle\psi_k|\psi_j\rangle = \delta_{jk}$, though we do have $p_k \geq 0, \sum_k p_k = 1$. We aim to find the limit of data compression; as always we use the source n times and consider $n \to \infty$. So we need to consider a sequence of density matrices $\rho^{(n)}$ acting on spaces $\mathcal{H}_n = \mathcal{H}^{\otimes n}$. $\rho^{(n)} = \sum_k p_k^{(n)} |\psi_k^{(n)}\rangle\langle\psi_k^{(n)}|$. Let $N_n = \dim \mathcal{H}_n$; this increases with n; in practice we will always be considering $\mathcal{H}$ a single qubit space so $N_n = 2^n$ [CHECK].

To compress, we have some function $\mathcal{C}^{(n)} : \mathcal{B}(\mathcal{H}_n) \to \mathcal{B}(\tilde{\mathcal{H}}_n)$ $|\psi_k^{(n)}\rangle\langle\psi_k^{(n)}| \mapsto \tilde{\rho}_n$. If $\dim \tilde{H}_n = d_c(n)$, we have compression if $d_c(n) < N_n = n$; compare this reduction in the dimension of the Hilbert space (i.e. the number of qubits) with a reduction in the number of classical bits. Our decompression function is $\mathcal{D}^{(n)} : \mathcal{B}(\tilde{\mathcal{H}}_n) \to \mathcal{B}(\mathcal{H}_n)$. We must have $\mathcal{C}^{(n)}, \mathcal{D}^{(n)}$ CPT. The rate R_n of a scheme $(\mathcal{C}^{(n)}, \mathcal{D}^{(n)})$ is $\frac{\log \dim \tilde{\mathcal{H}}_n}{\log \dim \mathcal{H}_n} = \frac{\log d_c(n)}{\log N_n} = \frac{\log d_c(n)}{n}$. $R_\infty = \lim_{n\to\infty} \log n = \lim_{n\to\infty} \frac{\log d_c(n)}{n}$.

We want to calculate $R_\infty = \lim_{n\to\infty} R^{(n)}$. Since $|\Psi_i\rangle, |\Psi_j\rangle$ are not necessarily orthogonal, i.e. not necessarily perfectly distinguishable, it would be "unfair" to use the same definition of reliable coding as in the classical case; in general perfectly reconstruction $|\Psi_k^{(n)}\rangle$ from its compressed version $\tilde{\rho}_k^{(n)}$ would be an impossible task. Instead, we require, for $\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)}) = \rho_k'^{(n)}$, $\rho_k'^{(n)}$ nearly indistinguishable from $|\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}|$. There are various measures of indistinguishability; we will use the ensemble average fidelity. This is defined as $F_n := \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)})|\Psi_k^{(n)}\rangle$. Note that this satisfies (the reader should verify) 1) $0 \leq F_n \leq 1$ 2) $F_n = 1 \Leftrightarrow \mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)}) = |\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}| \forall k$ [with $p_k > 0$]. Our compression-decompression scheme is reliable if $F_n \to 1$ as $n \to \infty$.

What is R_∞ for a reliable scheme? The key idea is that the Hilbert space $\mathcal{H}_n$ has a typical subspace; this notion was introduced by Ohya and Petz, but only applied to Quantum Information Theory by Schumacher, for a memoryless (IID) source - one for which $\mathcal{H}_n \simeq \mathcal{H}^{\otimes n}$ and $\rho^{(n)} = \pi^{\otimes n}$ for some density matrix π (acting on) $\mathcal{H}$. For such a source, let the spectral decompositions be $\pi = \sum_{i=1}^{d} q_i |\phi_i\rangle\langle\phi_i|$ ($d = \dim \mathcal{H}$), $\rho^{(n)} = \sum_{j=1}^{d^n} \lambda_j^{(n)} |\varphi_j^{(s)}\rangle\langle\varphi_j^{(n)}|$ (note that the $|\varphi_j\rangle$ are not the same as the $|\Psi_j\rangle$; in particular they are orthonormal), then we have $\lambda_j^{(n)} = q_{j_1} \dots q_{j_n}, |\varphi_j^{(n)}\rangle = |\phi_{j_1}\rangle \otimes \dots \otimes |\phi_{j_n}\rangle$. So we can identify j with the (classical) sequence $\mathbf{j} = (j_1, \dots, j_n)$. Then $\rho^{(n)} = \sum_{\mathbf{j}} \lambda_{\mathbf{j}}^{(n)} |\varphi_{\mathbf{j}}^{(n)}\rangle\langle\varphi_{\mathbf{j}}^{(n)}|$; the sum is over all possible sequences $\mathbf{j}$. So $S(\rho^{(n)}) = S(\pi^{\otimes n}) = nS(\pi)$. And $S(\pi) = H(\{q_i\})$.

We can try to define a typical set; recall that for an IID classical source, $U_1, \dots, U_n$ with $U_i \sim p(u), p(u_1, \dots, u_n) = \prod_{i=1}^n p(u_i)$. For a given $\epsilon > 0$, $\mathbf{u} = (u_1, \dots, u_n)$ is ϵ-typical if $| -\frac{1}{n} \log p(u_1, \dots, u_n) - H(U)| \leq \epsilon$, where $H(U) = -\sum p(u) \log p(u)$. So here we analagously define $T_\epsilon^{(s)}$, the set of typical sequences $\mathbf{j}$ (for given $\epsilon > 0$) by $| -\frac{1}{n} \log(q_{j_1} \dots q_{j_n}) - H(\{q_i\})| \leq \epsilon$ (recall $\lambda_{\mathbf{j}}^{(n)} = q_{j_1} \dots q_{j_n}$ is the probability of the sequence $\mathbf{j} = (j_1 \dots j_n)$). I.e. $\mathbf{j}$ is ϵ-typical if $| -\frac{1}{n} \log \lambda_{\mathbf{j}}^{(n)} - S(\pi)| \leq \epsilon$.

We proceed from $T_\epsilon^{(n)}$ to $\mathcal{T}_\epsilon^{(n)} \subset \mathcal{H}_n = \mathcal{H}^{\otimes n}$, the typical subspace, which is the

space spanned by those eigenvectors of $\rho^{(n)}$, $|\varphi_{\mathbf{j}}^{(n)}\rangle = |\phi_{j_1}\rangle\otimes\cdots\otimes|\phi_{j_n}\rangle$ for which $\mathbf{j} \in T_\epsilon^{(n)}$; note $\dim \mathcal{T}_\epsilon^{(n)} = |T_\epsilon^{(n)}|$. Then from the typical sequence theorem (for any $\delta > 0$ and n large enough, 1) $P\{T_\epsilon^{(n)}\} > 1-\delta$ and 2) $(1-\delta)2^{n(H(U)-\epsilon)} \le |T_\epsilon^{(n)}| \le 2^{n(H(U)+\epsilon)}$), we obtain the typical subspace theorem: for fixed $\epsilon > 0$, $\forall\delta > 0 \exists n_0(\delta) > 0$ such that $\forall n > n_0(\delta)$, a) $\mathrm{tr}(P_\epsilon^{(n)}\rho^{(n)}) > 1-\delta$ and b) $(1-\delta)2^{n(S(\pi)-\epsilon)} \le \dim \mathcal{T}_\epsilon^{(n)} \le 2^{n(S(\pi)+\epsilon)}$, where $P_\epsilon^{(n)}$ is the orthogonal projection operator onto $\mathcal{T}_\epsilon^{(n)}$. b) is immediate from 2) in the typical sequence theorem; for the first we claim $\mathrm{tr}(P_\epsilon^{(n)})$ is the probability of $\mathcal{T}_\epsilon^{(n)}$: it is $\mathrm{tr}(P_\epsilon^{(n)} \sum \lambda_{\mathbf{j}}^{(n)} |\varphi_{\mathbf{j}}^{(n)}\rangle\langle\varphi_{\mathbf{j}}^{(n)}|) = \sum_{\mathbf{j}} \lambda_{\mathbf{j}}^{(n)} \mathrm{tr}(P_\epsilon^{(n)}|\psi_{\mathbf{j}}^{(n)}\rangle\langle\psi_{\mathbf{j}}^{(n)}|P_\epsilon^{(n)}) = \sum_{\mathbf{j}\in T_\epsilon^{(n)}} \lambda_{\mathbf{j}}^{(n)}$ as required. But the probability of $\mathcal{T}_\epsilon^{(n)}$ is the same as the probability of $T_\epsilon^{(n)}$, so $> 1-\delta$ as required.

Conclusions: $\dim \mathcal{T}_\epsilon^{(n)} \approx 2^{nS(\pi)}$ and $\mathrm{tr}(P_\epsilon^{(n)}\rho^{(n)}) > 1-\delta$, so $\mathrm{tr}((I-P_\epsilon^{(n)})\rho^{(n)}) \le \delta$.

Our compression (technically compression-decompression) scheme is (sketch): we choose $\mathcal{C}^{(n)}$ such that its effect on a signal state $|\Psi_k^{(n)}\rangle = P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle + (I - P_\epsilon^{(n)})|\Psi_k^{(n)}\rangle$ is that the first term is left unchanged while the second ("junk") term is projected onto some fixed $|\phi_0\rangle \in \mathcal{T}_\epsilon^{(n)}$.

Define $\tilde{\rho}_k^{(n)} = \alpha_k^2 |\tilde{\Psi}_k^{(n)}\rangle\langle\tilde{\Psi}_k^{(n)}| + \beta_k^2 |\Phi_0\rangle\langle\Phi_0|$, where $|\tilde{\psi}_k^{(n)}\rangle = \frac{P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle}{\|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle\|}$ and $\alpha_k = \|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle\|$; $\beta_k = \|(I - P_\epsilon^{(n)})|\Psi_k^{(n)}\rangle\|$ and $|\Phi_0\rangle$ is a fixed vector in $\mathcal{T}_\epsilon^{(n)}$. We have $\tilde{\rho}_k^{(n)} \in \mathcal{T}_\epsilon^{(n)}$.

Then $\mathcal{D}^{(n)}$ is simply the extension of $\tilde{\rho}_k^{(n)}$ from the typical subspace to $\mathcal{H}^{\otimes n}$, $\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)}) = \tilde{\rho}_k^{(n)} \oplus 0$ (abusing notation slightly, we will also call this "matrix expanded with zeroes" $\tilde{\rho}_k^{(n)}$).

Recall we are using ensemble average fidelity F_n; here $F_n = \sum p_k^{(n)} \langle\Psi_k^{(n)}|\tilde{\rho}_k^{(n)}|\Psi_k^{(n)}\rangle$. (Since $\tilde{\rho}_k^{(n)} = |\tilde{\Psi}_k^{(n)}\rangle\langle\tilde{\Psi}_k^{(n)}| + \beta^2|\Phi_0\rangle\langle\Phi_0|$) this is $\sum p_k^{(n)}(\alpha_k^2 |\langle\Psi_k^{(n)}|\tilde{\Psi}_k^{(n)}\rangle|^2 + \beta_k^2 |\langle\Psi_k^{(n)}|\Phi_0\rangle|^2)$; the second term inside the bracket is ≥ 0, and $|\langle\Psi_k^{(n)}|\tilde{\Psi}_k^{(n)}\rangle|^2 = |\langle\Psi_k^{(n)}|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle|^2 = \|P_\epsilon^{(n)}\rangle\|^2 = \alpha_k^2$, so $F_n \ge \sum p_k^{(n)} \alpha_k^4$. Use that $(\alpha_k^2 - 1)^2 \ge 0$, i.e. $\alpha_k^4 \ge 2\alpha_k^2 - 1$, then $F_n \ge \sum_k p_k^{(n)}(2\alpha_k^2 - 1)$ or $F_n \ge 2\sum p_k^{(n)} \alpha_k^2 - 1$.

Schumacher Theorem

Let $\{\rho, \mathcal{H}_n\}$ be a memoryless quantum information source, $\mathcal{H}_n = \mathcal{H}^{\otimes n}$, $\rho_n = \pi^{\otimes n}$. Then 1) If $R > S(\pi)$ then there is a reliable compression scheme of rate R, 2) If $R < S(\pi)$ there is no such reliable scheme; we shall only proove the first part, as the second is somewhat fiddly. Recall $R = \frac{\log(\dim \tilde{H}_n)}{n}$ i.e. $\dim \tilde{\mathcal{H}}_n = 2^{nR}$. Choose $\epsilon > 0$ such that $R > S(\pi) + \epsilon$; for a given $\delta > 0$ consider n large enough and $\mathcal{T}_\epsilon^{(n)}$ the typical subspace of $\rho^{\otimes n}$ such that $\dim \mathcal{T}_\epsilon^{(n)} \le 2^{n(S(\pi)+\epsilon)} < 2^{nR} \Rightarrow \dim \mathcal{T}_\epsilon^{(n)} <$ [required $\dim \tilde{H}_n$]. $F_n \ge 2(\sum_k p_k^{(n)} \alpha_k^2) - 1 = 2\sum_k p_k^{(n)} \langle\Psi_k^{(n)}|P_\epsilon^{(n)}|\Psi_\epsilon^{(n)}\rangle - 1 = 2\mathrm{tr}(P_\epsilon^{(n)}\rho_n) - 1 \ge 2(1-\delta) - 1 = 1 - 2\delta$, so we have the result.

Quantum Error Correcting Codes

Consider a single qubit initially in a pure state $|\psi\rangle$, interacting with its environment initially in a pure state $|0_E\rangle$¿ The time-evolution is unitary, $U(|\psi_A\rangle \otimes |0_E\rangle)$. Since $|\psi\rangle = a|0\rangle + b|1\rangle$, the time evolution can be characterised by $U(|0\rangle \otimes |0_E\rangle) =$

$|0\rangle \otimes |e_{00}\rangle + |1\rangle \otimes |e_{01}\rangle, U(|1\rangle \otimes |0_E\rangle) = |0\rangle \otimes |e_{10}\rangle + |1\rangle \otimes |e_{11}\rangle$, for some states $|e_{ij}\rangle$ (not generally orthogonal, normalized or anything of the sort). Then $U(|\psi\rangle \otimes |0_E\rangle) = aU(|0\rangle \otimes |0_E\rangle) + bU(|1\rangle \otimes |0_E\rangle)$, which we can rearrange as $(a|0\rangle + b|1\rangle)\otimes\frac{1}{2}(|e_{00}\rangle+|e_{11}\rangle)+(a|0\rangle-b|1\rangle)\otimes\frac{1}{2}(|e_{00}\rangle-|e_{11}\rangle)+(a|1\rangle+b|0\rangle)\otimes\cdots+(a|1\rangle-b|0\rangle)\otimes\ldots$, which we write as $|\psi\rangle \otimes |e_0\rangle_E + \sigma_z|\psi\rangle \otimes |e_z\rangle_E + \sigma_x|\psi\rangle \otimes |e_x\rangle_E + \sigma_y|\psi\rangle \otimes |e_y\rangle_E$.

So heuristically, the effect of noise is one of: nothing $\sigma_0 = I$, a bit flip σ_x, phase flip σ_z or combined flip σ_y.

The unitary evolution of n qubits with environment is expressed in terms of the 4^n operators $\{\sigma_0, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$; we assume that errors are locally independent (i.e. errors on each of the n qubits are independent), and also sequentially independent (if we send the same qubit through multiple channels, the errors ocurring in each channel are independent).

If the initial state of the n qubits is $|\Psi\rangle$, environment initially $|0_E\rangle$ of course, then $U(|\Psi\rangle\otimes|0_E\rangle) = \sum_{a=1}^{4^n}(E_a|\Psi\rangle)\otimes|e_a\rangle$. The E_a are $2^n \times 2^n$ matrices, each a tensor product of n operators; the $|e_a\rangle$ are just some states. We have $E_a = E_a^\dagger, E_a^\dagger E_a = I$ since these are true for σ_i; these E_a are called Pauli operators.

We can relabel by $a \to \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. Each α_j is a "letter" from $\{I, X, Y, Z\}$, then we write $E_{\boldsymbol{\alpha}} = \bigotimes_{1\le j\le n} W_{\alpha_j}^{(j)}$ where W_{α_j} is respectively $\sigma_0 = I, \sigma_x, \sigma_y, \sigma_z$ depending on α_j and $^{(j)}$ indicates that it acts on the jth qubit (i.e. it is the matrix with a 2×2 σ-matrix in the jth position of the tensor product and identity elsewhere).

Given an $E_{\boldsymbol{\alpha}}$, what is its action on an n-qubit state? It suffices to consider $E_{\boldsymbol{\alpha}}|\boldsymbol{x}\rangle$ for $|\boldsymbol{x}\rangle$ a basis state of $\mathcal{H}^{\otimes n}$, $|\boldsymbol{x}\rangle = |x_1x_2\ldots x_n\rangle$ with $x_j \in \{0,1\}$.

Consider $\boldsymbol{\alpha}$ with nontrivial ($\neq I$) entries at locations $j_1, \ldots, j_r$, $\boldsymbol{\alpha} = (I\alpha_1 I\alpha_2 \ldots)$ (or similar). Write $E_{\boldsymbol{\alpha}} = E_{\alpha_1}^{j_1}E_{\alpha_2}^{j_2}\ldots E_{\alpha_r}^{j_r}$ (note this is the ordinary product of $2^n\times 2^n$ matrices).

$E_X^j|\boldsymbol{x}\rangle = |\boldsymbol{x} + \boldsymbol{e}_j\rangle$, where addition is mod 2 ($\boldsymbol{e}_j$ is the usual basis vector), because $\sigma_x|0\rangle = |1\rangle, \sigma_x|1\rangle = |0\rangle$ - this is a bit flip on the jth qubit. Similarly $E_Z^j|\boldsymbol{x}\rangle = (-1)^{x_j}|\boldsymbol{x}\rangle, E_Y^j|\boldsymbol{x}\rangle = i(-1)^{x_j}|\boldsymbol{x} + \boldsymbol{e}_j\rangle$.

Definition: the weight of a Pauly operator $E_{\boldsymbol{\alpha}}$ is the number of α_js which are $\neq I$, i.e. the number of local errors.

$E_{\boldsymbol{\alpha}}E_{\boldsymbol{\alpha}'} \propto E_{\boldsymbol{\alpha}\star\boldsymbol{\alpha}'}$; the proportionality constant is ± 1 or $\pm i$, and $\boldsymbol{\alpha}\star\boldsymbol{\alpha}'$ is componentwise product, $(\alpha_1\alpha_1', \alpha_2\alpha_2', \ldots)$ (where $\alpha\alpha'$ is defined to mimic the product of Pauli matrices: $I\alpha = \alpha I = \alpha, \alpha^2 = I, XY = Z, YZ = X$ etc).

A QECC χ, which we shall call an $[[n,k,d]]$ code or for now a $[[n,k]]$ code, is a linear subspace of dimension 2^k embedded in a space of dimension 2^n, for $n > k$. n is the block size of the code, k is the number of encoded qubits. E.g. $k = 1, n = 3$: χ_{rep} the quantum repitition code $|0\rangle \to |000\rangle, |1\rangle \to |111\rangle$ (But note that for this cote $|\psi\rangle = a|0\rangle + b|1\rangle \mapsto |\Psi\rangle = a|000\rangle + b|111\rangle$ not generally equal to $|\psi\rangle^{\otimes n}$ (so this does not violate the no cloning theorem)). Encoding is highly non-local - we have entanglement.

χ_{rep} protects against a single bit flip erro. The quantum analogue of the MBSC with probability p is a quantum channel Φ - a CPT map - given by $\Phi(\rho) = p\sigma_x\rho\sigma_x + (1-p)\rho$. (This is a CPT map: we can write $\phi(\rho) = \sum_{k=1}^2 A_k\rho A_k^\dagger$ by $A_1 = \sqrt{p}\sigma_x, A_2 = \sqrt{1-p}\sigma_0$).

There will be two parts to the error-correcting process: 1) syndrome dyagnosis - we will have to make a measurement to find out what error (if any) has

occurred. Note that since we do not want to disturb the state, this measurement will have to diagnose the error without telling us about the state. 2) Recovery - e.g. if we detect that σ_x acted on the 3rd qubit, we act with σ_x on the 3rd qubit to recover our original state.

Our syndrome diagnosis must be by a carefully chosen measurement that gives information about the error but no information about a and b. We perform a collective measurement on the three qubits: measure the four operators $A_0 = P_0 = |000\rangle\langle 000| + |111\rangle\langle 111|, A_1 = P_1 = |100\rangle\langle 100| + |011\rangle\langle 011|, \ldots$. The outcomes are syndromes S_0, S_1, S_2, S_3, each 1 or 0 (the P_i are projection operators, so when measuring P_i the projection operator onto the +1 eigenspace is P_i itself, and onto the 0 eigenspace is $I - P_i$). Also, the state $|\Psi'\rangle$ is unchanged, e.g. suppose we had a bit flip on the second qubit, $|\Psi'\rangle = a|010\rangle + b|101\rangle$. Then e.g. we measure P_1, so the outcome is 0 and the post measurement state is $\propto P_0^1|\Psi'\rangle = (I - P_1)|\Psi'\rangle = |\psi'\rangle$.

So we can determine whether there was no error or an error on the first, second or third qubit. In this case we discover an error on the second qubit. So then we act on $|\Psi'\rangle$ by $\sigma_0 \otimes \sigma_x \otimes \sigma_0$, to recover $|\psi\rangle = a|000\rangle + b|111\rangle$. Note we still have no information about a or b - a QECC can correct an error only if it does not disturb the superposition of the basis states, i.e. the coefficients must still be a, b.

What about a phase flip error, e.g. σ_z os the second qubit, $(\sigma_0 \otimes \sigma_z \otimes \sigma_0)|\Psi\rangle \to a|000\rangle - b|000\rangle$. χ_{rep} cannot correct a phase flip, as the reader may investigate.

Note that we could not use this method to obtain a [[2,1]] code for bit-flip errors; we would have $P_1 = |10\rangle\langle 10| + |01\rangle\langle 01| = P_2$, so we cannot distinguish between errors in the first or second bit.

The Schor code is a [[9,1]] code: $|0\rangle \mapsto |\mathbf{0}\rangle = \frac{1}{\sqrt{8}}(|000\rangle + |111\rangle)^{\otimes 3}, |1\rangle \to |\mathbf{1}\rangle = \frac{1}{\sqrt{8}}(|000\rangle - |111\rangle)^{\otimes 3}$; for convenience we write these as $|+\rangle|+\rangle|+\rangle$ and $|-\rangle|-\rangle|-\rangle$. Clearly this can correct a bit flip error. Consider a phase flip e.g. on the second qubit; we have $E_{\boldsymbol{\alpha}}$ (where $\boldsymbol{\alpha} = (IZI\ldots I)$) with $E_{\boldsymbol{\alpha}}|+\rangle|+\rangle|+\rangle = |-\rangle|+\rangle|+\rangle, E_{\boldsymbol{\alpha}}|-\rangle|-\rangle|-\rangle = |+\rangle|-\rangle|-\rangle$ (note that sometimes different error operators cause the same change to the codeword). Let $A = \sigma_x \otimes \sigma_x \otimes \sigma_x$, then $A|\pm\rangle = \pm|\pm\rangle$. We would like to measure e.g. $A \otimes I^{\otimes 3} \otimes I^{\otimes 3}$ etc., but this will not work; it would destroy the superposition. So instead we measure $E_{\boldsymbol{\alpha}}, E_{\boldsymbol{\beta}}$ where $\boldsymbol{\alpha} = XXXXXXIII, \boldsymbol{\beta} = IIIXXXXXX$. We have $E_{\boldsymbol{\alpha}}^2 = I = E_{\boldsymbol{\beta}}^2$, so the eigenvalues of each are ± 1. The reader may check $E_{\boldsymbol{\alpha}}|\Psi\rangle = |\Psi\rangle = E_{\boldsymbol{\beta}}|\Psi\rangle$. If e.g. $|\Psi'\rangle = a|-\rangle|+\rangle|+\rangle + b|+\rangle|-\rangle|-\rangle$, then $E_{\boldsymbol{b}}|\Psi'\rangle = |\Psi'\rangle, E_{\boldsymbol{\alpha}}|\Psi'\rangle = -|\Psi'\rangle$. Generally, S_α tells us whether the relative phase of the first and second sets is different, S_β does the same for the second and third, and together this tells us whether there have been no phase flips or in which sed of three the phase flip ocurred (assuming of course only one has occurred) - here, the first set of three. Then we can correct it, e.g. in this case if we act by $\sigma_z \otimes \sigma_0 \otimes \sigma_0$ on the first set of 3 qubits.

χ_{rep} is a non-degenerate code but χ_{shor} is degenerate; see later.

Clearly, if we only want to correct a phase flip with no possibility of bit flips, we can use $|0\rangle \mapsto |++\rangle, |1\rangle \mapsto |--\rangle$, where $|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$.

Definition and basic properties of QECC

Given χ to correct a set $\mathcal{E} = \{E_{\boldsymbol{\alpha}}\}$ of Pauli errors (e.g. χ_{rep} corrects $\{E_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} = III, XII, IXI, IIX\}$). We want a necessary and sufficient condition for error correction: suppose χ is an $[[n,k]]$ code, $\chi \subset \mathcal{H}^{\otimes n}$, $|\boldsymbol{i}\rangle, |\boldsymbol{j}\rangle$ mutually orthogonal basis codewords. It is necessary that the $E_{\boldsymbol{\alpha}}$s should not destroy the perfect distinguishability of such mutually orthogonal basis vectors: $\forall \boldsymbol{i} \neq \boldsymbol{j}, E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle \perp E_{\boldsymbol{\alpha}'}|\boldsymbol{j}\rangle$ i.e. $\langle \boldsymbol{j}|E^\dagger_{\boldsymbol{\alpha}'}E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle = 0$ (1) (and note the $E_{\boldsymbol{\alpha}}$ are self-adjoint, so we may or may not drop the †s). A sufficient condition is (2) $\langle j|E^\dagger_{\boldsymbol{\alpha}}E_{\boldsymbol{\alpha}}|i\rangle = \delta_{\alpha\alpha'}\delta_{ij}$, i.e. $E_{\boldsymbol{\alpha}}|i\rangle \perp E_{\boldsymbol{\alpha}'}|i\rangle$.

Suppose χ is such that (2) holds. A codeword $|\Psi\rangle = \sum_i c_i|\boldsymbol{i}\rangle$ passes through a noisy channel and is acted on by $E_{\boldsymbol{\alpha}}$, becoming $|\Psi'\rangle$. This lies $\in \chi^\perp$, since $\langle \boldsymbol{i}|\Psi'\rangle = 0 \forall$ basis codewords $|\boldsymbol{i}\rangle$. Then syndrome diagnosis is unambiguous, and we recover by $E_{\boldsymbol{\alpha}}|\Psi'\rangle = E_{\boldsymbol{\alpha}}E_{\boldsymbol{\alpha}}|\Psi\rangle = |\Psi\rangle$. This is called non-degenerate quantum error correction.

A general NASC for error correction is that $\langle \boldsymbol{j}|E^\dagger_{\boldsymbol{\alpha}'}E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle = c_{\alpha'\alpha}\langle j|i\rangle$ for some $c_{\alpha'\alpha} \in \mathbb{C}$ (independent of $\boldsymbol{i}, \boldsymbol{j}$ - i.e. $\langle i|(E^\dagger_{\alpha'}E_\alpha)|i\rangle$ is the same $\forall |i\rangle$). Then the non-degenerate case is $c_{\alpha\alpha'} = \delta_{\alpha\alpha'}$.

Recall the weight $w(\boldsymbol{\alpha})$ is the number of non-identity entries in $\boldsymbol{\alpha}$. If $\mathcal{E}$ contains all $E_{\boldsymbol{\alpha}}$ for $w(\boldsymbol{\alpha}) \leq$ some integer $\mathbb{E}$, we say χ is $\mathbb{E}$-error correcting. If $|\boldsymbol{i}\rangle, |\boldsymbol{j}\rangle \in \chi$ with $\langle \boldsymbol{i}|\boldsymbol{j}\rangle = 0$, χ a $\mathbb{E}$-error correcting code, then $\forall \boldsymbol{\alpha}, \boldsymbol{\alpha}'$ with $w(\boldsymbol{\alpha}), w(\boldsymbol{\alpha}') \leq \mathbb{E}$, $E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle, E_{\boldsymbol{\alpha}'}|\boldsymbol{j}\rangle$ orthogonal. The non-degenerate case is $E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle \perp E_{\boldsymbol{\alpha}'}|\boldsymbol{i}\rangle$ unless $\boldsymbol{\alpha} = \boldsymbol{\alpha}'$.

We say χ is D-error-detecting if $\forall E_{\boldsymbol{\alpha}}$ with $w(\boldsymbol{\alpha}) \leq D$, $\forall |\boldsymbol{i}\rangle, |\boldsymbol{j}\rangle \in \chi$, $\langle \boldsymbol{j}|E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle = r_{\boldsymbol{\alpha}}\langle \boldsymbol{j}|\boldsymbol{i}\rangle$ for some $r_{\boldsymbol{\alpha}} \in \mathbb{C}$, i.e. $E_{\boldsymbol{\alpha}}|i\rangle = r_{\boldsymbol{\alpha}}|i\rangle + |\phi_{\alpha i}\rangle$ for some $|\phi_{\alpha i}\rangle \in \chi^\perp$. The non-degenerate case is $\langle \boldsymbol{j}|E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle = r_\alpha\langle \boldsymbol{j}|\boldsymbol{i}\rangle$ where $r_\alpha = 0 \forall \alpha \neq \mathbb{I}$.

For non-degenerate error correction $E_{\boldsymbol{\alpha}}|i\rangle \in \chi^\perp \therefore \langle \boldsymbol{i}|E_{\boldsymbol{\alpha}}|\boldsymbol{i}\rangle = 0$.

The reader should check: A QECC is $\mathbb{E}$-correcting if $\forall |\psi\rangle, |\psi'\rangle \in \chi, \langle \psi'|E^\dagger_{\alpha'}E_\alpha|\psi\rangle = c_{\alpha'\alpha}\langle \psi'|\psi\rangle$, non-degenerate if $c_{\alpha'\alpha} = \delta_{\alpha'\alpha}$, and error detecting if $\langle \psi'|E_\alpha|\psi\rangle = r_\alpha\langle \psi'|\psi\rangle$.

Distance of a code χ: d (the 3rd term in the characterisation of a code as $[[n,k,d]]$). d is defined as the minimal weight of a Pauli operator $E_{\boldsymbol{\alpha}}$ such that $\exists |\boldsymbol{i}\rangle, |\boldsymbol{j}\rangle \in \chi$ with $\langle \boldsymbol{j}|E_\alpha|\boldsymbol{i}\rangle \neq r_\alpha\langle j|i\rangle$.

Theorem 1: 1) If χ corrects $\mathbb{E}$ errors it detects $2\mathbb{E}$ errors, 2) If χ is D-error detecting it is $\lfloor \frac{D}{2} \rfloor$ error detecting: for 1), $\forall \boldsymbol{\alpha}$ with $w(\boldsymbol{\alpha}) \leq 2\mathbb{E}$, we can write $E_{\boldsymbol{\alpha}} = E^\dagger_{\boldsymbol{\beta}'}E_{\boldsymbol{\beta}}$ (non-uniquely) with $w(\beta), w(\beta') \leq E$. Then $\langle \psi'|E_{\boldsymbol{\alpha}}|\psi\rangle = \langle \psi'|E_{\beta'}E_\beta|\psi\rangle = c_{\beta'\beta}\langle \psi'|\psi\rangle$ since the code is $\mathbb{E}$-error-correcting, so set $r_\alpha := c_{\beta'\beta}$. For 2), $\forall \boldsymbol{\alpha}, \boldsymbol{\alpha}'$ [of weight] $\leq \lfloor \frac{D}{2} \rfloor$, $E'_{\boldsymbol{\alpha}}E_{\boldsymbol{\alpha}} = k_{\alpha'\alpha}E_{\boldsymbol{\alpha}\star\boldsymbol{\alpha}'}$. But $w(\boldsymbol{\alpha} \star \boldsymbol{\alpha}') \leq D$, so $\langle \psi'|E^\dagger_{\boldsymbol{\alpha}'}E_{\boldsymbol{\alpha}}|\psi\rangle = k_{\boldsymbol{\alpha}\alpha'}\langle \psi'|E_{\boldsymbol{\alpha}\star\boldsymbol{\alpha}'}|\psi\rangle = k_{\boldsymbol{\alpha}\alpha'}r_{\boldsymbol{\alpha}\star\boldsymbol{\alpha}'}\langle \psi|prime|\psi\rangle$, so $c_{\alpha'\alpha} := k_{\alpha\alpha'}r_{\alpha\star\alpha'}$ works.

Theorem: an $[[n,k,d]]$ code detects $d-1$ errors and corrects $\frac{d-1}{2}$ errors.

Theorem: an $[[n,k,d]]$ code corrects $d-1$ errors of known location. Wlog assume the errors are on qubits $1, 2, \ldots, d-1$. Then $\forall E_{\boldsymbol{\alpha}}, E_{\boldsymbol{\alpha}'}$ of the form $\ldots I \ldots I$, $E^\dagger_{\alpha'}E_\alpha \propto E_{\alpha\star\alpha'}$, but $w(\boldsymbol{\alpha} \star \boldsymbol{\alpha}') \leq d-1$, so $\langle \psi'|E_{\alpha'}E_\alpha|\psi\rangle = k_{\alpha\alpha'}r_\beta\langle \psi|\psi\rangle$; call this $c_{\alpha\alpha|prime}\langle \psi'|\psi\rangle$.

Exercise: check these results for χ_{rep} and χ_{shor}.

Recall: A $[[n,k,d]]$ QECC χ is a 2^k-dimensional subspace of the n-qubit space $\mathcal{H}^{\otimes n}$ which can detect $d-1$ errors and correct $\lfloor \frac{d-1}{2} \rfloor$ errors. (This is also sometimes called a $[[n,k,t]]$ code where t is the number of errors it can correct.)

Quantum Hamming Bound

For a non-degenerate $[[n,k,t]]$ code: In a block of n qubits, there are $\binom{n}{j}$ possible ways for there to be errors on j qubits, so the total number of possible errors of weight $\leq t$ is $N(t) := \sum_{j=0}^{t} 3^j \binom{n}{j}$ (there being three possible errors X, Y, Z on a single qubit).

Recall: a non-degenerate code has $J_\alpha|i\rangle, E_{\alpha'}|j\rangle$ always linearly independent for $|i\rangle \neq |j\rangle \in \chi, \alpha \neq \alpha'$. So $\mathcal{H}^{\otimes n}$ must be large enough to accomodate $N(t)2^k$ linearly independent vectors, so $N(t)2^k \leq 2^n \Rightarrow N(t) = \sum_{j=0}^{t} 3^j \binom{n}{j} \leq 2^{n-k}$. This is the Quantum Hamming Bound.

Consider $k = 1, t = 1$. Then the QHB is $1 + 3n \leq 2^{n-1}$, satisfied by $n \geq 5$. This is exact: $1 + 15 = 16$, so any non-degenerate $[[5,1,3]]$ code (if such exists) will be perfect.

Does a degenerate $[[4,1,3]]$ code exist? (No, as will be proven on the example sheet using the no cloning theorem). There are many other bounds, e.g. the Quantum Singleton Bound (which is true for a general, possibly degenerate QECC): $n - k \geq 2(d-1)$; see the final example sheet for this course.

Recall: All possible quantum operations are given by CPT maps Φ. Such a Φ is called a Quantum channel. A channel is called unital if $\Phi(I) = I$. It is useful to consider a qubit channel as a map on the Bloch sphere.

We have already seen the bit flip channel $\Phi(\rho) = (1-p)\rho + p\sigma_x\rho\sigma_x$. We found the Krauss operators $A_1 = \sqrt{1-p}, A_2 = \sqrt{p}\sigma_x$. Write $\Phi(\rho) = \frac{1}{2}(I + \boldsymbol{s}' \cdot \boldsymbol{\sigma}), \rho = \boldsymbol{s} \cdot \boldsymbol{\sigma}$, then consider $\Phi : \boldsymbol{s} \to \boldsymbol{s}'$. $\boldsymbol{s}$ is called the spin polarization vector. Substituting we find (as the reader should check) $\boldsymbol{s}' = (s_x, (1-2p)s_y, (1-2p)s_z)$ - the Bloch sphere is deformed into an ellipsoid. Similarly, the reader should verify that the phase flip channel $\phi(\rho) = p\sigma_z\rho\sigma_z + (1-p)\rho$ corresponds to a compression of the sphere by a factor of $(1-2p)$ in the XY plane.

Depolarizing Channel

$\Phi(\rho) = (1-p)\rho + \frac{p}{3}(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z)$. $A_0 = \sqrt{1-p}\sigma_0, A_1 = \sqrt{\frac{p}{3}}\sigma_x, A_2 = \sqrt{\frac{p}{3}}\sigma_y, A_3 = \sqrt{\frac{p}{3}}\sigma_z$. This is unital: $\Phi(I) = I$. Acting on the Bloch sphere, $s \mapsto (\alpha s_x, \alpha s_y, \alpha s_z)$ where $\alpha = 1 - \frac{4p}{3}$ - the sphere contracts uniformly by this factor.

An alternative, and more common, formulation, is $\Phi(\rho) = (1-q)\rho + q\frac{I}{2}$. These are equivalent, because $\frac{1}{2} = \frac{1}{4}(\rho + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z$¡ so the above is $(1-q)\rho + q\frac{\rho}{4} + \frac{q}{4}\sum_{\alpha=x,y,z}\sigma_\alpha\rho\sigma_\alpha = (1 - \frac{3}{4}q)\rho + \frac{q}{4}\sum\sigma_\alpha\rho\sigma_\alpha$, i.e. $\frac{p}{3} = \frac{q}{4}, q = \frac{4}{3}p$. This second form generalizes to a qdit: for $\dim\mathcal{H} = d, \rho \in \mathcal{B}(\mathcal{H}), \Phi_d(\rho) = (1-q)\rho + q\frac{I}{d}$. Generalizing the first form is harder.

As a physical scenario, consider an atom A with two states, a ground state 0 and excited state 1 - a qubit. It is possible for there to be "spontaneous emission", where if the atom is in the excited state, with probability p it will emit a photon and shift to the lower state. We want to write this as a channel - the Amplitude Damping Channel.

Let $|0_A\rangle$ be the ground state of the atom, $|1_A\rangle$ the excited state. The environment is the EM field, initially in the vacuum state $|0_E\rangle$, and possibly entering a 1-photon state $|1_E\rangle$.

The channel is given by: $\Phi(\rho) = \sum_{i=1}^{2} A_i \rho A_i^\dagger$, where $A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, A_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$. Recall that these A_i act only on $\mathcal{H}_A$. The reader should check $A_2|0_A\rangle = 0, A_2|1_A\rangle = \sqrt{p}|0_A\rangle, A_1|0_A\rangle = |0_A\rangle, A_1|1_A\rangle = \sqrt{1-p}|1_A\rangle$, so $\Phi(\rho) = \begin{pmatrix} \rho_{00} + p\rho_{11} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{pmatrix}$.

The initial state $\rho \otimes |0_E\rangle\langle 0_E| = \rho_{00}|00\rangle\langle 00| + \rho_{01}|00\rangle\langle 01| + \rho_{10}|10\rangle\langle 00| + \rho_{11}|10\rangle\langle 10|$ We act by a unitary operator, $U(\dots)U^\dagger$; we know $U|00\rangle = |00\rangle, U|10\rangle = \sqrt{p}|01\rangle + \sqrt{1-p}|10\rangle$, which $= \rho_{00}|00\rangle\langle 00| + \rho_{01}|00\rangle(\sqrt{p}\langle 01| + \sqrt{1-p}\langle 01|) + \dots$. $\Phi(\rho) = \mathrm{tr}_E(\dots) = \rho_{00}|0\rangle\langle 0| + \rho_{01}\sqrt{1-p}|0\rangle\langle 1| + \dots$. Φ is not unital, $\Phi_{\mathrm{AD}}(I) \neq I$) (in fact it is $\begin{pmatrix} 1+p & 0 \\ 0 & 1-p \end{pmatrix}$.

Consider applying this charrel n times in succession. The probability of transition per unit time is q, so in a time δt will be $p = q\delta t$. Then the probability that an excited state <u>does not</u> decay in a time $t = n\delta t$ is $(1-p)^n = (1-q\delta t)^{\frac{t}{\delta t}} \overset{\delta t \to 0}{\to} e^{-qt}$. So $\Phi^n(\rho) \overset{n\to\infty}{\to} \begin{pmatrix} \rho_{00} + \rho_{11} & 0 \\ 0 & 0 \end{pmatrix}$, and this is a <u>pure state</u>. This is surprising - normally, Φ would add noise to a pure state, increasing $S(\rho)$ from 0 to some value ¿0. But here, we went from a mixed state with $S(\rho) > 0$ to a pure state $S(\rho) = 0$. Looking at it, we have really lost information in so doing - so $S(\rho)$ is not monotonous under CPT maps, and not a good characterization of quantum information.

Transmission of Classical Information through a Quantum Channel

The scenario is: Alice wants to transmit some mesage $X \sim p_x$ with $x \in J$. She encodes by $x \mapsto \rho_x$, in general a mixed state, then sends this through a noiseless quantum channel to Bob, who performs a POVM characterised by $\{E_y\}$, obtaining some outcome y. So the outcome variable Y is a classical random variable.

The probability of Bob measuring E_y is $\mathrm{tr}(E_y\rho_x)$. If x is the emitted symbol, the probability that Bob infers x correctly is then $P(Y = x \mid X = x) = \mathrm{tr}(E_x\rho_x)$.

How much information can B gain about X through his knowledge of Y? Clearly, $H(X : Y)$. So to get the maximum possible information, Bob chooses a measurement which maximises this.

Definition: the <u>accessible information</u> I_{acc} is the maximum information Bob can gain through any possible measurement, i.e. $\max H(X : Y)$ where the maximum is taken over all possible measurements.

We'll find an upper bound on this: the <u>Holero Bound</u> is that $I_{\mathrm{acc}} \leq \chi(\{p_x, \rho_x\})$, the Holero chi-quantity. I.e. we will show $H(X : Y) \leq \chi(\{p_x, \rho_x\})$ (1). We define $\chi(\{p_x, \rho_x\}) := S(\sum_x p_x\rho_x) - \sum p_x S(\rho_x)$, and define $\rho := \sum p_x\rho_x$. Notice that we have equality in (1) if all the ρ_x commute with each other and B performs a measurement in the simultaneous eigenbasis. Note also that $\chi(\{p_x, \rho_x\})$ depends not only on the average state ρ, but also on its preparation in terms of p_x and ρ_x (aside: $S(\rho)$ for $\rho = \{p_x, |\psi_x\rangle\langle\psi_x|\} = \sum p_x|\psi_x\rangle\langle\psi_x| = \sum \lambda_i|\phi_i\rangle\langle\phi_i|$ depends only on ρ, not on the particular choice of decomposition). Write $\mathcal{E} = \{p_x, \rho_x\}$ and talk

about $\chi(\mathcal{E})$. We have $\chi(\mathcal{E}) \to S(\rho)$ if ρ is an ensemble of pure states, i.e. all the ρ_x are pure.

Proof of the Holero Bound: We'll use strong subadditivity (SSA): $S(ABC) + S(B) \leq S(AB) + S(BC)$. We saw that this implies (1) $S(A|BC) \leq S(A|B)$ - "conditioning reduces entropy", (2) $S(A:B) \leq S(A:BC)$ - "discarding a quantum system cannot increase mutual information", and (3) $S(A':B') \leq S(A:B)$ for $\rho_{A'B'} = (I_A \otimes \Phi_B)\rho_{AB}$ - "quantum operations cannot increase mutual information".

(I) Embed X into a dummy quantum system A, with $\{|x\rangle\}_{x\in J}$ an ONB for $\mathcal{H}_A$ - this is a "register". (II) Let Q be the quantum system in whose state ρ_x Alice encodes x. (III) B is the quantum system representing Bob's measuring device. The initial state is wlog $|0\rangle\langle 0|_B$, uncorrelated with A.

AQB is now a tripartite system, with state space $\mathcal{H}_A \otimes \mathcal{H}_Q \otimes \mathcal{H}_B$. The initial state $\rho_{AQB} = \sum_x p_x |x\rangle\langle x|_A \otimes \rho_{x_Q} \otimes |0\rangle\langle 0|_B$. We can write this as $\rho_{AQ} \otimes |0\rangle\langle 0|_B$, uncorrelated.

Bob will perform a measurement. A measurement characterised by M_k can be seen as a CPT map if we store the outcome in an ancilla rather than looking at it: $\Phi(\rho \otimes |0\rangle\langle 0|) = \sum_k M_k \rho M_k^\dagger \otimes |k\rangle\langle k|$. So define: for σ_Q a state of Q, $\Phi(\sigma_Q \otimes |0\rangle\langle 0|_B) = \sum_y \sqrt{E_y}\sigma_Q\sqrt{E_y} \otimes |y\rangle\langle y|$ (one way to write $E_y = M_y^\dagger M_y$ is by $M_y = \sqrt{E_y}$). Exercise: the reader may prove this is a CPT map, by writing the RHS as $\sum_y A_y(\sigma_Q \otimes |0\rangle\langle 0|)A_y^\dagger$ with $\sum_y A_y^\dagger A_y = I$.

We have $\Phi : AQB \to A'Q'B'$; $\rho_{A'Q'B'} = \sum_{x,y} p_x |x\rangle\langle x| \otimes \sqrt{E_y}\rho_x\sqrt{E_y} \otimes |y\rangle\langle y|$.

1) $S(A:Q) = S(A:QB)$, as the reader should check, because $\rho_{AQB} = \rho_{AQ} \otimes |0\rangle\langle 0|_B$. 2) $S(A:QB) \geq S(A':Q'B')$. 3)$S(A|prime:Q'B') \geq S(A':B')$. So $S(A':B') \leq S(A:Q)$, and this is the Holero bound - the LHS is $H(X:Y)$ and the RHS is $\chi(\{p_x,\rho_x\})$. ($S(A:Q) = S(A) + S(Q) - S(AQ)$; $S(A) = H(\{p_x\})$ since $\rho_A = \sum_x |x\rangle\langle x|$, $\rho_Q \sum p_x\rho_x = \rho$ so $S(Q) = s(\rho)$, and $\rho_{AQ} = \sum p_x |x\rangle\langle x| \otimes \rho_x$. $S(AQ) = H(\{p_x\}) + \sum_x p_x S(\rho_x)$ (using the third example sheet: if $\rho = \sum p_i\rho_i$ where the ρ_i have mutually orthogonal supports, then $S(\rho) = H(\{p_i\}) + \sum p_i S(\rho_i)$), so $S(A:Q) = H(\{p_x\}) + S(\rho) - H(\{p_x\}) - \sum p_x S(\rho_x) = \chi(\{p_x\rho_x\})$. For the LHS, $S(A':B')$: $\rho_{A'B'Q'} = \sum p_x |x\rangle\langle x| \otimes \sqrt{E_y}\rho_x\sqrt{E_y} \otimes |y\rangle\langle y|$. We want $S(A') + S(B') - S(A'B')$; $\mathrm{tr}(\sqrt{E_y}\rho_x\sqrt{E_y}) = \mathrm{tr}(E_y\rho_x) = p(y \mid x)$, so $\rho_{A'B'} = \sum_{x,y} p_x p(y \mid x)|x\rangle\langle x| \otimes |y\rangle\langle y| = \sum_{x,y} p(x,y)|xy\rangle\langle xy|$, where $|xy\rangle = |x\rangle \otimes |y\rangle$. The $|xy\rangle$ form an ONB for $\mathcal{H}_A \otimes \mathcal{H}_B$, so $S(A'B')$ is just $H(XY)$. $\rho_{A'} = \sum_{x,y} p(x,y)|x\rangle\langle x| = \sum_x p(x)|x\rangle\langle x|$, $\rho_{B'} = \sum_y p_y |y\rangle\langle y|$. So $S(A') = H(X)$ $(= H(\{p_x\}))$, $S(B') = H(Y)$ as required.)

Properties of the χ quantity $\chi(\{p_x,\rho_x\}) := S(\sum p_x\rho_x) - \sum_x p_x S(\rho_x)$:

I) Non-negativity $\chi(\mathcal{E}) \geq 0$, by concavity of von Neumann entropy $S(\sum p_x\rho_x) \geq \sum p_x S(\rho_x)$.

II) We can express it as a relative entropy: $\chi = \sum_x p_x S(\rho_x\|\rho)$, since $S(\rho_x\|\rho) = \mathrm{tr}(\rho_x \log \rho_x) - \mathrm{tr}(\rho_x \log \rho)$ $\therefore$ $\sum_x p_x S(\rho_x\|\rho) = -\sum_x p_x S(\rho_x) - \sum_x p_x \mathrm{tr}\rho_x \log \rho = -\sum_x p_x S(\rho_x) - \mathrm{tr}((\sum_x p_x\rho_x)\log\rho) = -\sum_x p_x S(\rho_x) - \mathrm{tr}(\rho\log\rho) = S(\rho - \sum_x p_x S(\rho_x)$

Lindblad-Uhlmann monotonicity: (A) $S(\Phi(\rho)\|\Phi(\sigma)) \leq S(\rho\|\sigma) \forall$ CPT maps Φ; see the example sheet. So compare $\chi(\mathcal{E})$ with $\chi(\mathcal{E}')$ where $\mathcal{E}' = \{p_x, \Phi(\rho_x)\}$. $\chi(\mathcal{E}) = \sum_x p_x S(\rho_x\|\rho)$. (a) implies this is $\geq \sum_x p_x S(\Phi(\rho_x)\|\Phi(\rho)) = \chi(\mathcal{E}')$, i.e. the Hulero χ-quantity cannot increase under any CPT map (recall this was not so for von Neumann entropy $S(\rho)$).

Recall: Alice encodes a classical RV $X \sim p_x, x \in J$ as a quantum state $x \mapsto \rho_x$, which is transmitted to Bob who performs a POVM $\{E_y\}$ and obtains a resulting RV Y. The maximum amount of information Bob can get is $I_{\mathrm{acc}} \leq \chi(\{p_x,\rho_x\})$. If

the noiseless channel between Alice and Bob is replaced by a noisy channel Φ, this becomes $\chi(\{p_x, \Phi(\rho_x)\})$.

Our Q. channel is assumed memoryless $\Phi^{(n)} = \Phi^{\otimes n}$. For such a channel, product state inputs $\rho_1 \otimes \cdots \otimes \rho_n$ remain product states $\Phi(\rho_1) \otimes \cdots \otimes \Phi(\rho_n)$. What is the maximum amount of (classical) information that we can send through a memoryless channel per use of it? (Consider $n \to \infty$ as before) (Channels with memory are much harder, and an open research area).

Alice has a set $\mathcal{M}$ of classical messages. Say Φ is a qubit channel, $M \in \mathcal{M} \to \rho_M^{(n)}$, M encoded as a state of n qubits. This is sent through the channel, $\to \sigma_m^{(n)} = \Phi^{\otimes n}(\rho_M^{(n)})$. Bob performs a POVM, a collective measurement on n qubits, characterised by $\{E_M^{(n)}\}$¿ Say A has sent M; the probability of an error is $1-\mathrm{tr}(E_M^{(n)}\sigma_M^{(n)})$. What is the average probability of an error? Assume the messages are equiprobable, then $p_{\mathrm{av}}^{(n)} = \frac{1}{|M|}\sum_{M\in\mathcal{M}}(1 - \mathrm{tr}(E_M^{(n)}\sigma_M^{(n)}))$. We say information transmission is reliable if $p_{\mathrm{av}}^{(n)} \to 0$ as $n \to \infty$.

The limiting rate of information transmission is $R := \lim_{n\to\infty} \frac{\log|\mathcal{M}|}{n}$. We say a rate R is achieveable if there is an encoding/decoding scheme of rate R such that $p_{\mathrm{av}}^{(n)} \to 0$ as $n \to \infty$. The capacity $C_{\mathrm{cl}}(\Phi) = \sup R$.

Recall: For a classical memoryless channel, $C = \max_{p_x} H(X : Y)$. For a quantum channel, there are various scenarios: is information classical or quantum? Is the input $\rho^{(n)}$ a product state or entangled? Is the output measurement on single qubits or collective?

Holevo-Schumacher-Westmoreland: For classical messages, product state inputs, and collective measurement, the "product state capacity" $C^{(1)}(\Phi) = \max_{p_x,\rho_x} \chi(\{p_x, \Phi(\rho_x)\})$. This is often written as $\chi^\star(\Phi)$, and sometimes called the Holevo capacity.

Additivity, a major new result from this August: $C_{\mathrm{classical}}(\Phi) =\sim_{n\to\infty} \frac{1}{n}\chi^\star(\Phi^{\otimes n})$.

Can the classical capacity of a memoryless Q. channel increase by entangled inputs? The conjecture is addativity of the $\chi^\star(\Phi) : \chi^\star(\Phi_1\otimes\Phi_2) = \chi^\star(\Phi_1)+\chi^\star(\Phi_2)$. Exercise (three lines' worth): $\chi^\star$ is superaddative $\chi^\star(\Phi_1\otimes\Phi_2) \geq \chi^\star(\Phi_1)+\chi^\star(\Phi_2)$. This tells us $C_{\mathrm{classical}}(\Phi) \geq \lim_{n\to\infty} \frac{1}{n} n\chi^\star(\Phi) = \chi^\star(\Phi)$; addativity would give equality here, which would imply classical capacity cannot be increased by entangled inputs; it would just be $C^{(1)}(\Phi)$. However, a counterexample to addativity was recently round at Los Alamos.

The reader should try to prove: HSW implies that every non-constant quantum channel can be used to transmit classical information.

This seems tobe the end of the course.