# Quantum Information Theory

Benjamin Schumacher
*Department of Physics*
*Kenyon College*
*Gambier, Ohio USA*

May-June 1998

Many thanks to:

- University of Innsbruck – H. Weinfurter, ...
- Collaborators – M. Nielsen, R. Jozsa, M. Westmoreland, W. Wootters, ...

## Outline of lectures

1. Introduction
2. States and dynamics
3. Entropy and fidelity
4. Classical information via quantum channels
5. Noisy quantum channels
6. Coherent information and quantum error correction
7. Applications of the ideas

## Some things <u>not</u> discussed

- General definitions of **quantum entanglement**
- Entanglement **purification** protocols
- Quantum **teleportation**
- Quantum **error-correcting codes**
- Quantum **conditional entropy** formalism
- Quantum computer **algorithms** (including computational **complexity** issues)
- Quantum **cryptography**
- **Experiments**
- **Infinite-dimensional** Hilbert spaces

## Physics of information

**I.** No information without representation!

*Information must be "embodied" in the state of a physical system.*

**II.** No processing without a process!

*Information processing must be accomplished by dynamical evolution of a physical system.*

## What is information?

Information is always associated with the ability to perform some specific task.

## How much information?

What physical resources are required to perform a given task successfully?

In each case we must:

- Specify task
- Specify type of resource
- Define “success”

## Two fundamental tasks

**I.** Using quantum systems to send ordinary "classical" information.

Success: Probability of error $P_E \to 0$

**II.** Using quantum systems to send "quantum information".

Success: Fidelity $F \to 1$.

Issues to consider: Noiseless and noisy channels, capacities, relations between problems, etc.

## “Classical” information theory

Basic resource: binary digit or **bit** (0 or 1)

Alice sends: $n$ equally probable messages
$X = x_1, \ldots, x_n$ with $p_k = 1/n$

Measure of information: **entropy**

$$\boxed{H(X) = \log_2 n}$$

Suppose $n = 16$, so that $H(X) = 4$.

Encode the message using strings of 4 bits.

$$\begin{array}{ccc} x_1 & \to & 0000 \\ x_2 & \to & 0001 \\ \cdots & & \cdots \\ x_{16} & \to & 1111 \end{array}$$

Now suppose $n = 3$, so that $H(X) = 1.585$.

What does it mean to represent a message by 1.585 bits?

Consider **blocks** of 5 independent messages: $(X_1 X_2 X_3 X_4 X_5)$. # of blocks is $3^5 = 243$.

Use strings of 8 bits. # of strings is $2^8 = 256$.

**Code:**

$$\begin{array}{ccc} x_1 x_1 x_1 x_1 x_1 & \to & 00000000 \\ \cdots & & \cdots \\ x_3 x_3 x_3 x_3 x_3 & \to & 11110010 \end{array}$$

$$\frac{8 \text{ bits}}{5 \text{ messages}} = 1.6 \text{ bits per message}$$

## Shannon entropy

In general, not all messages are equally likely!

Suppose message $x_k$ has probability $p_k$.

Define the **entropy** $H(X)$ as

$$\boxed{H(X) = -\sum_k p_k \log_2 p_k}$$

Agrees with previous definition of $H(X)$ when all the $p_k$'s are equal.

Obvious analogy to the Boltzmann-Gibbs entropy of (classical) statistical mechanics!

## Noiseless Coding Theorem

$H(X)$ is the number of bits needed to faithfully represent the message $X$.

We'll need: coding using large blocks
small probability of error

How it works: **Typical sequences**

Consider sequences of $N$ independent $X$'s: $(X_1 \cdots X_N)$

**Fact:** For sufficiently large $N$, the set of $N$-sequences is the union of two disjoint subsets:

- A subset $\mathbf{T}$ of "typical" sequences, containing about $2^{NH(X)}$ elements.
- A subset $\mathbf{T}^{C}$ of "atypical" sequences, having very small total probability.

(Hidden details: "about", "very small")

Our strategy:

- Code typical sequences (in $\mathbf{T}$ ) by binary strings of length about $NH(X)$. (There are about $2^{NH(X)}$ of these.)
- Code atypical sequences (in $\mathbf{T}^{C}$) by $00000\cdots000$.

If the sequence is atypical, we will make an error when we decode—but these sequences are unlikely ($P \to 0$ as $N \to \infty$).

We represent the messages with high reliability using about $H(X)$ bits per message.

## Properties of entropy

- If $X$ can take on $n$ values,

$$0 \leq H(X) \leq \log_2 n$$

- **Joint** entropy:

$$H(X,Y) = -\sum_{j,k} p(x_j, y_k) \log_2 p(x_j, y_k)$$

$H(X,Y,Z)$, etc., defined similarly.

- **Conditional** entropy:

$$H(X|y_k) = -\sum_{j} p(x_j|y_k) \log_2 p(x_j|y_k)$$

Average conditional entropy:

$$H(X|Y) = \sum_{k} p(y_k) H(X|y_k)$$

$$
\begin{aligned}
H(X|Y) &= -\sum_{j,k} \underbrace{p(y_k)p(x_j|y_k)}_{p(x_j,y_k)} \log_2 p(x_j|y_k) \\
&= -\sum_{j,k} p(x_j,y_k) \log_2 \left( \frac{p(x_j,y_k)}{p(y_k)} \right) \\
&= -\sum_{j,k} p(x_j,y_k) \log_2 p(x_j,y_k) \\
&\qquad + \sum_{j,k} p(x_j,y_k) \log_2 p(y_k) \\
H(X|Y) &= H(X,Y) - H(Y).
\end{aligned}
$$

- $H(X) \leq H(X,Y)$
- **Useful inequality**. Suppose $p_k$ and $q_k$ are sets of probabilities. Then

$$\sum_k p_k \log_2 q_k \leq \sum_k p_k \log_2 p_k$$

Prove from $\ln x \leq x - 1$, with $x = q_k / p_k$.

- Subadditivity:

$$H(X,Y) \le H(X) + H(Y)$$

  Prove using the useful inequality above.

- $H(X,Y) = H(X) + H(Y)$ if and only if $X$ and $Y$ are independent.

  $X$ and $Y$ independent means that $p(x_j, y_k) = p(x_j)p(y_k)$

- $H(X|Y) \le H(X)$
- "Strong" subadditivity:

$$H(X,Y,Z) + H(X) \le H(X,Y) + H(X,Z)$$

  Follows from subadditivity of the conditional entropy.

## Mutual information

$$
\begin{aligned}
H(X:Y) &= H(X) + H(Y) - H(X,Y) \\
&= H(X) - H(X|Y)
\end{aligned}
$$

Mutual information tells how much information is transferred through a noisy channel.

Let $X =$ input of channel, with probabilities $p(x_j)$

$Y =$ output of channel, with conditional probabilities $p(y_k|x_j)$

Channel properties contained in $p(y_k|x_j)$.

$H(X:Y)$ is the average amount of information gained about the input $X$ when the output $Y$ is specified.

## Noisy coding theorem

Given $p(y_k|x_j)$. Define **channel capacity** $C$ by

$$C = \max H(X:Y)$$

where the maximum is taken over all input distributions $p(x_j)$. Then

- Up to $C$ bits of information can be sent through the channel with arbitrarily low probability of error.

  (To achieve this, we'll need to send large blocks of messages using many similar channels.)

- If we try to send more than $C$ bits through the channel, the probability of error is bounded away from zero—and in fact, $P_E \to 1$ as the channel is used repeatedly.

## Quantum entropy

Consider a quantum system $Q$ in a state $\rho^Q$.

Define **entropy** $S(\rho^Q)$

$$S(\rho^Q) = -\mathrm{Tr}\,\rho^Q \log_2 \rho^Q$$

(Sometimes we will write $S^Q$ instead of $S(\rho^Q)$.)

- Let $d = \dim \mathcal{H}^Q$. Then $0 \leq S^Q \leq \log_2 d$.
- $S^Q = 0$ if and only if $\rho^Q$ is a pure state—that is, $\rho^Q = |\psi\rangle\langle\psi|$ for some $|\psi\rangle$.
- For a composite system $RQ$, $S^Q$ can be greater than $S^{RQ}$. (For example, $RQ$ may be in a pure entangled state.)
- $S^Q$ is related to the entropy of quantum statistical mechanics.

## Next time

- Density operators
- Entanglement and ensembles
- Generalized (positive-operator) observables
- Generalized time evolution (open systems)
- Representation theorems for trace preserving, completely positive linear maps

$\boxed{\text{“The Church of the Larger Hilbert Space”}}$