FAITHFUL TRANSPORT OF

QUANTUM INFORMATION

T. C. Dorlas (D.I.A.S.)

Maynooth, 10 February 2006.

Classical Information Theory

Given a set of random symbols x from an alphabet A occurring with probabilities p_x, Shannon introduced the concept of *average information gain*:

$$H\left(\{p_x\}_{x\in A}\right) = -\sum_{x\in A} p_x \, \log_2 p_x .$$

Note that $-\log\, p_x$ is the number of bits of information gained if one finds that x occurs. This is now called the **Shannon entropy**.

If a source of information emits a stream of symbols from the alphabet A independently with probabilities p_x, the average information per symbol is given by H.

The Shannon entropy has an operational meaning for **data compression**. Data compression is a map $\mathcal{C}^n : A^n \to \{0,1\}^{nR}$, where $R < 1$ is called the rate of compression. To decompress, one needs a decompression map $\mathcal{D}^n : \{0,1\}^{nR} \to A^n$.

Theorem. **(Shannon)** Consider a memoryless source of information with probability distribution μ on an alphabet A. Let $\epsilon > 0$ and suppose $R > H(\mu)$. Then there exist compression schemes $\mathcal{C}^n : A^n \to \{0,1\}^{nR}$ and corresponding decompression schemes $\mathcal{D}^n : \{0,1\}^{nR} \to A^n$ such that for large enough n,

$$\mu_n \left[\{ \underline{x} \in A^n \,|\, \mathcal{D}^n \left(\mathcal{C}^n (\underline{x}) \right) \neq \underline{x} \} \right] < \epsilon.$$

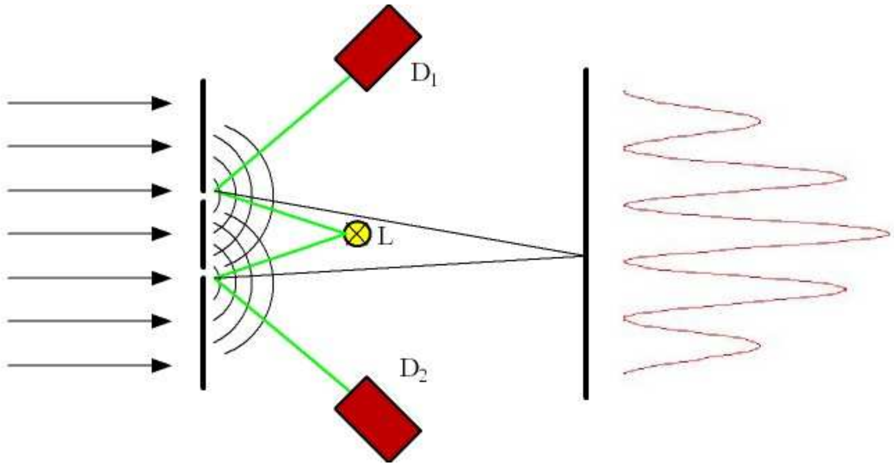
Conversely, if $R < H(\mu)$ then such schemes do not exist for any $\epsilon < 1$.

QUANTUM INFORMATION

The analogue of **bits** of information is **qubits** of quantum information. Remember that the state of a quantum system is given by a vector in a Hilbert space $\mathcal{H}$. A *qubit* is a state vector in a 2-dimensional Hilbert space, i.e. a complex 2-component vector.

Remember also that a measurement in quantum mechanics corresponds to a hermitian operator A on $\mathcal{H}$. The eigenvalues λ_i ($i = 1, \ldots, d$ if $d = \dim \mathcal{H}$) of A are the possible numerical values a measurement of A can take, and if the system is in a state $\psi \in \mathcal{H}$ then the probability that the result is λ_i is given by $\dfrac{||P_i\psi||^2}{||\psi||^2}$, where P_i is the projection onto the eigenspace for λ_i.

I'll now argue that we need to extend the concept of *state*. Consider the following 2-slit experiment:



One can show that the probability of detecting a particle at position x on the screen is given by

$$P(x) = |a\phi_1(x) + b\phi_2(x)|^2 + |a\phi_2(x) + b\phi_1(x)|^2,$$

where $\phi_i = P_i\phi$ is the projection of the wave function corresponding to hole i, and a resp. b is the probability amplitude of detecting a photon at detector 1, given that the particle went through hole 1 resp. 2, or a photon at detector 2 given that the particle went through hole 2 resp. 1. This is not of the form above.

The act of measurement changes the outcome of the experiment *even if it is not recorded*! To describe this situation, we need a concept of probabilistic mixture of vector states. This given by a **density matrix**, i.e. a positive-definite matrix with trace 1. If P_i is the projection onto the eigenspace of A with eigenvalue λ_i then the probability is given by

$$\mathrm{Trace}(\rho\, P_i).$$

Now suppose a quantum information source outputs states $\psi_k \in \mathcal{H}$ sequentially and independently with probabilities p_k $(k = 1, \ldots, m)$. This can be described by a density matrix

$$\rho = \sum_{k=1}^{m} p_k \, |\psi_k\rangle\langle\psi_k|.$$

Note that the states ψ_k are not necessarily orthogonal.

The analogue of the Shannon entropy is then the **Von Neumann entropy** given by

$$S(\rho) = -\mathrm{Trace}\,(\rho\,\log\rho).$$

Notice that if we diagonalise ρ:

$$\rho = \sum_{i=1}^{d} \lambda_i\,|\phi_i\rangle\langle\phi_i| \quad \text{ then } \quad S(\rho) = -\sum_{i=1}^{d} \lambda_i\,\log\lambda_i.$$

QUANTUM DATA COMPRESSION

Quantum data compression is given in general by a map $\mathcal{C}^n$ mapping the states $\psi_{\underline{k}}^{(n)}$ which are tensor products of states ψ_k, onto density matrices $\tilde{\rho}_{\underline{k}}^{(n)}$ of some new Hilbert space $\tilde{\mathcal{H}}_n$ of much lower dimension $d_c(n)$ than that of $\mathcal{H}^{\otimes n}$, that is d^n. In fact, it is more convenient to consider $\mathcal{C}^n$ as a map $\mathcal{C}^n : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\tilde{\mathcal{H}}_n)$ such that $\mathcal{C}^n \left(|\psi_{\underline{k}}^{(n)}\rangle\langle\psi_{\underline{k}}^{(n)}| \right) = \tilde{\rho}_{\underline{k}}^{(n)}$.

This map must be **completely positive**. This means that it must map positive definite matrices into positive definite matrices, even if the system is a subsystem of a larger system. More precisely, a map $\mathcal{E} : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ is completely positive if for any Hilbert space $\mathcal{K}$ and any positive definite $A \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{K})$, $(\mathcal{E} \otimes \mathrm{id})(A)$ is also positive definite.

The **asymptotic rate of compression** is given by:

$$R = \lim_{n\to\infty} \frac{\log d_c(n)}{n}.$$

The **fidelity** of the compression is defined by

$$F_n = \sum_{\underline{k}} \lambda_{\underline{k}} \langle \psi_{\underline{k}}^{(n)} \mid \tilde{\rho}_{\underline{k}}^{(n)} \psi_{\underline{k}}^{(n)} \rangle.$$

It satisfies $0 \leq F_n \leq 1$ and
$F_n = 1$ if and only if $\tilde{\rho}_{\underline{k}}^{(n)} = |\psi_{\underline{k}}^{(n)}\rangle\langle\psi_{\underline{k}}^{(n)}|$ for all $\underline{k}$. The

compression scheme is said to be **reliable** if $F_n \to 1$ as $n \to \infty$.

The analogue of Shannon's theorem is:

Theorem (Schumacher) *If $R > S(\rho)$ then there exists a reliable compression scheme of rate R; if $R < S(\rho)$ then any compression scheme of rate R is unreliable.*

ERROR-CORRECTING CODES

The simplest classical error-correcting code is the repetition code: simply repeat each bit 3 times and decide by majority rule. There is no direct quantum analogue because of the **No-Cloning Theorem**:

Theorem. There does not exist a unitary operator $U : \mathbb{C}^4 \to \mathbb{C}^4$ such that for a fixed $\psi_0 \in \mathbb{C}^2$ and all $\psi \in \mathbb{C}^2$, $U(\psi \otimes \psi_0) = \psi \otimes \psi$.

There are other difficulties as well:

- *Errors in qubits are continuous* and
- *Measurement destroys quantum information.*

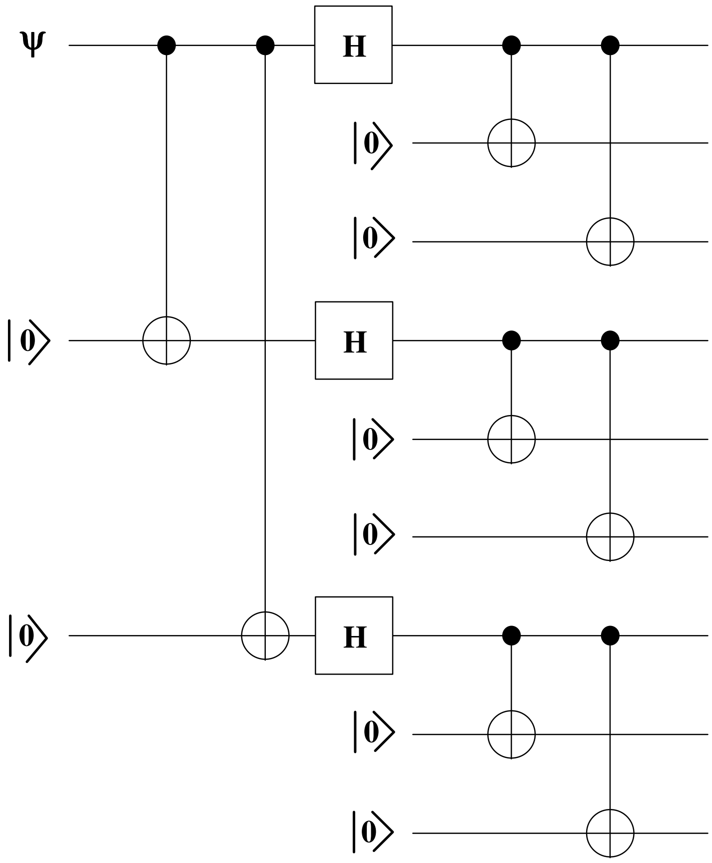
Peter Shor found a way out of these problems. His 9-qubit code is as follows:

$$|0\rangle \mapsto |0_L\rangle = \frac{1}{2\sqrt{2}}(|000\rangle + |111\rangle) \otimes (|000\rangle + |111\rangle)$$

$$\otimes (|000\rangle + |111\rangle) \text{ and}$$

$$|1\rangle \mapsto |1_L\rangle = \frac{1}{2\sqrt{2}}(|000\rangle - |111\rangle) \otimes (|000\rangle - |111\rangle)$$

$$\otimes (|000\rangle - |111\rangle).$$



Circuit for Shor's error-correcting code

To understand this why this code corrects arbitrary errors in a single qubit, consider first the simpler code

$$|0\rangle \mapsto |000\rangle \text{ and } |1\rangle \mapsto |111\rangle.$$

This code can correct single-qubit errors of the bit-flip kind, i.e. $|0\rangle \leftrightarrow |1\rangle$. Error-detection is done by a measurement with respect to the **error syndrome** given by the 4 projection operators P_0, P_1, P_2 and P_3

given by

$$P_0 = |000\rangle\langle 000| + |111\rangle\langle 111|$$

$$P_1 = |100\rangle\langle 100| + |011\rangle\langle 011|$$

$$P_2 = |010\rangle\langle 010| + |101\rangle\langle 101|$$

$$P_3 = |001\rangle\langle 001| + |110\rangle\langle 110|$$

If a bit-flip error occurs in the first qubit, a measurement of P_1 will result in the value 1, and we can recover the original state by applying $\sigma_x^{(1)}$.

A more complete description of a bit-flip error-prone channel is given by a map $\mathcal{E} \otimes \mathcal{E} \otimes \mathcal{E}$ on the $2^3 \times 2^3$ matrices where

$$\mathcal{E}(\rho) = (1-p)\rho + p\sigma_x \rho \sigma_x$$

introduces a bit-flip error with probability $p \ll 1$. To correct for such errors one then operates on the resulting state with the map $\mathcal{R}$ given by

$$\mathcal{R}(\rho) = \sum_{i=0}^{3} \sigma_x^{(i)} P_i \rho P_i \sigma_x^{(i)}.$$

(Here $\sigma_x^{(0)} = \mathbf{1}$.) Starting with a pure state $\rho_0 = |000\rangle\langle 000|$ for example, one obtains

$$\begin{aligned}\mathcal{R}(\mathcal{E}(\rho_0)) = {} & [(1-p)^3 + 3p(1-p)^2]\rho_0 \\ & + [3p^2(1-p) + p^3]\rho_1.\end{aligned}$$

We can compute the fidelity:

$$F = \langle 000|\, \mathcal{R}(\mathcal{E}(\rho_0))\, |000\rangle = 1 - 3p^2 + 2p^3.$$

The Shor code corrects *arbitrary* independent errors in the 9 individual qubits. Such errors can be described by a map $\mathcal{E} \otimes \ldots \otimes \mathcal{E}$ where $\mathcal{E}$ is a completely positive trace-preserving map $\mathcal{E} : \mathcal{M}_2 \to \mathcal{M}_2$ on the 2×2-matrices. Such maps can be written in the Kraus form

$$\mathcal{E}(\rho) = \sum_k E_k \, \rho E_k^*,$$

where E_k are matrices such that $\sum_k E_k^* \, E_k = \mathbf{1}$.

Each E_k can be written as $E_k = a_{k,0}\mathbf{1} + a_{k,1}\sigma_x + a_{k,2}\sigma_y + a_{k,3}\sigma_z$. Now, σ_x is a bit-flip error, and σ_z is a *phase-flip error*, while σ_y is a combination of the two. The error syndrome is given by the 2^8 projections

$$
\begin{aligned}
Q_0(i_1, i_2, i_3) &= P^+_{i_1} \otimes P^+_{i_2} \otimes P^+_{i_3} + P^-_{i_1} \otimes P^-_{i_2} \otimes P^-_{i_3} \\
Q_1(i_1, i_2, i_3) &= P^-_{i_1} \otimes P^+_{i_2} \otimes P^+_{i_3} + P^+_{i_1} \otimes P^-_{i_2} \otimes P^-_{i_3} \\
Q_2(i_1, i_2, i_3) &= P^+_{i_1} \otimes P^-_{i_2} \otimes P^+_{i_3} + P^-_{i_1} \otimes P^+_{i_2} \otimes P^-_{i_3} \\
Q_3(i_1, i_2, i_3) &= P^+_{i_1} \otimes P^+_{i_2} \otimes P^-_{i_3} + P^-_{i_1} \otimes P^-_{i_2} \otimes P^+_{i_3},
\end{aligned}
$$

where $P_i^{\pm}$ is the projection onto the state

$$\frac{1}{\sqrt{2}}\sigma_x^{(i)}(|000\rangle \pm |111\rangle).$$

Provided $\sum_k |a_{k,0}|^2$ is close to 1 (small error probability), the fidelity can be shown to be closer to 1 than without error-correction.

To increase the fidelity further, one has to increase the length of the code. For classical codes, Shannon derived a bound on the minimum length needed for a given channel. This called the **capacity** of the channel. Moreover, he showed that this bound can in principle be achieved.

Shannon's formula for the capacity is given by

$$R = \max_X [H(X, Y) - H(X) - H(Y)],$$

where Y is the output of the channel with given input X.

Holevo, Schumacher and Westmoreland derived a quantum version of this result in the case of classical input to a quantum channel:

$$\chi(\mathcal{E}) = \sup_{\{p_j,\rho_j\}} \left[S\left(\sum_j p_j \mathcal{E}(\rho_j) \right) - \sum_j p_j S(\mathcal{E}(\rho_j)) \right].$$

With Nilanjana Datta (Cambridge) we have a new proof of this formula in the form of a quantum version of the Feinstein lemma:

Theorem Let $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a quantum channel with Holevo capacity $\chi(\mathcal{E})$. Given $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ there exists $N \geq 2^{n(\chi(\mathcal{E})-\epsilon)}$ and there exist product states $\rho_1^{(n)}, \ldots, \rho_N^{(n)}$ on $\mathcal{H}^{\otimes n}$, and positive operators $Q_1^{(n)}, \ldots, Q_N^{(n)} \in \mathcal{B}(\mathcal{K}^{\otimes n})$ such that $\sum_{k=1}^N Q_k^{(n)} \leq \mathbf{1}$ and

$$\mathrm{Trace}\left[\mathcal{E}^{\otimes n}(\rho_k^{(n)})Q_k^{(n)}\right] > 1 - \epsilon$$

for each k.