Topics in Quantum Information Theory

Lent Term 2006

Sebastian Ahnert

Lecture notes on: http://www.tcm.phy.cam.ac.uk/∼sea31/
Corrections and comments welcome under: sea31@cam.ac.uk

Warning: These notes are designed to assist lectures given at a blackboard, and therefore do not contain some of the more detailed explanations and figures given in the actual lectures.

1 From Hamming Space to Hilbert Space

1.1 Hamming Space

Bit strings are the simplest possible way of encoding information, which is why it is the most common representation of information in information theory.

The *Hamming space* of dimension n is the space of bit strings of length n, denoted as $\{0,1\}^n$.

It is a discrete space with 2^n points, which can be mapped to the vertices of an n-dimensional hypercube. By moving along an edge of this hypercube we change ('flip') a single bit in the bit string.

1.2 Hamming distance

The **Hamming distance** between two bit strings is the number of digits in which they differ.

More formally, but no more complicated:

$$d(\mathbf{x},\mathbf{y}) = \#\{i : x_i \neq y_i\} \quad \mathbf{x},\mathbf{y} \in \{0,1\}^n$$

The Hamming distance of a bit string $\mathbf{x}$ from the origin $\mathbf{0}$ – and therefore, the number of ones in it – is its *weight* $w(\mathbf{x})$:

$$w(\mathbf{x}) = d(\mathbf{x},\mathbf{0})$$

1.3 Shannon entropy

Consider a more general string of symbols taken from an alphabet of size m.

Furthermore let's assume that the jth symbol of the alphabet has a probability p_j of occurring in the string.

Then the *amount of information gained* by seeing the jth symbol of the alphabet when it occurs in the bit string is:

$$I_j = -\log_2 p_j$$

In other words, rare characters gives a lot of information. As an example consider as strings the words of the English language, and in particular two words beginning with:

$$xyl\ldots$$

$$pro\ldots$$

The first word, with rare letters, is much easier to complete, since we have a lot more information about the rest of the word.

The **Shannon entropy** S of a general string of length n containing m different symbols (for bitstrings $m = 2$) is the *average amount of information* I_j *per character* contained in the string:

$$S = \sum_j^m p_j I_j = -\sum_j^m p_j \log_2 p_j$$

Note that $0 \leq S \leq \log_2 m$.

The Shannon entropy S tells us how compressible a string of symbols is. We will return to this issue in the next lecture in more detail.

1.4 Joint entropy

Consider two strings of symbols $\mathbf{a}$ and $\mathbf{b}$, and consider the nth letter in both of them.

The probability that the nth letter in string $\mathbf{a}$ will be the jth letter of the alphabet and the nth letter in string $\mathbf{b}$ will be the kth letter of the alphabet gives us the joint probability distribution p_{jk}, and thus a joint entropy $S(\mathbf{a}, \mathbf{b})$:

$$S(\mathbf{a}, \mathbf{b}) = -\sum_{j,k} p_{jk} \log_2 p_{jk}$$

If the probabilities for the two strings are independent, then $p_{jk} = p_j^a p_k^b$ and the joint entropy is just the sum of the Shannon entropies of the strings, i.e. $S(\mathbf{a}, \mathbf{b}) = S(\mathbf{a}) + S(\mathbf{b})$.

1.5 Conditional entropy

Similarly, if there is a conditional probability $p(j|k)$ that if the nth character of string $\mathbf{b}$ is symbol k in the alphabet, the nth symbol in $\mathbf{a}$ will be symbol j, then we can define a conditional entropy $S(\mathbf{a}|\mathbf{b})$:

$$S(\mathbf{a}|\mathbf{b}) = -\sum_{j,k} p_{jk} \log_2 p_{j|k}$$

Note the *joint* probability inside the sum, which can be understood from the point of view of average information gain. Also, using Bayes' theorem we obtain:

$$S(\mathbf{a}|\mathbf{b}) = -\sum_{j,k} p_{jk} (\log_2 p_{jk} - \log_2 p_k) = S(\mathbf{a}, \mathbf{b}) - S(\mathbf{b})$$

1.6 Mutual information

The **mutual information** M is an important measure which indicates how much information is shared between two random variables, or in our language, two strings of symbols with probability distributions p_j^a and p_k^b and with a joint distribution p_{jk}. Hence $M(\mathbf{a}, \mathbf{b})$ is defined as:

$$M(\mathbf{a}, \mathbf{b}) = S(\mathbf{a}) + S(\mathbf{b}) - S(\mathbf{a}, \mathbf{b})$$

Note that $M(\mathbf{a}, \mathbf{b}) = M(\mathbf{b}, \mathbf{a})$, and that $M = 0$ for independent distributions.

1.7 Relative entropy

Relative entropy is a distance measure between probability distributions. For two probability distributions $\{p_j\}$ and $\{q_k\}$ it is defined as:

$$D(p||q) = \sum_l p_l \log_2 \left(\frac{p_l}{q_l} \right)$$

Note that $D(p||q) \neq D(q||p)$ and, less trivially, $D(p||q) \geq 0$.

To gain a better understanding of this quantity we can rewrite it as:

$$D(p||q) = \sum_l p_l (\log_2 p_l - \log_2 q_l)$$

which is minus the *average difference* of the information gain between the two distributions *when sampling one of the distributions*.

Example: The letter u is about three times more common in French than in English. Therefore, when reading a u in an English word, one gains $\log_2 3$ more bits of information, but this happens with a smaller probability $p_E(u) \approx \frac{1}{3} p_F(u)$. In a French sentence, we lose $\log_2 3$ bits of information, and this is much more likely to happen, as the probability of a u is higher.

1.8 The qubit

In quantum information theory, the equivalent of the bit is the **qubit**. Instead of encoding information using 0 and 1, it is encoded using two orthogonal quantum states, $|0\rangle$ and $|1\rangle$.

Unlike classical bits, qubits can exist in a superposition state:

$$|\Psi\rangle = a|0\rangle + b|1\rangle$$

where $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$. We can write these states as vectors in two-dimensional Hilbert space $\mathcal{H}$:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad |\Psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \quad \langle\Psi| = \begin{pmatrix} a & b \end{pmatrix}$$

1.9 The physical qubit

Two possible qubit states of physical systems:

- spin-$\frac{1}{2}$ particles, e.g. $|0\rangle = |\uparrow\rangle$ and $|1\rangle = |\downarrow\rangle$
- photon polarization, e.g. $|0\rangle = |H\rangle$ and $|1\rangle = |V\rangle$

There are many proposals of how to realize quantum computers, involving ion traps, semiconductors, cavities, nuclear magnetic resonance, 2D electron gases, single photons and many more. So far there is no obvious choice. It is important to find solutions which are **easily scalable**, have a **long decoherence time**.

1.10 Hilbert space

The Hilbert space $\mathcal{H}_n$ of n qubits is of dimension 2^n and is the n-fold tensor product of the two-dimensional Hilbert space $\mathcal{H}$:

$$\mathcal{H}_n = \mathcal{H}^{\otimes n} \equiv \underbrace{\mathcal{H} \otimes \mathcal{H} \otimes \ldots \otimes \mathcal{H}}_{n \text{ times}}$$

For example, the general two qubit vector is:

$$|\Psi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

where $a, b, c, d \in \mathbb{C}$ and $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$.

1.11 The density matrix

A density matrix ρ is a representation of a quantum state or a statistical ensemble of quantum states. It can also be used to describe part of a composite system.

The quantum states we have talked about so far are called **pure states**. The density matrix of a pure state $|\Psi\rangle$ is given by the *outer* product of the state vector with itself:

$$\rho = |\Psi\rangle\langle\Psi| = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} a^* & b^* \end{pmatrix} = \begin{pmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{pmatrix}$$

where we have used the general two-dimensional state vector of $|\Psi\rangle$.

1.12 Mixed states

Density matrices can also describe **ensembles** of pure states $\{|\Psi_i\rangle\}$ with probabilities $\{p_i\}$ by constructing a linear combination of pure states:

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|$$

where $\sum_i p_i = 1$.

These are **mixed states**, and if $\rho = \frac{\mathbf{I}}{n}$, they are **maximally mixed**, without any quantum superpositions.

1.13 Properties of density matrices

Recall:

$$\rho = \textstyle\sum_i p_i |\Psi_i\rangle\langle\Psi_i| \qquad \sum_i p_i = 1$$

$$|\Psi_i\rangle\langle\Psi_i| = \begin{pmatrix} |a_i|^2 & a_i b_i^* \\ a_i^* b_i & |b_i|^2 \end{pmatrix} \quad |a_i|^2 + |b_i|^2 = 1$$

From this it follows that ρ has the following properties:

$\mathrm{tr}\rho = 1$ (unit trace)

$\rho = \rho^\dagger$ (Hermitian)

$\langle\phi|\rho|\phi\rangle \geq 0 \ \forall \ |\phi\rangle \in \mathcal{H}$ (positive definite)

1.14 Bloch sphere

We can rewrite any density matrix in terms of a linear combination of Pauli matrices, and thus in terms of a **Bloch vector n**:

$$\sum_i p_i \begin{pmatrix} |a_i|^2 & a_i b_i^* \\ a_i^* b_i & |b_i|^2 \end{pmatrix} = \frac{1}{2}(\mathbf{I} + n_x\sigma_x + n_y\sigma_y + n_z\sigma_z) = \frac{1}{2}(\mathbf{I} + \mathbf{n}\cdot\boldsymbol{\sigma})$$

where:

$$\mathbf{n} = \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \sum_i p_i \begin{pmatrix} 2\,\mathrm{Re}(a_i^* b_i) \\ 2\,\mathrm{Im}(a_i^* b_i) \\ (|a_i|^2 - |b_i|^2) \end{pmatrix}$$

Note that $|\mathbf{n}| \leq \mathbf{1}$ (with equality for pure matrices) so that the Bloch vector lies inside the unit sphere around the origin, the **Bloch sphere**.

1.15 Eigenvalues of the density matrix

As $\mathrm{tr}\rho = 1$ and ρ is positive definite and Hermitian, the eigenvalues $\{\lambda_i\}$ of ρ obey $\forall i \ : \ 0 \leq \lambda_i \in \mathbb{R} \leq 1$ and $\sum_i \lambda_i = 1$.

In other words, they are a *probability distribution*, corresponding to the coefficients of the expansion in terms of the orthogonal eigenbasis projectors$|\Phi_i\rangle\langle\Phi_i|$, which is just another of many possible expansions of a given density matrix ρ.

$$\rho = \sum_j p_j |\Psi_j\rangle\langle\Psi_j| = \sum_i^n \lambda_i |\Phi_i\rangle\langle\Phi_i|$$

where n is the dimension of ρ.

In the case of single qubits we can visualize orthogonal states as antiparallel Bloch vectors.

Therefore, the Bloch vector of a mixed state ρ points in the same direction as the eigenstate with the *larger* eigenvalue, and the length of the Bloch vector is the difference between the eigenvalues $|\lambda_1 - \lambda_2|$.

1.16 Von Neumann entropy

We have seen that the eigenvalues are probabilities in the expansion of the orthogonal eigenstate projectors. Hence the distribution of these probabilities gives a measure of *unpredictability*. If the distribution is uniform, i.e. $\lambda_i = \frac{1}{n} \forall\, i$, we have the maximally mixed case and cannot make any predictions. For a pure state we have one eigenvalue $\lambda_j = 1$ and all others are zero ($\lambda_i = 0 \forall\, i \neq j$).

The Shannon entropy of this probability distribution is the **von Neumann entropy** which is a very important quantity in quantum information theory:

$$S(\rho) = -\sum_i^n \lambda_i \log \lambda_i = \mathrm{tr}(-\rho \log \rho)$$

the last equality can be shown by considering a polynomial expansion of the log function and the fact that $\mathrm{tr}(\rho) = \mathrm{tr}(\mathbf{U}^\dagger \rho \mathbf{U})$.

1.17 Expectation values

For an observable M and a density matrix ρ the expectation value is given by:

$$\begin{aligned} \mathrm{tr}(M\rho) &= \mathrm{tr}(M \sum_i p_i |\Psi_i\rangle\langle\Psi_i|) = \sum_i p_i \mathrm{tr}(M|\Psi_i\rangle\langle\Psi_i|) \\ &= \sum_i p_i \mathrm{tr}(\langle\Psi_i|M|\Psi_i\rangle) = \sum_i p_i \langle\Psi_i|M|\Psi_i\rangle \end{aligned} \tag{1}$$

which is what we would expect (!) for the average expectation value for an ensemble of states ρ.

1.18 The partial trace

Now we want to obtain the expectation value of an observable in a subsystem of a composite system. Consider a two-qubit composite system described by a density matrix $\rho_{AB} \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. Our observable in subsystem A is denoted by

the operator $M_A \otimes \mathbf{1}_B$ where the operator $\mathbf{M}_A$ only acts in subsystem A, while $\mathbf{1}_B$ leaves subsystem B unchanged.

For a pure composite density matrix $\rho_{AB} = |\Psi\rangle_{AB}\langle\Psi|_{AB}$ where $|\Psi\rangle_{AB} = \sum_{\alpha,\beta} c_{\alpha\beta}|\alpha_A\rangle \otimes |\beta_B\rangle$ and $\{|\alpha_A\rangle\}$ and $\{|\beta_B\rangle\}$ are orthogonal bases in $\mathcal{H}_A$ and $\mathcal{H}_B$.

$$\begin{aligned} \mathrm{tr}\big((M_A \otimes \mathbf{1}_B)\rho_{AB}\big) &= \mathrm{tr}\left((M_A \otimes \mathbf{1}_B)\sum_{\alpha',\beta'} c_{\alpha'\beta'}|\alpha'_A\rangle \otimes |\beta'_B\rangle \sum_{\alpha,\beta} c^*_{\alpha\beta}\langle\alpha_A| \otimes \langle\beta_B|\right) \\ &= \mathrm{tr}\left(M_A \sum_{\alpha,\alpha',\beta} c_{\alpha'\beta}c^*_{\alpha\beta}|\alpha'_A\rangle\langle\alpha_A|\right) = \mathrm{tr}(M_A\rho_A) \end{aligned} \tag{2}$$

where we have defined the **partial trace** as

$$\rho_A = \mathrm{tr}_B(\rho_{AB}) = \sum_{\alpha,\alpha',\beta} c_{\alpha'\beta}c^*_{\alpha\beta}|\alpha'_A\rangle\langle\alpha_A|$$

1.19 Partial trace - example

Consider one of the four *Bell states*:

$$|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Then:

$$\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|)$$

and:

$$\rho_A = \mathrm{tr}_B\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) = \frac{\mathbf{I}_A}{2}$$

Hence by measuring a subsystem of the pure state ρ_{AB} we have turned the remaining part of the system into a (maximally) mixed state. This transformation is closely connected with *entanglement* which we shall return to later.

1.20 No-cloning theorem

Unlike classical information, quantum information cannot be copied perfectly. This result is the **no-cloning theorem**, which states that in any Hilbert space $\mathcal{H}$ (of any dimension) there exists no unitary operator $\mathbf{U}$ such that

$$\mathbf{U}(\phi \otimes \psi) = (\phi \otimes \phi) \;\; \forall \phi, \psi \in \mathcal{H}$$

Proof: Consider $\phi \to c\phi$ with $c \in \mathbb{C}$. The LHS becomes linear in c while the RHS becomes quadratic.

2 Entanglement and Non-Locality

2.1 Entanglement

Entanglement is a property of a quantum state of more than one qubit. In general a state $|\Psi\rangle_{AB}$ of two qubits A and B is **entangled** if it is not **separable**, i.e. *cannot* be written as the tensor product of two single particle states:

$$|\Psi\rangle_{AB} \neq |\Psi\rangle_A \otimes |\Psi\rangle_B \ \ \forall |\Psi\rangle_A, |\Psi\rangle_B$$

An example are the four Bell states

$$|\Psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \quad |\Phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle)$$

which form an orthogonal basis in two-qubit Hilbert space.

2.2 Subadditivity

Recall the **von Neumann entropy**:

$$S(\rho) = -\sum_i^m \lambda_i \log \lambda_i = \mathrm{tr}(-\rho \log \rho)$$

where $\{\lambda_i\}$ are the eigenvalues of ρ. For a pure state, $S(\rho) = 0$, for a maximally mixed state $\rho = \frac{\mathbf{I}}{m}$ and $S(\rho) = \log m$. For a composite system ρ_{AB} we can calculate a total entropy $S(\rho_{AB}) = \mathrm{tr}(-\rho_{AB} \log \rho_{AB})$ and the entropy of a subsystem $S(\rho_A) = \mathrm{tr}(-\rho_A \log \rho_A)$. The total quantum entropy is **subadditive**, just like the classical entropy:

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$$

2.3 Entropy and Entanglement

Unlike the classical *conditional* entropy $S(\mathbf{a}|\mathbf{b}) = S(\mathbf{a}, \mathbf{b}) - S(\mathbf{b})$ which is always positive, its quantum equivalent

$$S(\rho_A|\rho_B) = S(\rho_{AB}) - S(\rho_B)$$

is not. The state ρ_{AB} is **entangled** if $S(\rho_A|\rho_B) < 0$

Note that $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$ but also that $\rho_{AB} = \sum_i \rho_A^{(i)} \otimes \rho_B^{(i)}$ are also always separable. Thus if $S(\rho_A|\rho_B) = S(\rho_A)$ we have a separable state and no correlations, if $0 < S(\rho_A|\rho_B) < S(\rho_A)$ we have classical correlations between A and B and if $S(\rho_A|\rho_B) < 0$ we have quantum correlations.

2.4 Schmidt decomposition

Any pure two-qubit state $|\Psi_{AB}\rangle$ can be written in terms of orthogonal single qubit bases $\{\psi_A^0, \psi_A^1\}$ and $\{\psi_B^0, \psi_B^1\}$ such that:

$$\Psi_{AB} = a|\psi_A^0\rangle|\psi_B^0\rangle + b|\psi_A^1\rangle|\psi_B^0\rangle + c|\psi_A^0\rangle|\psi_B^1\rangle + d|\psi_A^1\rangle|\psi_B^1\rangle$$

The powerful **Schmidt decomposition** allows us to write *any* pure two-qubit state in terms of two new orthogonal bases:

$$\Psi_{AB} = a'|\psi'^0_A\rangle|\psi'^0_B\rangle + b'|\psi'^1_A\rangle|\psi'^1_B\rangle$$

where a', b' are the non-negative and *real* **Schmidt coefficients**, obeying $a'^2 + b'^2 = 1$.

2.5 Consequences of the Schmidt decomposition

As the Schmidt decomposition

$$\Psi_{AB} = a'|\psi'^0_A\rangle|\psi'^0_B\rangle + b'|\psi'^1_A\rangle|\psi'^1_B\rangle$$

holds for all pure states, it follows that

$$\rho_A = a'^2|\psi'^0_A\rangle\langle\psi'^0_A| + b'^2|\psi'^1_A\rangle\langle\psi'^1_A|$$

and

$$\rho_B = a'^2|\psi'^0_B\rangle\langle\psi'^0_B| + b'^2|\psi'^1_B\rangle\langle\psi'^1_B|$$

so that the eigenvalues of both ρ_A and ρ_B are equal, namely a'^2 and b'^2. Thus for any pure two-qubit state $S(\rho_A) = S(\rho_B)$.

2.6 Bell state measurements

Just as one can do measurements on a single qubit by projecting into some two-dimensional orthogonal basis, e.g. $\{|0\rangle, |1\rangle\}$, one can project two-qubit states into some four-dimensional orthogonal basis. This could be $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ which would amount to two separate single qubit measurements.

On could however also choose the **Bell state** basis $\{|\Psi^+\rangle, |\Psi^-\rangle, |\Phi^+\rangle, |\Phi^-\rangle, \}$ where

$$|\Psi^\pm\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \quad |\Phi^\pm\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle)$$

The Bell states are entangled, and thus measurements in this basis cannot be achieved by separate single-qubit measurements.

2.7 Superdense coding

Entanglement is a peculiar property of quantum systems. For instance we can send **two** classical bits using one qubit, if that qubit is part of an entangled pair, and the person we send it to has the other member of the pair. Note that by using local operations represented by the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ to manipulate *one qubit* we can change a Bell state into any other Bell state. Thus, if two parties (Alice and Bob) share a particular Bell state, e.g. $|\Phi^+\rangle$ to start with, then Alice can transfer two bits of classical information to Bob by first manipulating her qubit and producing whichever Bell state she wishes, and then sending her qubit to Bob. Bob, who has left his qubit untouched, performs a Bell state measurement and can find out which state Alice has sent him. Since Alice had the choice of four states, she was able to transmit two bits.

2.8 Teleportation

Another application of entanglement and Bell state measurements is **teleportation**. Again Alice and Bob share a Bell pair of qubits. If Alice wants to teleport a single-qubit quantum state $|\psi\rangle_C = \alpha|0\rangle + \beta|1\rangle$ to Bob, she performs a Bell state measurement on this qubit together with her part of the Bell state she shares with Bob. She then tells Bob which Bell state her two qubits collapsed to. Bob's Bell state qubit has also been projected by Alice's measurement, so that by using Alice's information to perform a local Pauli matrix manipulation he can recreate the qubit $|\psi_C\rangle$, although all that travelled from Alice to Bob were two bits of classical information telling Bob which of the four Bell states occurred! The key realization is that we can rewrite the combined state of the three qubits A, B, C:

$$|\psi\rangle_C|\Phi^+_{AB}\rangle = |\Phi^+_{CA}\rangle|\psi\rangle_B + |\Phi^-_{CA}\rangle\sigma_z|\psi\rangle_B + |\Psi^+_{CA}\rangle\sigma_x|\psi\rangle_B + |\Psi^-_{CA}\rangle(i\sigma_y)|\psi\rangle_B$$

2.9 Entanglement swapping

We can use a mechanism similar to teleportation in order to entangle two particles which have never interacted. This is termed **entanglement swapping** and requires two Bell states to start with. A Bell state measurement is then performed on *one photon from each Bell pair*. This projects the remaining two photons into a Bell state, even if they are light-years apart and have never interacted. We can summarize this procedure as:

$$|\Phi^+\rangle_{12}|\Phi^+\rangle_{34} = \frac{1}{2}(|\Psi^+\rangle_{13}|\Psi^+\rangle_{24} + |\Psi^-\rangle_{13}|\Psi^-\rangle_{24} + |\Phi^+\rangle_{13}|\Phi^+\rangle_{24} + |\Phi^-\rangle_{13}|\Phi^-\rangle_{24})$$

and the Bell state measurement projects out one of the terms on the RHS.

2.10 Tripartite entanglement

While for two particles any entangled state can be converted into any other state of the same entanglement through local operations, for three particles there are

two distinct classes of states. A state from one class cannot be converted into a state of the other class using only local operations. The two classes are defined by the **GHZ states** and **W states**:

$$\begin{aligned} |\Psi_{GHZ}\rangle &= \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \\ |\Psi_{W}\rangle &= \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle) \end{aligned}$$

Note that for the GHZ state, measuring one qubit gives an unentangled state of the remaining two, while for the W state there is a $\frac{2}{3}$ probability of obtaining a Bell state.

2.11 Bell's inequalities

Imagine Alice and Bob waiting to be interviewed for a job. They will be interviewed simultaneously but in different rooms. They know that one of two questions will come up: "Are you left-handed?" (L) or "Are you short-sighted?" (S) but not necessarily the same question for both of them. Let's say the interviewers write down a 1 if the answer is yes, and -1 if the answer is no, and let's call the answers L_1, S_1, L_2, S_2, bearing in mind that only one will be asked. It is easy to show that in any case:

$$B = L_1L_2 + L_1S_2 + S_1L_2 - S_1S_2 = L_1(L_2 + S_2) + S_1(L_2 - S_2) = \pm 2$$

Therefore, the average of this quantity over many interviewees obeys

$$\langle B\rangle = \langle L_1L_2\rangle + \langle L_1S_2\rangle + \langle S_1L_2\rangle - \langle S_1S_2\rangle \leq 2$$

If however Alice and Bob had secret microphones and earplugs *during* the interview, they could reach averages higher than 2 – why?

Because the answers can now depend on what the interviewer is asking the *other person*. Alice and Bob can agree that they should give *opposite* answers if *both* are asked the S question and otherwise they should give *the same* answer. For many Alices and Bobs this gives $< B >= 4 \geq 2$.

Alice and Bob can however also beat the inequality without microphones and earplugs. Instead they share a Bell pair beforehand and take their photon into the interview. They agree on four measurement bases, two for each of them. Relative to some reference frame Alice's two bases are rotated by 0° and 45° while Bob's are rotated by 67.5° and 22.5°. If the interviewer asks them the S question, they are to use the first of their bases, if they are asked the L question they should use the second basis. If the photon is projected into the 0°/22.5°/45°/67.5° state, they are to answer 'yes' (1), if it is projected into the perpendicular state they should say 'no' (-1).

The expectation values of their answers are now:

$$\langle L_1L_2\rangle = \cos^2 22.5^\circ - \sin^2 22.5^\circ = \cos 45^\circ = \frac{1}{\sqrt{2}}$$

$$\langle S_1 L_2 \rangle = \cos^2 22.5^\circ - \sin^2 22.5^\circ = \cos 45^\circ = \frac{1}{\sqrt{2}}$$
$$\langle L_1 S_2 \rangle = \cos^2 22.5^\circ - \sin^2 22.5^\circ = \cos 45^\circ = \frac{1}{\sqrt{2}}$$
$$\langle S_1 S_2 \rangle = \cos^2 67.5^\circ - \sin^2 67.5^\circ = \cos 135^\circ = -\frac{1}{\sqrt{2}}$$

Thus $\langle L_1 L_2 \rangle + \langle L_1 S_2 \rangle + \langle S_1 L_2 \rangle - \langle S_1 S_2 \rangle = 2\sqrt{2} > 2$ and we have beaten the inequality!

3 Codes and Compression

3.1 Classical Codes

A **code** C_N is a subset of the Hamming space $\{0,1\}^N$. It has the following properties:

length N (dimension of Hamming space)

size $r = \#C_N$ (number of codewords)

minimum distance δ (min. Hamming distance between codewords)

This information is often summarized in the notation $[N, r, \delta]$. From these properties a further one can be derived, namely the **information rate** ρ:

$$\rho = \frac{\log_2 r}{N}$$

3.2 Examples of classical codes

Two simple examples of codes:

repetition code $[N, 2, N]$ which consists of the words $000\ldots0$ and $111\ldots1$.

parity code $[N, 2^{N-1}, 2]$ which consists of all words with an even number of ones ($0 = \sum_i x_i \bmod 2$)

3.3 Error detection and correction

An $[N, r, \delta]$ code C_N can **detect** $D \leq \delta - 1$ errors, and can **correct** $E \leq \frac{\delta-1}{2}$ (rounded down for even δ) errors.

To see why, consider "balls" of radius R in Hamming space around the codewords, which includes all bitstrings of distance $d \leq R$ around the word at the origin of the ball.

If a codeword is affected by e errors, then if $e < D$ we cannot be at another codeword yet, and thus are in the space between codewords which signifies an error. If $e < E$ we are still within a *disjoint* ball of radius $R = E$ around the original codeword which is uniquely identifiable and thus the error can be corrected.

3.4 The Hamming Bound and Perfect Codes

The volume of a ball of radius r in N-dimensional Hamming space is given by:

$$v_N(r) = \sum_{i=0}^{r} \left(\begin{array}{c} N \\ i \end{array} \right)$$

which is just the number of bit strings with r or less ones. Thus, if a code corrects E errors, its size r is limited by:

$$r \leq \frac{2^N}{v_N(E)}$$

This is the **Hamming bound**, and a code which achieves the equality is termed a **perfect code**.

3.5 Linear codes

A code is **linear** if for each pair of codewords $\mathbf{x}, \mathbf{y} \in C_N$, the string $z_i = x_i + y_i \mod 2$ is also a codeword.

Linear codes are a k-dimensional subspace of Hamming space which can be spanned by a set of k linearly independent basis words. Thus the size of a linear code is $r = 2^k$

Linear codes are denoted using *round* brackets, and by mentioning the *rank* k instead of the size r. Furthermore the distance in linear codes is denoted as d so that we talk of (N, k) and (N, k, d) codes.

3.6 Generator and Parity Check Matrix

The **generator G** of a linear code is simply a list of all basis words in form of a $N \times k$ matrix.

The **parity check matrix H** of a linear code is a list of linearly independent binary vectors which are *orthogonal to any basis words.*

Note that for two binary vectors $\mathbf{x}$ and $\mathbf{y}$ to be orthogonal in Hamming space means that $0 = \mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i \mod 2$. This means that some non-zero vectors are orthogonal to themselves.

3.7 The Hamming (7,4) code

As an example consider the **Hamming (7,4) code** which is determined by its parity check matrix **H**, which is taken to be the lexicographical ordering of all non-zero bit strings of length 3. Its generator **G** is:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The **syndrome** $\mathbf{s}(\mathbf{x})$ of a given bit string $\mathbf{x}$ is $\mathbf{s}(\mathbf{x}) = \mathbf{xH}$, which is the zero vector for a codeword and one of the seven rows of $\mathbf{H}$ for all seven error strings of weight one, which specifies the position of the error exactly.

3.8 Quantum errors

While a classical bit can only suffer from a *bit flip* error ($0 \leftrightarrow 1$), a qubit can be subjected to three different errors which are equivalent to operations of the Pauli matrices on the qubit:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

These are **bit flip** (σ_x), **combined bit/phase flip** (σ_y) and **phase flip** (σ_z).

3.9 Quantum codes

Consider a general error operator $\mathbf{E}_{\boldsymbol{\alpha}}$ where $\boldsymbol{\alpha}$ is a vector with entries $\in \{I, X, Z\}$ standing for the unit matrix and the σ_x and σ_z matrices (we can form σ_y from the other two). The element α_j signifies the error on the jth qubit, or in the case of I, that there is no error.

3.10 Quantum error detection and correction

A quantum code $\mathcal{X}$ is D**-error detecting** if $\forall \boldsymbol{\alpha}$ such that $w(\boldsymbol{\alpha}) \leq D$ and $\forall \psi, \psi' \in \mathcal{X}$:

$$\langle \psi' | \mathbf{E}_{\boldsymbol{\alpha}} | \psi \rangle = c_{\boldsymbol{\alpha}} \langle \psi' | \psi \rangle$$

where $c_{\boldsymbol{\alpha}} \in \mathbb{C}$.

It is E**-error correcting** if $\forall \boldsymbol{\alpha}, \boldsymbol{\alpha}'$ such that $w(\boldsymbol{\alpha}), w(\boldsymbol{\alpha}') \leq E$ and $\forall \psi, \psi' \in \mathcal{X}$:

$$\langle \psi' | \mathbf{E}_{\boldsymbol{\alpha}'} \mathbf{E}_{\boldsymbol{\alpha}} | \psi \rangle = b_{\boldsymbol{\alpha}', \boldsymbol{\alpha}} \langle \psi' | \psi \rangle$$

where $b_{\boldsymbol{\alpha}', \boldsymbol{\alpha}} \in \mathbb{C}$.

3.11 Non-degenerate detection and correction

Considering the formulae again:

$$\langle \psi' | \mathbf{E}_{\boldsymbol{\alpha}} | \psi \rangle = c_{\boldsymbol{\alpha}} \langle \psi' | \psi \rangle \quad \langle \psi' | \mathbf{E}_{\boldsymbol{\alpha}'} \mathbf{E}_{\boldsymbol{\alpha}} | \psi \rangle = b_{\boldsymbol{\alpha}', \boldsymbol{\alpha}} \langle \psi' | \psi \rangle$$

If $b_{\boldsymbol{\alpha}', \boldsymbol{\alpha}} = 0$ unless $\boldsymbol{\alpha}' = \boldsymbol{\alpha}$ and/or $c_{\boldsymbol{\alpha}} = 0$ unless $\mathbf{E}_{\boldsymbol{\alpha}} = \mathbf{I}$ then the code is said to be **non-degenerate** D-detecting and/or E-correcting respectively.

For detection this means that all error operators with $w(\boldsymbol{\alpha}) \leq D$ project $|\psi\rangle$ into an orthogonal subspace, whereas for correction all pairs of error operators with $w(\boldsymbol{\alpha}), w(\boldsymbol{\alpha}') \leq E$ and $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}'$ project $|\psi\rangle$ into mutually orthogonal subspaces.

3.12 Properties of quantum codes

As in classical codes, we can define a code as (N, k) or (N, k, d) code, where N is its length in qubits, k is the dimension of the code subspace of the Hilbert space and therefore $\log_2 r$, and d is the distance, defined as the minimum $w(\boldsymbol{\alpha})$ at which, for some ψ, ψ':

$$\langle \psi' | \mathbf{E}_{\boldsymbol{\alpha}} | \psi \rangle \neq 0$$

As with classical codes, $D = d - 1$. Also, a code which corrects E errors, detects $2E$ errors, and a code which detects D errors, corrects $D/2$. Furthermore, if the location of the errors is known (but not the type), a code which detect D errors, corrects D errors.

3.13 The Shor code

The **Shor code** is a (9,1,3) code consisting of the two 9-qubit states:

$$\begin{aligned} |\Psi_0\rangle &= \frac{1}{2\sqrt{2}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle) \\ |\Psi_1\rangle &= \frac{1}{2\sqrt{2}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle) \end{aligned} \tag{3}$$

A single bit flip is corrected because there are always three qubits together (000, 111). A single phase flip is corrected because there are three groups.

3.14 Dual classical codes

A classical linear code $\mathcal{C}$ has a **dual code** $\mathcal{C}^\perp$ which has the transposed parity check matrix H^T of code $\mathcal{C}$ as its generator and the transposed generator G^T of code $\mathcal{C}$ as its check matrix. If $\mathcal{C}$ is a (N, k) code, $\mathcal{C}^\perp$ will be a $(N, N - k)$ code.

By construction, all code words of $\mathcal{C}$ are orthogonal to those in $\mathcal{C}^\perp$ but due to the mod 2 scalar product, we can have $\mathcal{C} \subseteq \mathcal{C}^\perp$ (weakly self dual) and $\mathcal{C} = \mathcal{C}^\perp$ (strictly self-dual) codes. A simple example is the repetition code of length 4, which has $G = (1111)$ and

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

The parity code is the dual code of the repetition code.

3.15 Cosets

Consider a linear code $\mathcal{C}$ with k generators, and a subset of k' of those generators, which give rise to a second code $\mathcal{C}' \subset \mathcal{C}$.

All possible $2^{k-k'}$ words generated by the $k - k'$ generators *outside* the code $\mathcal{C}'$ give rise to the $2^{k-k'}$ **cosets** of the code $\mathcal{C}'$ when combined with the elements of $\mathcal{C}'$.

3.16 CSS codes

The *Calderbank-Shor-Steane (CSS) codes* are an important class of quantum codes constructed using linear classical codes. Taking two linear classical codes $\mathcal{C}, \mathcal{C}'$ of ranks k and k', such that $\mathcal{C}'$ is a linear subspace of $\mathcal{C}$, one can construct $2^{k-k'}$ quantum code words using the cosets by writing, for any $\mathbf{x} \in \mathcal{C}$:

$$\Psi_{\mathbf{x}} = \frac{1}{2^{k'/2}} \sum_{\mathbf{y} \in \mathcal{C}'} |\mathbf{x} + \mathbf{y}\rangle$$

It can correct $(d-1)/2$ bit flip errors and $(d^{\perp}-1)/2$ phase flip errors, where d is the distance of $\mathcal{C}$ and $d^{\perp}$ is the distance of $\mathcal{C}^{\perp}$, the dual code of $\mathcal{C}$.

Note that Hamming codes are very well suited for constructing CSS codes.

3.17 Classical data compression

Recall that if m types of symbol are distributed in a text with a probability distribution $\{p_i\}$ then the average amount of information (in bits) per character is the Shannon entropy

$$S = -\sum_{i}^{m} p_i \log_2 p_i$$

From this follows **Shannon's noiseless channel coding theorem** which states that we can compress a string of N characters at most down to $N \times S$ bits.

3.18 Quantum messages

We already know that we can use two orthogonal quantum states of a qubit $|\Psi_0\rangle$ and $|\Psi_1\rangle$ as our '0' and '1' for encoding classical bits in a quantum system. Thus we can create an n-qubit state carrying an n-bit message:

$$|\Psi^n\rangle = |\Psi_0\rangle \otimes |\Psi_1\rangle \otimes |\Psi_0\rangle \otimes |\Psi_0\rangle \otimes |\Psi_1\rangle \otimes |\Psi_1\rangle \ldots$$

Similar to a probability distribution of symbols in a text, we now have a probability distribution of two density matrices, $\rho_0 = |\Psi_0\rangle\langle\Psi_0|$ and $\rho_1 = |\Psi_1\rangle\langle\Psi_1|$ which gives us a mixed density matrix $\rho = p_0\rho_0 + p_1\rho_1$ where p_0 and p_1 are the probabilities of the two density matrices occurring.

3.19 Schumacher's coding theorem

Recall that the probabilities p_0 and p_1 for the orthogonal decomposition are the eigenvalues of ρ. If you then also recall that the von Neumann entropy $S(\rho)$ is the Shannon entropy of the eigenvalues of ρ, then it should come as no surprise that $S(\rho)$ is the average amount of classical information per qubit which can be transmitted.

This in turn gives us Schumacher's coding theorem which states that in the n-qubit Hilbert space $\mathcal{H}^{\otimes n}$ of the message there exists a subspace $\mathcal{X}_n$ of

dimension $nS(\rho)$, such that the n-qubit message (with symbols occurring at frequencies p_0 and p_1) can be projected onto $\mathcal{X}_n$ with unit probability.

In other words, only $nS(\rho)$ qubits are required to transmit an n-qubit message.

3.20 Compression and decompression

Consider again a pure message state of n qubits $|\Psi^n\rangle = |\Psi_0\rangle \otimes |\Psi_1\rangle \otimes |\Psi_0\rangle \otimes \ldots$ in which $|\Psi_0\rangle$ and $|\Psi_1\rangle$ appear with frequencies p_0 and p_1. To compress and decompress the following three steps are necessary:

1. The operator $\mathbf{P}_{\mathcal{X}_n}$ projects an n-qubit message state onto the subspace $\mathcal{X}_n$ of dimension $2^{nS(\rho)}$.

2. Using a unitary operator $\mathbf{U}_{\mathcal{X}_n}$ on $\mathcal{H}^{\otimes n}$ the message is brought into the form $|\Psi_{\text{comp}}\rangle \otimes |\mathbf{0}\rangle$ where $|\Psi_{\text{comp}}\rangle \in \mathcal{H}^{\otimes nS(\rho)}$ and $|\mathbf{0}\rangle \in \mathcal{H}^{\otimes[n-nS(\rho)]}$. The message is now compressed and can be sent using the $nS(\rho)$ qubits of the state $|\Psi_{\text{comp}}\rangle$

3. Upon receipt of the compressed message, one re-appends the string $|\mathbf{0}\rangle$ and applies the unitary operator $\mathbf{U}_{\mathcal{X}_n}^{-1}$ to arrive at the state $\mathbf{P}_{\mathcal{X}_n}|\Psi^n\rangle$.

4 Measurements and Operations

4.1 Orthogonal Projective Quantum Measurements

The quantum measurements one learns about first are **orthogonal projective** quantum measurements. Usually an *observable* A is defined by a Hermitian operator $\hat{A}$. A given state $|\psi\rangle$ collapses into one of the eigenstates $|\Psi_i\rangle$ of $\hat{A}$ with probability $p_i = |\langle\psi|\Psi_i\rangle|^2$. The corresponding eigenvalue a_i of $\hat{A}$ is the value of A in this measurement.

The choice of A corresponds thus to a choice of $\{a_i\}$ and of $\{|\Psi_i\rangle\}$. As far as the actual measurement of $|\psi\rangle$ is concerned, only the eigenstates $\{|\Psi_i\rangle\}$ are of interest. The eigenvalues of $\hat{A}$ are of physical importance, but do not influence what state $|\psi\rangle$ collapses to. Thus quantum measurements in general are not defined by observables but only by a set of measurement operators, which in the simplest case are projectors onto the set of orthogonal eigenstates $\{|\Psi_i\rangle\}$.

4.2 Positive operator value measures

More generally one can define a set of positive-definite, hermitian measurement operators $\{F_i\}$ which obey $\sum_i F_i = I$. In other words the $\{F_i\}$ form a **positive definite partition of unity**. In the orthogonal measurement case we have $F_i = |\Psi_i\rangle\langle\Psi_i|$ for a set of orthogonal states $\{|\Psi_i\rangle\}$.

In general we can write $F_i = 2\lambda_i\rho_i$, or in terms of a Bloch vector $\mathbf{n}$:

$$F_i = \lambda_i(\mathbf{I} + \mathbf{n}^{(i)} \cdot \boldsymbol{\sigma})$$

where $\sum_i \lambda_i = 1$ and $\sum_i \lambda_i \mathbf{n}^{(i)} = 0$.

4.3 Kraus operators

We can rewrite POVM operators $\{F_i\}$ as **Kraus operators** $\{M_i\}$ which are a (non-unique) decomposition of the POVM operators:

$$F_i = M_i^\dagger M_i$$

In a POVM measurement a density matrix ρ collapses to ρ_i with probability p_i, where:

$$\rho_i = \frac{M_i \rho M_i^\dagger}{\mathrm{tr}(M_i \rho M_i^\dagger)}$$

and $p_i = \mathrm{tr}(M_i \rho M_i^\dagger) = \mathrm{tr}(F_i \rho)$. This is the most general form of quantum measurement.

4.4 Superoperators

We have established that in the most general quantum measurement, outcome i with state $\rho_i = \frac{M_i \rho M_i^\dagger}{\mathrm{tr}(M_i \rho M_i^\dagger)}$ occurs with probability $p_i = \mathrm{tr}(M_i \rho M_i^\dagger)$. Therefore the complete ensemble of outcomes can be written as a mixed density matrix.

$$\rho' = \mathfrak{F}(\rho) = \sum_i p_i \rho_i = \sum_i M_i^\dagger \rho M_i$$

which we can write as the result of the action of the **superoperator** $\mathfrak{F}$ on ρ.

Formally a superoperator is defined as the action of $\mathfrak{F} = \{M_i\}$ in a Hilbert space $\mathcal{H}$ of dimension m on a $m \times m$ complex matrix $\mathbf{A}$, so that the map $\mathbf{A} \to \mathfrak{F}(\mathbf{A}) = \sum_i M_i^\dagger \mathbf{A} M_i$ is:

1. **linear:** $\mathfrak{F}(\mathbf{A_1} + \mathbf{A_2}) = \mathfrak{F}(\mathbf{A_1}) + \mathfrak{F}(\mathbf{A_2})$
2. **completely positive:** $\sum_i (M_i^\dagger \otimes \mathbf{I}_{\mathcal{H}'}) \tilde{\mathbf{A}} (M_i \otimes \mathbf{I}_{\mathcal{H}'})$ is positive-definite $\forall\, \tilde{\mathbf{A}}$ acting on $\tilde{\mathcal{H}} = \mathcal{H} \otimes \mathcal{H}'$
3. **trace- and unity preserving:** (unital superoperators only) $\mathrm{tr}\mathfrak{F}(\mathbf{A}) = \mathrm{tr}\mathbf{A}$ and $\mathfrak{F}(\mathbf{I}) = \mathbf{I}$

While so-called **unital superoperators** have to be trace- and unity preserving (property 3 above), general superoperators only have to obey $\mathrm{tr}\mathfrak{F}(\mathbf{A}) \leq \mathrm{tr}\mathbf{A}$, and $\mathbf{I} - \mathfrak{F}(\mathbf{I})$ has to be positive definite.

For the general superoperators $\sum_i M_i M_i^\dagger \leq \mathbf{I}$, while the kraus operators of unital superoperators have to obey the usual POVM constraint of $\sum_i = M_i M_i^\dagger = \mathbf{I}$.

4.5 Depolarizing Channel

We can use the superoperator formalism to describe the action of noisy quantum channels on a density matrix that is being transmitted. An example is the **depolarizing channel** which, with probability $p/3$ performs a bit flip (σ_x), a phase flip (σ_z) or a combined flip (σ_y), and with probability $1-p$ leaves the density matrix unchanged.

The Kraus operators of this channel are:

$$M_I = \sqrt{1-p}\,\mathbf{I} \quad M_X = \sqrt{\tfrac{p}{3}}\sigma_x$$
$$M_Y = \sqrt{\tfrac{p}{3}}\sigma_y \quad M_Z = \sqrt{\tfrac{p}{3}}\sigma_z$$

4.6 Erasure Channel

Another quantum channel is the **erasure channel** which relies on an extension of Hilbert space by an additional dimension, which corresponds to the "erased" state. Its operators are:

$$M_0 = \begin{pmatrix} \sqrt{1-p} & 0 & 0 \\ 0 & \sqrt{1-p} & 0 \\ 0 & 0 & 0 \end{pmatrix} M_1 = \begin{pmatrix} 0 & 0 & \sqrt{p} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{p} \\ 0 & 0 & 0 \end{pmatrix}$$

And the action of the channel can be summarized as:

$$\rho \to (1-p)\rho + p|e\rangle\langle e|$$

where $|e\rangle$ is the erased state.

4.7 Phase-damping channel

This channel is characterized by the following Kraus operators:

$$M_0 = \begin{pmatrix} (1-p) & 0 \\ 0 & (1-p) \end{pmatrix} M_1 = \begin{pmatrix} \sqrt{p} & 0 \\ 0 & 0 \end{pmatrix} M_2 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix}$$

which means that the matrix is made "more diagonal" and therefore more mixed:

$$\rho \to \begin{pmatrix} \rho_{00} & (1-p)\rho_{01} \\ (1-p)\rho_{10} & \rho_{11} \end{pmatrix}$$

4.8 Amplitude-damping Channel

Yet another channel:

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix} M_1 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$$

which means that the matrix tends towards $\rho_{00} = 1$ and $\rho_{ij} = 0$ otherwise:

$$\rho \to \begin{pmatrix} \rho_{00} + p\rho_{11} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{pmatrix}$$

4.9 Quantum Zeno Effect

The **Quantum Zeno Effect (QZE)** is a measurement effect which, in essence says that *a continously observed quantum system will not evolve in time.*

Imagine a time evolution which rotates a state $|0\rangle$ to a state $\cos\omega t|0\rangle + \sin\omega t|1\rangle$. If we measure the state at time intervals of $\Delta t = \frac{1}{N}$, the state is going to be projected into $|0\rangle$ with probability $\cos^2\frac{\omega}{N}$, and the probability of the state being $|0\rangle$ after N measurements is $\cos^{2N}\frac{\omega}{N}$ which for $N \to \infty$ goes to one.

4.10 Quantum Anti-Zeno Effect

Perhaps even more fascinating is the so-called **Quantum Anti-Zeno** or **Inverse Zeno Effect**. Here we do a succession of measurements in which the measurement basis is slightly rotated each time. This way we can "drag" the state from $|0\rangle$ to $|1\rangle$ with a probability approaching one, as the number of steps goes to infinity.

4.11 Interaction-Free Measurement

One of the most striking examples of the way in which quantum mechanics violates everyday notions of locality and realism is **interaction-free measurement (IFM)**. Consider a Mach-Zehnder interferometer with a dark and a bright output. Now one arm is blocked by an object. The idea of the IFM is that a single photon entering the Mach-Zehnder apparatus has a probability of $\frac{1}{4}$ of exiting in the dark output, because the obstruction in one of the arms destroys its ability to *interfere with itself* and cause the destructive and constructive interference which results in the "dark" and "bright" exits of the unobstructed interferometer. So we can detect the presence of an object without a photon "interacting" with it.

4.12 Perfect IFM

We can do better than just detecting an object 25% of the time without interaction. In fact we can detect it with unit probability, by employing the Quantum Zeno Effect! The idea is to start with a horizontally polarized photon, rotate its polarization slightly and let it traverse a polarizing beamsplitter which reflects vertical polarization and transmits horizontal polarizetion. These two paths are reunited in another polarizing beamsplitter. If the path with vertical polarization is obstructed, then the photon will still exit the second beamsplitter with a high probability, but in the horizontal state. If the vertical path is not obstructed, the photon polarization will remain rotated. The photon is then recycled through the apparatus until its polarization would be rotated by 90° if the arm is not obstructed. At this point a simple measurement of polarization will establish whether the arm is obstructed or not.

4.13 Hardy's paradox

An elegant non-locality proof closely related to interaction-free measurement is **Hardy's paradox**. In this gedankenexperiment an electron and a positron each traverse a Mach-Zehnder interferometer. The interferometers overlap in one arm. If no annihilation is observed, the state of the two particles is projected into the entangled state:

$$|\psi\rangle = \frac{1}{\sqrt{3}}(|NO\rangle|O\rangle + |O\rangle|NO\rangle + |NO\rangle|NO\rangle)$$

We then look at four scenarios...

These are:

1. The second (reunifying) beamsplitters are present in both interferometers, i.e. both interferometers are complete.
2. The positron's second beamsplitter is absent, the electron's is present.
3. The electron's second beamsplitter is absent, the positron's is present.
4. Both second beamsplitters are absent.

The four quantum states are:

$$\begin{aligned}
|\Psi_1\rangle &= \frac{1}{4}[-2|\gamma\rangle - 3|O^+\rangle|O^-\rangle + i|O^+\rangle|NO^-\rangle + i|NO^+\rangle|O^-\rangle - |NO^+\rangle|NO^-\rangle] \\
|\Psi_2\rangle &= \frac{1}{2\sqrt{2}}[-\sqrt{2}|\gamma\rangle - |O^+\rangle|O^-\rangle + i|O^+\rangle|NO^-\rangle + 2i|NO^+\rangle|O^-\rangle] \\
|\Psi_3\rangle &= \frac{1}{2\sqrt{2}}[-\sqrt{2}|\gamma\rangle - |O^+\rangle|O^-\rangle + 2i|O^+\rangle|NO^-\rangle + i|NO^+\rangle|O^-\rangle] \\
|\Psi_4\rangle &= \frac{1}{2}[-|\gamma\rangle + i|O^+\rangle|NO^-\rangle + i|NO^+\rangle|O^-\rangle + |NO^+\rangle|NO^-\rangle] \qquad (4)
\end{aligned}$$

Then one constructs local quantities which tell whether (1) or not (0) a particle is in the overlapping (ov) or nonoverlapping (nov) arm, and whether the second beamsplitter for that particle is present (0) or absent (∞). From the four states we can deduce:

$$\begin{aligned}
e^{\infty}_{ov}p^{\infty}_{ov} = 0 \qquad & e^{0}_{nov} = 1 \rightarrow p^{\infty}_{ov} = 1 \\
e^{0}_{nov}p^{0}_{nov} = 1 \qquad & e^{\infty}_{ov} = 1 \rightarrow p^{0}_{nov} = 1 \qquad (5)
\end{aligned}$$

Which is a contradiction, meaning that such local quantities cannot exist.