COMPSCI 755: Unconventional Models of Computation

Lecturers:

Joshua Arulanandham, Room 576, ext 87595,
e-mail : hi_josh@hotmail.com

Professor Cristian S. Calude, Room 575, ext 85751,
e-mail : cristian@cs.auckland.ac.nz

Topics

- A short guide to literature
- Why 'unconventional models of computation'?
- Fundamental mathematical constraints on computation
- Natural algorithms
- DNA computation
- P systems
- Fundamental physical constraints on computation
- The billiard-ball model of computation
- Quantum computation
- Relativistic computation
- Cellular automata
- Potential future computing technologies
- Implications for the mind theories

Format

The course will consist of:

1. Lectures introducing each topic and reviewing the readings
2. Reading assignments from the primary research literature
3. Written assignments to encourage and verify participation
4. Final project
5. Written exam with questions from the topics discussed in class

Assessment

You will have *weekly reading assignments*, *fortnightly written assignments*, each worth 4%, and *a project* worth 35%.

You will be given a few papers from the primary research literature to read every week. Skim through the readings, and read more thoroughly, at your leisure, the ones that you think you will get most out of. Don't worry if you don't understand every bit of what you read. In this course we will be reading materials that span a wide range of levels of depth and sophistication, and not everyone will understand every phrase and formula in every paper.

The *exam* is worth 40%.

In the written assignments (1-2 pages) you may do one of the following:

1. *Summarize* what you learned from the fortnight's lectures and/or readings.
2. *Write* a summary, review, or critique of one or more of the articles/chapters that you read.
3. *Describe and elaborate on* any creative or interesting ideas/thoughts relating to the subject matter that might have been stimulated in your mind as you were listening to/reading/reflecting on the material.
4. *Set up and carry out* any interesting analysis, calculation or simulation relating to any of the quantitative/technical ideas covered during the fortnight.
5. *Correct* any statement that was made in class or in one of the readings which in your opinion is wrong or inaccurate (explain why).
6. *Do a bit of research* on your own. Summarize what you learned and cite your references.

A Short Guide to Literature

Textbook

C. S. Calude, G. Păun. *Computing with Cells and Atoms*, Taylor & Francis Publishers, London, 2001.

Recommended Books

- J. Gruska. *Quantum Computing,* McGraw-Hill, London, 1999.

  The most comprehensive textbook in Quantum Computing.

- J. G. Hey and R. W. Allen, (eds.). *Feynman Lectures on Computation,* Addison-Wesley, Reading, Massachusetts, 1996.

  Feynman's lecture notes from the course "The Potentialities and Limitations of Computing Machines" taught at Caltech in the early eighties.

- J. G. Hey (ed.). *Feynman and Computation. Exploring the Limits of Computers,* Perseus Books, Reading, Massachusetts, 1999.

  Companion volume to *Feynman Lectures on Computation*, this book collects old and recent articles on the physics of computing by Feynman and his colleagues in physics, electrical engineering, and computer science who were guest lecturers in his course.

- Gh. Păun, G. Rozenberg, A. Salomaa. *DNA Computing. New Computing Paradigms,* Springer-Verlag, Berlin, 1998.

  The best textbook in DNA Computing.

- C. P. Williams, S. H. Clearwater. *Explorations in Quantum Computing,* Springer-Verlag, New York, 1997.

  A very good book in Quantum Computing which comes with software which some may be interested in playing with.

- C. P. Williams, S. H. Clearwater. *Ultimate Zero and One: Computing at the Quantum Frontier,* Springer-Verlag, Heidelberg, 2000.

  A continuation of *Explorations in Quantum Computing.*

Projects

- Improve the P system implementation of Bead-Sort by reducing the number of symbols involved (in symport, antiport) close to 2 (as in biology). Also, do a more precise time complexity analysis of Bead-Sort by taking into consideration, factors such as air friction which affect the bead fall.
- Simulate using a CA, the physical system used to solve SAT.
- An analog electrical circuit for solving the 'Graph Connectivity' problem was discussed in class. Describe a similar approach to solve the 'Maze Problem': Given a maze, find a route leading from the source to the destination.
- Simulate using a CA, the liquid-based natural algorithm for finding the average of n integers.
- Is it possible to imagine a 'natural' representation of text? If so, construct natural algorithms to do fast text pattern matching.
- Present a review of DNA computing approaches to solve at least five interesting "hard" computing problems (other than those discussed in class).
- Explore natural physical phenomena (other than beads falling down) that exhibit 'sorting' capability. Describe in detail one such phenomenon and a possible implementation.
- An approach to solve the Travelling Salesman problem (TSP) using a Particle Physics metaphor was discussed in class. Imagine yet another physics metaphor to solve TSP and simulate it.
- Explore problems (other than sorting) that can be solved using

beads and rods. Describe a beads-and-rods solution to one such problem (it should exhibit some 'natural law' in action).

- Critical review of Physical Church-Turing Thesis.
- Critical review of an attempt to break the Turing barrier.
- Program for a non-trivial quantum algorithm.
- Analysis of the capability of a restricted class of quantum algorithms.
- Analysis of randomness in quantum computation.
- Is quantum computation relevant for the brain activity?

Why Unconventional Models of Computation?

The computer seems to be the only important instrument ever to get exponentially better as it gets cheaper. Its capacity for handling information has been growing about ten million times faster than it did in nervous systems during our entire evolution. The power

- doubled every two years up until 1980s,
- doubled every 18 months in the 1980s (Gordon Moore's 1965 law), and
- is now doubling each year.

By 1993 personal computers provided 10 MIPS (MIPS = million of instructions per second), by 1995 it was 30 MIPS, in 1997 it was over 100 MIPS, now it's about 200 MIPS.

For the sake of a comparison: the human retina uses about 1,000 MIPS to handle edge and motion detectors, while the whole human brain—which is roughly 100,000 times larger than the retina—is worth perhaps 100 million MIPS.

Computers are reading text, recognizing speech, and robots are driving themselves across Mars.

Yet, this exponential race will not guarantee solutions to the many intractable/undecidable problems challenging computer science.

Even worse, it is predictable that this trend of conventional technology will hit the wall in less than 20 years. This is a reason to believe that conventional computation is approaching a critical phase where new technologies will be required to provide significant further progress.

Fundamental Mathematical Constraints on Computation

Church-Turing Thesis

Church-Turing Thesis, a prevailing paradigm in classical computation theory, states that no realizable computing device can be "globally" more powerful, that is, aside from relative speedups, than a universal Turing machine. The modern form of Church-Turing Thesis states that

> *any "reasonable" model of computation can be effectively simulated by a (probabilistic) Turing machine.*

The above statement is a *thesis*, and not a theorem, as it relates an informal notion–a realizable computing device–to the mathematical notion of (probabilistic) Turing machine. Here are some reasons supporting Church-Turing Thesis:

- *Philosophical argument*: Due to Turing's analysis it seems very difficult to imagine some other method which falls outside the scope of his description.
- *Mathematical evidence*: Every mathematical notion of computability which has been proposed was proven equivalent to Turing computability.
- *Sociological evidence*: No example of classical computing device which cannot be simulated by a Turing machine has been given, i.e., the thesis has not been disproved despite having been proposed for more than 60 years.

Church-Turing's Thesis includes a syntactic as well a physical claim. In particular, it specifies which types of computations are physically realisable. According to Deutsch (1982):

> *The reason why we find it possible to construct, say, electronic calculators, and indeed why we can perform mental arithmetic, cannot be found in mathematics or logic.* The reason is that the laws of physics "happen" to permit the existence of physical models for the operations of arithmetic *such as addition, subtraction and multiplication. If they did not, these familiar operations would be non-computable functions. We might still know* of *them and invoke them in mathematical proofs (which would be presumably called "non-constructive") but we could not perform them.*

Church-Turing Thesis was challenged by logicians (Kalmar, Davis, Kreisel), computer scientists (Rosen, Hogarth, Siegelmann) and physicists (Landauer, Svozil). For example, Davis asks himself:

> *"...how can we ever exclude the possibility of our presented, some day (perhaps by some extraterrestrial visitors), with a (perhaps extremely complex) device or "oracle" that "computes" an uncomputable function?"*

Thinking is an essential, if not the most essential, component of human life–it is a mark of "intelligence". Descartes placed the essence of being in thinking. Church-Turing Thesis has been used to approach formally the notion of "intelligent being". In simple terms, Church-Turing Thesis was stated as follows:

> *What is human computable is computable by a universal Turing machine.*

FAQ

Q: What has computing to do with physics?

A: Information, essential for any form of computing, is not a pure abstract entity. In fact, measuring, communicating and computing are *all* about exchanging information. Information is inevitably tied to a physical embodiment or representation; it can be engraved on stone tablets, represented by holes punched in a card, or by a present/absent charge or by a spin up or down.

R. Landauer: "The computer has made us aware that information is a physical entity".

D. Deutsch: "The reason why we find it possible to construct, say, electronic calculators, and indeed why we can perform mental arithmetic ... *is that the laws of physics "happen" to permit the existence of physical models for the operations of arithmetic* such as addition, subtraction and multiplication.

Q: What is quantum computing?

A: Quantum computing is the quest to understand what sort of machines do useful computation in a universe described by quantum mechanics. Today the subject is mostly theoretical, but tentatively, slowly and hesitantly groping towards some practical applications.

Q: What is quantum mechanics?

A: Quantum mechanics describes the behaviour of very small objects, the size of atoms or smaller, in contrast with relativity theory which describes the laws of larger everyday objects. Interestingly, particles do not behave in the same way as larger everyday objects, such as billiard balls. If we strike a billiard ball in a very precise way and we know its exact initial position, then we can predict with (theoretical) certainty where it will go. The same is not true for particles.

Q: If classical mechanics is wrong, why do we still use it?

A: Classical mechanics is flawed *only* when dealing with the very small (atomic size) or the very fast (near the speed of light). For everyday things, classical physics does an excellent job.

Q: What are the main features of quantum mechanics?

A: Here are five:

- *Quantisation*: observable quantities do not vary continuously, but come in discrete chunks called quanta.
- *Randomness*: physical reality is irreducibly random.
- *Interference*: the outcome of a quantum process depends on all the possible histories of that process.
- *Superposition*: the ability of carrying out computations with "blends" of states, superpositions.
- *Entanglement*: two spatially separated and non-interacting quantum systems, that have interacted in the past could have some locally inaccessible information in common – information which cannot be accessed in any experiment performed on either of them alone.

Q: Are these features useful for quantum computing?

A: Quantisation makes quantum computing possible at all. Randomness, superposition and interference make quantum computers more powerful than classical ones. Entanglement is useful in quantum cryptography.

Q: In contrast with classical computers which use bits, quantum computers process quantum bits. What's the difference?

A: We have four concepts:

- The mathematical bit (0 or 1).
- A classical system representing a bit, called Cbit.
- The mathematical quantum bit.
- A quantum system (event) in which we have two possible mutually exclusive outcomes realising a quantum bit, called Qbit.

All knowledge of the quantum system is based upon acts of observation. The information derived from an elementary act of observation is no more than a single bit, but ... Before measurement, the system can be in any intermediate quantum state, that is in a superposition of 0 and 1, in a (sort of) mixture of 0 and 1 containing both classical (contradicting) states at once; after observation, we get either 0 or 1 with some probability.

Q: Are Qbits responsible for the famous "exponential explosion"?

A: Yes. Any classical register composed of three Cbits can store in a given moment of time only one out of eight different numbers because the register can be in only *one* out of eight possible configurations:

$$000, 001, 010, 011, 100, 101, 111.$$

A quantum register composed of three Qbits can store in a given moment of time *all* eight numbers in a quantum superposition. If we increase the number of Qbits to the register, then we increase its storage capacity exponentially: three Qbits can store eight different numbers at once, four Qbits can store sixteen different numbers at once, in general n Qbits can store 2^n numbers at once.

Note that it would require vast resources to simulate even a small quantum system on a conventional computer, as such a simulation would require keeping track of exponentially many states.

Q: What can you do with superpositions?

A: We can perform operations on them. During such an operation each number in the superposition is affected and as the result we obtain a massive parallel computation albeit in just one piece of quantum hardware. As in the solution of the Merchant's Problem, we can act at once on all stacks of coins. A quantum computer offers an enormous gain in time and memory.

Q: Where is the catch?

A: Qbits suffer from a major limitation which doesn't affect Cbits: given a superposition of Qbits in some state, there is nothing one can do to the Qbits to be able to extract what that state is in.

Q: Is this the only limitation?

A: No. There are also limited possibilities to extract the information contained in a Qbit. Learning the value of a combination of Cbits is so easy (you print it out) that it is not even explicitly regarded as a part of the computation. More importantly, Cbits are not altered by "reading" them. Not anymore with Qbits: we can extract the information from a Qbit *only* by measurement, a process which:

(a) is probabilistic (recall the intrinsic randomness of quantum mechanics), and

(b) affects the state of the Qbit; simple operations, like copying a Cbit into another Cbit, are not available in quantum computing.

Q: So, under these wretched conditions, what are Qbits good for?

A: The art is to produce a superposition in which the useful information has a high probability of being indicated by measurement and the unimportant information can be expected to appear with probability close to zero. To make the result safe, one has to be able to easily *confirm* the result of the computation ...

Q: Can you give an example?

A: Peter Shor has shown in 1994 that quantum factoring integers is dramatically faster than any *known* classical algorithm. The obvious method of factoring a number N represented by n bits requires about $2^{n/2}$ trials.

A much smarter algorithm (based on sophisticated mathematical results) does the job in approximately $2^{c\sqrt[3]{n}}$ steps, where c is a constant; still, factoring a number of a million of bits would require a time larger than the age of the Universe.

Shor has observed that the factoring problem can be rephrased in terms of a search for how often some "period" of a finite sequence is repeating itself within the sequence. For example, the sequence

123412341234

has 1234 as period which repeats itself three times. Periods may be seen as waves, undulating streams.

The quantum algorithm is polynomial-time in the number of bits necessary to represent the number to be factored. Confirming the result is easy.

Q: What about Grover's quantum algorithm?

A: Start with an example. Searching a telephone directory containing n names in alphabetic ordering requires about $\log_2 n$ steps. Searching the name in the telephone directory, when the telephone number is known, is much more difficult because the list is unsorted with respect to telephone numbers. We need about $n/2$ steps on average and n steps in the worst case.

Looking up a name given a number is exponentially more difficult than looking up a number given a name.

Grover's quantum algorithm searches an unsorted list very fast; his procedure needs roughly $\pi/4\sqrt{n}$ quantum steps.

Q: What will quantum computers be good at?

A: These are the most important applications currently known:

- *Cryptography*: RSA code breaking, perfectly secure communication.
- *Searching*: fast searching (Grover's algorithm).
- *Simulating*: efficient simulation of quantum-mechanical systems.

Q: Can I learn quantum computing without understanding quantum mechanics?

A: Yes, you can. Recently, L. Fortnow has published in *Theoretical Computer Science* (292 (2003), 597–610) a nice paper titled

"One complexity theorist's view of quantum computing"

in which he shows that a large part of quantum computing can be understood without any knowledge of quantum mechanics.

Arguably, the amount of quantum mechanics required for the mainstream quantum computing is limited : this parallels the situation of classical computing, where computer scientists need not know much about transistors and the way they work.

Q: How soon a quantum computer might be built?

A: Lab experiments show that the basic principles of quantum computing are sound. To realistically compete with classical computing, quantum computing must be carried out on significantly larger scales ... It is unreasonable to make predictions; however, it is reasonable to expect that small milestones will continue to appear.

Information and Computation Are Physical

An operation is "logically reversible" if it can be run *backwards*, that is, if its inputs can always be deduced from the outputs. Most logical gates are irreversible; a typical example is the NAND gate

$$(a, b) \mapsto \neg(a \wedge b) \tag{1}$$

which has two input bits and only *one* output bit. We cannot recover a unique input from the output bit because the result 1 can be obtained from three distinct inputs: $(0, 0), (0, 1), (1, 0)$.

Assume we operate the gate NAND with two Boolean variables, a, b, and suppose that the four initial states, $(0, 0), (0, 1), (1, 0), (1, 1)$, have the same probability distribution, $\frac{1}{4}$. Then, the initial entropy, which is calculated with Shannon's formula:

$$H = -\sum_i p_i \cdot \log p_i,$$

is then

$$H_{initial} = -4 \cdot \left(\frac{1}{4} \log \frac{1}{4}\right) = 2 \text{ bits.}$$

The result will be a system with only two possible states, 0 and 1, the outcome 0 appearing with probability $\frac{1}{4}$ and the outcome 1 appearing with probability $\frac{3}{4}$. Consequently, the final entropy is

$$H_{final} = -\left(\frac{3}{4} \log \frac{3}{4} + \frac{1}{4} \log \frac{1}{4}\right) = 2 - \frac{3}{4} \log 3 \text{ bits,}$$

which means a loss of $H_{initial} - H_{final} = \frac{3}{4} \log 3$ bits.

Assume now that we operate the gate

$$(a, b) \mapsto (a \vee b, a \wedge b),$$

and, again, suppose that the four initial states of the Boolean variables a, b have the same probability distribution, $\frac{1}{4}$. This gate has finally only *three* final states, namely $(0, 0), (1, 0), (1, 1)$, two of them with probability $\frac{1}{4}$ and one with probability $1/2$. Consequently, the final entropy is

$$H_{final} = -\left(2 \cdot \frac{1}{4} \log \frac{1}{4} + \frac{1}{2} \log \frac{1}{2}\right) = 1.5 \text{ bits}.$$

In this case, the gate decreases the entropy by 0.5 bits.

The first gate is "more irreversible" than the second one, since it decreases more the entropy.

In thermodynamics the entropy is defined by

$$S = -k \cdot \sum_i p_i \cdot \ln\,(p_i),$$

where $k \approx 1.38 \times 10^{-23}$ joule/°kelvin is Boltzmann's constant.

This notion is coupled to energy through the temperature T of the system: when the entropy of a system is decreased by some amount, the system dissipates energy equal to the amount of entropy reduction times the temperature. Von Neumann noticed that the two entropies are related by some constant factor, so they are in fact the same notion. When the probability distribution of the system is changed so that the entropy H is decreased by 1 bit, then the entropy S is decreased by $k \cdot \ln 2$ joule/°kelvin, and the system dissipates $kT \cdot \ln 2$ joules of energy in the form of heat.

Does the above analysis apply to computation?

In 1961 Landauer produced evidence for the affirmative answer. To operate a computer we have to make sure that distinct logical states are represented by distinct physical states. A set of n bits has n degrees of freedom; they correspond to 2^n physical states. If we erase n bits, say we reset all to 0, then we have compressed 2^n logical states into a single state, a loss of entropy. The irreversible loss information increases temperature of the system, which means, heat dissipation. Consequently, operations which are not one-to-one, which map distinct logical states into a common one, *cost energy.* This cost is expressed by *Landauer's principle*:

erasure of information is a dissipative process.

Here is a simple "home" example. We need two basketballs to design a system of representing information. Put one on the floor by your left foot and hold the other in your (right) hand.

Zero (0) is represented by the ball on the floor; one (1) is represented by the ball in your hand.

Assume that we want to *erase the bit 1*, that is the bit in your hand. To do this you have to *drop the ball.* Simple?

Not really, as the ball does not get *directly* into the floor (to become a 0), but in fact bounces for a while. With a perfectly elastic basketball and a good hard floor the ball may bounce close to your hand, i.e., to the 1 position!

To settle down into 0 the ball has to encounter friction, with the air molecules and the floor. Eventually friction slows down the ball, so 1 has been erased. We could do it because *the energy from bouncing the ball has been transmitted to the floor and the air.* In a vacuum with a perfect friction-less floor erase would be impossible! Energy is consumed in the process of erasure.

Irreversible operations, as the NAND gate, the binary addition $(a, b) \mapsto (a \oplus b, a \wedge b)$ (sum and carry) and the real addition $(x, y) \mapsto x + y$, dissipate energy. Recall that $a \oplus b$ is 1 only if a and b have different values, i.e., $a = 0, b = 1$ or $a = 1, b = 0$.

The above irreversible operations can be easily simulated by reversible ones. A reversible version of the NAND gate is, for example, Toffoli's gate

$$(a, b, c) \mapsto (a, b, c \oplus (a \wedge b)). \tag{2}$$

Indeed, (2) is a reversible 3-bit gate that flips the third bit if the first two both take the value 1 and does nothing otherwise. Hence, the third output bit becomes the NAND of a and b in case $c = 1$. The price paid to get reversibility consisted in adding a new variable c.

Question: Why NAND? Reason: a single NAND gate is as good as having both AND and NOT, they are *universal*! (Prove it!)

Similar tricks can be used to produce reversible versions of the binary addition, $(a, b) \mapsto (a, a \oplus b, a \wedge b)$ and real addition $(x, y) \mapsto (x + y, x - y)$. In the first case we replicated the first variable a; in the second case we added a new component storing some additional value.

A computer may be fully reversible and yet dissipate energy! The important point is that the laws of physics allow for technologies to make reversible computers operate with negligible dissipation. To build a reversible computer one needs only two types of logical gates, say AND and NOT. Clearly, the NOT gate is reversible as its composition with itself gives the initial input. However, the AND gate is irreversible.

To make a reversible variant of the gate AND we need to ensure that we have the same number of output lines as input ones, so, in principle, we can just add some "garbage" output lines to solve the problem. However, this may not be enough, as we want to guarantee also universality! One possibility is Toffoli's reversible 3-bit gate which uses in addition to a, b a control bit c. Input bits a and b do not change their states; the control bit, however, will change its state, but only when $a = b = 1$. Toffoli's truth table is the following:

input			**output**		
a	b	c	a	b	c
0	0	0	0	0	0
0	1	0	0	1	0
1	0	0	1	0	0
1	1	0	1	1	1

input			**output**		
a	b	c	a	b	c
0	0	1	0	0	1
0	1	1	0	1	1
1	0	1	1	0	1
1	1	1	1	1	0

Toffoli's gate.

Fredkin's reversible 3-bit gate also uses in addition to a, b a control bit c in the following way: a) if $c = 0$, then the values of a, b are transmitted unaltered, i.e., the output is the pair (a, b), b) if $c = 1$, then the values of a, b are switched to the opposite output, i.e., the output is the pair (b, a). Its truth table is the following:

input			**output**		
a	b	c	a	b	c
0	0	0	0	0	0
0	1	0	0	1	0
1	0	0	1	0	0
1	1	0	1	1	0

input			**output**		
a	b	c	a	b	c
0	0	1	0	0	1
0	1	1	1	0	1
1	0	1	0	1	1
1	1	1	1	1	1

Fredkin's gate.

Fredkin's gate is universal (prove it!). Fredkin's gate has often been used for photon based gates where a 1 represents a photon and a 0 simply denotes the absence of a photon; non-linear optics is used to control the output of an interferometer. The number of ones cannot change as the number of photons cannot change–absorption is not allowed for reversible gates.

To achieve reversibility we added more outputs than are required for the computed functions: these outputs, called *garbage* bits, are a necessary consequence of reversible logic. Consequently, one may wonder whether we have only postponed the energy cost; garbage bits can be irreversibly erased, but that would require to pay Landauer's price ...

We do not need to erase the garbage bits!

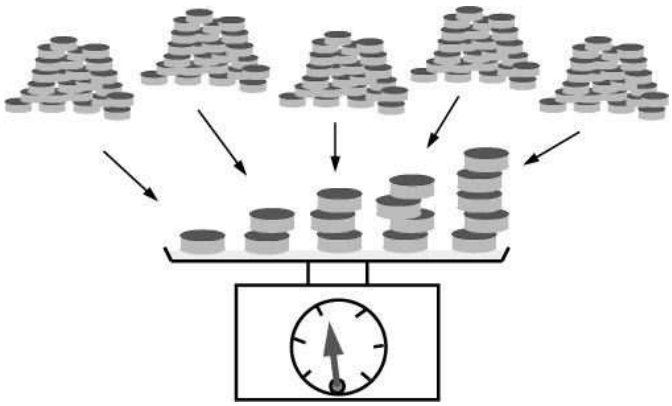
Bennett realised that reversible computer can run forward to the end of a computation, print out a copy of the answer (a logically reversible operation) and then reverse all of its steps to return to its initial configuration. So we can remove the garbage without any energy cost.

In practice, Landauer's limit seems not to be an important engineering principle. But as computing hardware continues to shrink in size, it may become important to beat Landauer's limit, for example, to prevent the components from melting. Then reversible computation may be one, if not the only one, option.

The following problem illustrates the "quantum" approach to problem solving.

Merchant's Problem

A merchant learns than one of his five stacks of coins contains only false coins, 0.01 grams heavier than normal ones. Can he find the odd stack by a single "weighing"?



Coin selection

What about the case when more than one stack of coins contains false coins: can we, again with only one single weighting find all stacks containing false coins?

Solution: choose 1,2,4,8,16 coins from each stack!

- What are the limits of the above solutions?
- What about the case when we are allowed to take only just one coin from each stack?
- What about the case when we have infinitely many stacks?

A Bit of Mathematics

Real and Complex Numbers

Let $\mathcal{R}$ denote the set of real numbers. By $\mathcal{C}$ we denote the set of complex numbers, i.e. pairs of reals: $z = (x, y)\,(x, y \in \mathcal{R})$. Here are the rules to operate with complex numbers:

- addition: $z_1 + z_2 = (x_1 + x_2, y_1 + y_2)$, if $z_1 = (x_1, y_1)$, $z_2 = (x_2, y_2)$,
- multiplication: $z_1 \cdot z_2 = (x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1)$,
- conjugation: $z^* = (x_1, -y_1)$,
- norm: $|z| = \sqrt{x^2 + y^2}$.

Matrices

A 2×2 *matrix* is a table of the form

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

An $m \times n$ matrix has m *rows* and n *columns*. The first column of M is the 2×1 matrix

$$\begin{pmatrix} \alpha \\ \gamma \end{pmatrix},$$

and the second column is the the 2×1 matrix

$$\begin{pmatrix} \beta \\ \delta \end{pmatrix},$$

Similarly, the first row is the 1×2 matrix (α, β) and the second row is the 1×2 matrix (γ, δ).

The i, j*–component* of a matrix is the element sitting on the column i and row j. For example, the 1,2–element of M is β; the 2,2–element is δ.

The rules to operate with matrices are:

- transposition: $(\alpha_{i,j})^t_{i=1,\ldots,m,j=1,\ldots,n} = (\alpha_{j,i})_{j=1,\ldots n,i=1,\ldots,m}$,
- addition: $(\alpha_{i,j})_{i=1,\ldots,m,j=1,\ldots,n} + (\beta_{i,j})_{i=1,\ldots,m,j=1,\ldots,n} = (\alpha_{i,j} + \beta_{i,j})_{i=1,\ldots,m,j=1,\ldots,n}$,
- product: $(\alpha_{i,j})_{i=1,\ldots,m,j=1,\ldots,n} \cdot (\beta_{j,k})_{j=1,\ldots,n,k=1,\ldots,r} = (\alpha_{i,1}\beta_{1,k} + \alpha_{i,2}\beta_{2,k} + \ldots + \alpha_{i,n}\beta_{n,k})_{i=1,\ldots,m,k=1,\ldots,r}$,
- scalar multiplication: $a(\alpha_{i,j})^t_{i=1,\ldots,m,j=1,\ldots,n} = (a\alpha_{i,j})_{i=1,\ldots,m,j=1,\ldots,n}$,
- dagger: $((\alpha_{i,j})_{i=1,\ldots,m,j=1,\ldots,n})^\dagger = (\alpha_{j,i})^*_{j=1,\ldots,n,i=1,\ldots,m}$.

Examples

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^t = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix},$$

$$\begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{pmatrix} + \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{pmatrix} = \begin{pmatrix} \alpha_{1,1}+\beta_{1,1} & \alpha_{1,2}+\beta_{1,2} \\ \alpha_{2,1}+\beta_{2,1} & \alpha_{2,2}+\beta_{2,2} \end{pmatrix},$$

$$\begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{pmatrix} . \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_{1,1}\beta_{1,1}+\alpha_{1,2}\beta_{2,1} & \alpha_{1,1}\beta_{1,2}+\alpha_{1,2}\beta_{2,2} \\ \alpha_{2,1}\beta_{1,1}+\alpha_{2,2}\beta_{2,1} & \alpha_{2,1}\beta_{1,2}+\alpha_{2,2}\beta_{2,2} \end{pmatrix},$$

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^\dagger = \begin{pmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{pmatrix},$$

$$I_1 = ((1,0)),\ I_2 = \begin{pmatrix} (1,0) & (0,0) \\ (0,0) & (1,0) \end{pmatrix}.$$

A *unitary* matrix of type $m \times n$ is a matrix M such that $M^\dagger M = I_n$. Here I_n is the identity matrix of type $n \times n$; obviously, I_n is a unitary matrix.

Properties

1. *Multiplying a matrix by the appropriate identity matrix leaves the matrix unchanged.*

2. *For every matrices A, B of types $m \times n$ and $n \times r$, respectively, we have:*

$$(AB)^\dagger = B^\dagger A^\dagger.$$

3. *Let*

$$V = \begin{pmatrix} v_{1,1} \\ v_{2,1} \\ \vdots \\ v_{n,1} \end{pmatrix},$$

be an $n \times 1$ unitary matrix (a column). Then,

$$|v_{1,1}|^2 + |v_{2,1}|^2 + \ldots + |v_{n,1}|^2 = 1.$$

Rudiments of Quantum Theory

Quantum mechanics, one of the pillars of the 20th century physics, describes *closed* (i.e. perfectly isolated from the world) physical systems, usually small systems, of the size of atoms or smaller.

We will use the following postulates:

A. State Postulate: *The state of a closed physical system is completely described by a unitary* $n \times 1$ *matrix of complex numbers.*

Explicitly, a state is given by a column of n complex numbers

$$V = \begin{pmatrix} v_{1,1} \\ v_{2,1} \\ \vdots \\ v_{n,1} \end{pmatrix},$$

such that $V^\dagger V = I_1$. The column V is a unitary matrix of type $n \times 1$ (or, normalised).

State vectors are typically written with a special angular bracket notation, the "ket vector" $V = |\psi\rangle$. The word "ket" was invented by Paul Dirac, one of the founders of quantum mechanics.

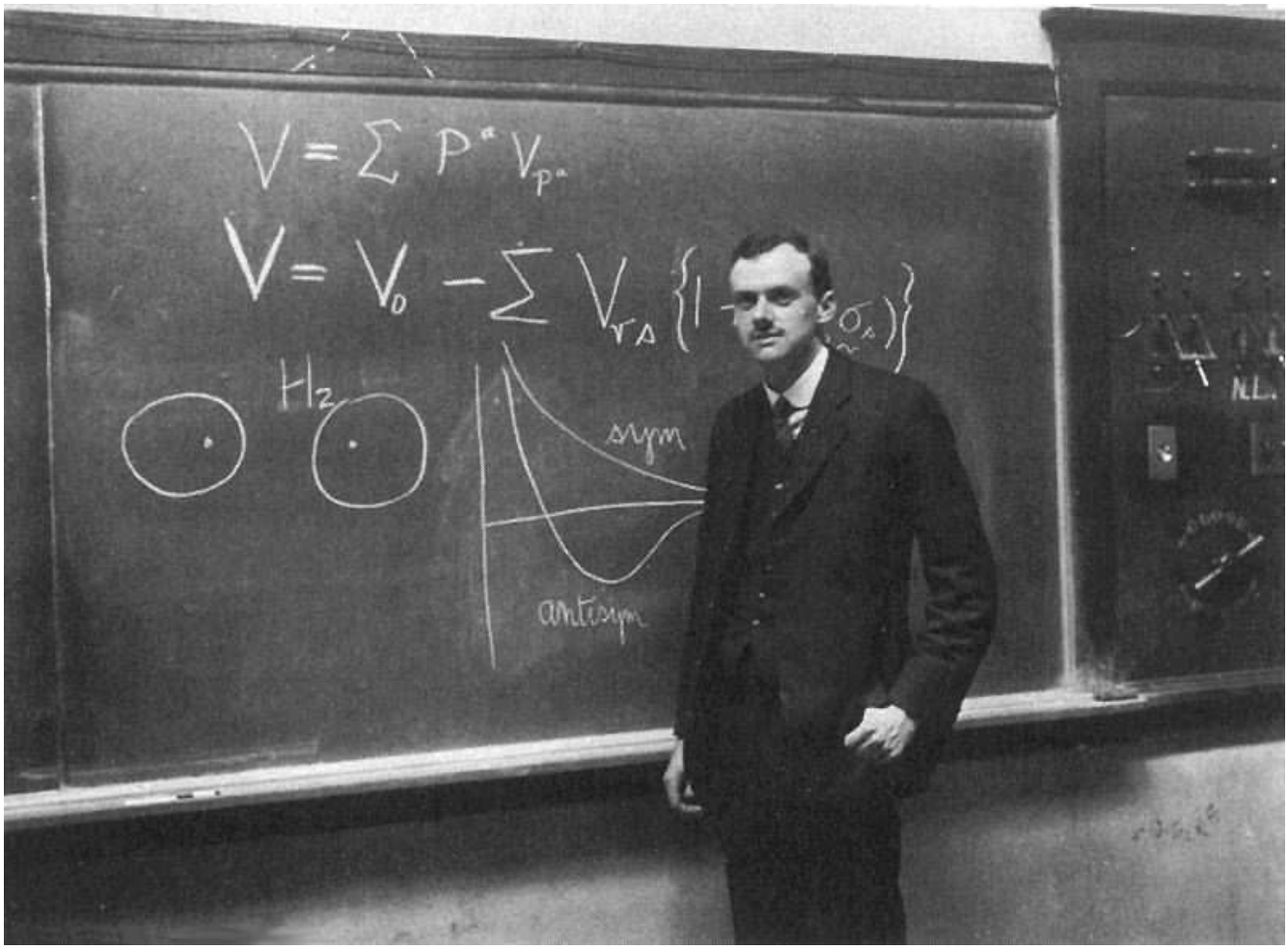


Figure 1: Paul Dirac

Example: Quantum-like coin. A coin can be in two states, head $\odot$ and tail $\oplus$ represented by the following two columns:

$$\odot = \begin{pmatrix} (1,0) \\ (0,0) \end{pmatrix}, \ \oplus = \begin{pmatrix} (0,0) \\ (1,0) \end{pmatrix}.$$

If the coin was shut into a perfectly closed box, then it would start behaving like a truly quantum coin, hence

$$\odot + \oplus = \begin{pmatrix} (\frac{1}{\sqrt{2}},0) \\ (\frac{1}{\sqrt{2}},0) \end{pmatrix},$$

is a legitimate quantum state: the coin is in a *superposition* of $\odot$ and $\oplus$, that is, it can be both head and tail *at the same time.*

B. Evolution Postulate: *A closed physical system in state V evolves in time into a new state $W = UV$, where U is an $n \times n$ unitary matrix.*

In other words, the system changes its states in time and each change is obtained by multiplying the current state with a square unitary matrix U (recall, this means $U^\dagger U = I_n$).

We *have* to check that the resulting column vector W is also a valid quantum state, i.e. $W^\dagger W = I_1$.

C. Born's Measurement Postulate: *When a closed physical system in state*

$$V = \begin{pmatrix} v_{1,1} \\ v_{2,1} \\ \vdots \\ v_{n,1} \end{pmatrix},$$

is measured *it yields outcome i with probability $|v_{i,1}|^2$. Whenever outcome i occurs, the system is left in the state*

$$V = \begin{pmatrix} (0,0) \\ \vdots \\ (0,0) \\ (1,0) \\ (0,0) \\ \vdots \\ (0,0) \end{pmatrix} \leftarrow i^{th}\text{row}$$

(which contains no trace of the information in the pre-measurement state).

Example continued: Quantum-like coin. Assume the coin is in the closed box in the state

$$\odot + \oplus = \begin{pmatrix} (\frac{1}{\sqrt{2}}, 0) \\ (\frac{1}{\sqrt{2}}, 0) \end{pmatrix},$$

and we *measure* its state: with probability $p_1 = |\frac{1}{\sqrt{2}}|^2 = \frac{1}{2}$ we will get the outcome 1, which means that the system will be left in the state

$$\odot = \begin{pmatrix} (1,0) \\ (0,0) \end{pmatrix},$$

and with probability $p_2 = |\frac{1}{\sqrt{2}}|^2 = \frac{1}{2}$ we will get the outcome 2, which means that the system will be left in the state

$$\oplus = \begin{pmatrix} (0,0) \\ (1,0) \end{pmatrix}.$$

We *have* to check that Born's Measurement Postulate is probabilistically sound, i.e. probabilities sum up to one:

$$p_1 + p_2 + \ldots + p_n = 1.$$

Two essential consequences of the Born's Measurement Postulate are:

- *Randomness*: A measurement is fundamentally a probabilistic process: When a physical state that is in a superposition of states is measured, then it collapses into one of its possible states in a completely unpredictable way—we can only evaluate the probability of obtaining various possible outcomes. According to Milburn, *physical reality is irreducible random.*
- *State Change*: A measurement irrevocably disturbs the state. If the state is initially unknown, then there is no way to *determine* it with any conceivable measurement.

From Cbits to Qbits

Classical **bits** are abstractly denoted by 0 and 1. A classical physical system used to represent a bit will be called **Cbit**. For example, the position of gear teeth in Babbage's differential engine, a memory element or wire carrying a binary signal, in contemporary machines, are examples of Cbits.

A Cbit is a system comprising many atoms. Typically, the system is described by one or more continuous parameters, for example, voltage. Such a parameter is used to separate the space into two well-defined regions chosen to represent 0 and 1. Manufacturing imperfections, local perturbations may affect, so signals are periodically restored toward these regions to prevent them from drifting away.

An n-bit register of memory can exist in any of 2^n logical states, from $00\ldots0$ (n zeros) to $11\ldots1$ (n ones).

There are only two unary reversible operations with bits: the identity ($0 \mapsto 0, 1 \mapsto 1$) and the flip ($0 \mapsto 1, 1 \mapsto 0$). The "erase" operation ($0 \mapsto 0, 1 \mapsto 0$) is irreversible. Less trivial reversible operations are available for two bits.

The information derived from an elementary act of observation of a quantum system with only two possible outcomes is no more than a single bit, but *there is more on it than that.* To mark this difference Schumaker has coined the name quantum bit.

An abstract **quantum bit** is a unitary column in $\mathcal{C}^2$:

$$|x\rangle = \begin{pmatrix} a \\ b \end{pmatrix}. \tag{3}$$

If we adopt the notation:

$$|0\rangle = \begin{pmatrix} (1,0) \\ (0,0) \end{pmatrix}, \ |1\rangle = \begin{pmatrix} (0,0) \\ (1,0) \end{pmatrix},$$

then we can write

$$|x\rangle = a|0\rangle + b|1\rangle, \tag{4}$$

where $a, b \in \mathcal{C}$ are such that $|a|^2 + |b|^2 = 1$.

For brevity we will sometime write:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Mathematically, the set $\{|0\rangle, |1\rangle\}$ forms a *basis* for $\mathcal{C}^2$: every $|x\rangle$ can be written in an unique way in the form (4).

The quantum-like coin is an example of quantum bit:

$$\begin{pmatrix} (\frac{1}{\sqrt{2}}, 0) \\ (\frac{1}{\sqrt{2}}, 0) \end{pmatrix}.$$

A **Qbit** is a (typically microscopic) system, such that when measured is always found to be in one of two possible states. For example, an atom or nuclear spin or polarised photon. For example, the state of a spin-$\frac{1}{2}$ particle, represented as

$$|+\frac{1}{2}\rangle \text{ (spin-up) or } |-\frac{1}{2}\rangle \text{ (spin-down).}$$

Unlike the intermediate states of a classical bit (for example, any voltages between the "standard" representations of 0 and 1) which can be distinguished from 0 and 1, but do not exist from an informational point of view, quantum intermediate states cannot be reliably distinguished, even in principle, from the basis states, but do have an informational "existence".

To handle systems of more than one Qbit we need a new operation, the *tensor product*:

$$\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix},$$

A two Qbit system can be represented by a unitary vector in the tensor product of two copies of $\mathcal{C}^2$, i.e., the space $\mathcal{C}^2 \otimes \mathcal{C}^2$. If $|0\rangle$ and $|1\rangle$ is the basis for $\mathcal{C}^2$, then denoting

$$|00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$|10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

we obtain a basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ for $\mathcal{C}^2 \otimes \mathcal{C}^2$ (hint: $|0\rangle \otimes |0\rangle = |00\rangle, |0\rangle \otimes |1\rangle = |01\rangle, |1\rangle \otimes |0\rangle = |10\rangle, |1\rangle \otimes |1\rangle = |11\rangle$).

For three Qbits we have:

$$\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} \otimes \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} ace \\ acf \\ ade \\ adf \\ bce \\ bcf \\ bde \\ bdf \end{pmatrix},$$

so for example

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

In general, a system containing exactly $n \geq 2$ Qbits is represented by n copies of $\mathcal{C}^2$ tensored together. Therefore, the state space is 2^n dimensional.

A natural basis for this space consists of 2^n tensor products:

$$|0\rangle \otimes |0\rangle \otimes \ldots \otimes |0\rangle,$$

$$|0\rangle \otimes |0\rangle \otimes \ldots \otimes |1\rangle,$$

$$\vdots$$

$$|1\rangle \otimes |1\rangle \otimes \ldots \otimes |1\rangle.$$

A classical string of bits $i_1 i_2 \ldots i_n$ with $i_k \in \{0, 1\}$, $1 \leq k \leq n$, corresponds to the quantum state $|i_1\rangle \otimes |i_2\rangle \otimes \ldots \otimes |i_n\rangle$ which is simply denoted by $|i_1 i_2 \ldots i_n\rangle$.

The set

$$\{|i_1 i_2 \ldots i_n\rangle | i_k \in \{0, 1\}, 1 \leq k \leq n\}$$

is a basis in $\mathcal{C}^2 \otimes \mathcal{C}^2 \otimes \ldots \otimes \mathcal{C}^2$.

Note that a Cbit i is "mirrored" in the quantum system via the map $i \mapsto |i\rangle$.

An n Qbit system can exist in any superposition of the form

$$\Psi = \sum_{x=00\ldots0}^{11\ldots1} c_x |x\rangle, \tag{5}$$

where c_x are (complex) numbers such that $\sum_x |c_x|^2 = 1$. The exponential "explosion" represented by formula (5) distinguishes quantum systems from classical ones: *In a classical system we need 2^n states to represent n Cbits, but, quantum mechanically, we need only n states.*

In contrast with the classical physics, where the state of a system is completely defined by describing the state of each of its component pieces separately, in a quantum system the state cannot always be described considering only the component pieces.

The state

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

cannot be written as a tensor product of two single Qbits.

Indeed, let us assume for the sake of a contradiction, that there exist two Qbits $|x\rangle$ and $|y\rangle$ in $\mathcal{C}^2$ such that

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = |x\rangle \otimes |y\rangle.$$

Since each single Qbit is in a superposition of $|0\rangle$ and $|1\rangle$, there exist four complex numbers a_1, b_1, a_2, b_2 such that

$$|x\rangle = a_1|0\rangle + b_1|1\rangle \text{ and } |y\rangle = a_2|0\rangle + b_2|1\rangle.$$

It follows that

$$\begin{aligned} |x\rangle \otimes |y\rangle &= (a_1|0\rangle + b_1|1\rangle) \otimes (a_2|0\rangle + b_2|1\rangle) \\ &= a_1a_2|00\rangle + a_1b_2|01\rangle + b_1a_2|10\rangle + b_1b_2|11\rangle, \end{aligned}$$

hence $a_1b_2 = 0$ and $a_1a_2 = \frac{1}{\sqrt{2}} = b_1b_2$, which is impossible.

A state that cannot be expressed as a tensor product is called an **entangled state**. One can easily find entangled states in an n Qbit system, for any integer $n \geq 2$.

More about Evolution

According to the Evolution Postulate the quantum evolution (quantum transformation, operator) of (on) a Qbit is described by multiplication with a unitary matrix.

Considering the basis $\{|0\rangle, |1\rangle\}$, the transformation is fully specified by its effect on the basis vectors. In order to obtain the associated matrix of an operator $\tilde{U} : \mathcal{C}^2 \to \mathcal{C}^2$, we put the coordinates of $\tilde{U}|0\rangle$ in the first column and the coordinates of $\tilde{U}|1\rangle$ in the second one. So, the general form of a transformation that acts on a single Qbit is a 2×2 unitary matrix

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

which transforms the Qbit state

$$\alpha|0\rangle + \beta|1\rangle$$

into the state

$$(\alpha a + \beta b)|0\rangle + (c\alpha + d\beta)|1\rangle : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha a + \beta b \\ c\alpha + d\beta \end{pmatrix}.$$

Examples

For $\theta \in [0, 2\pi)$, the rotation R_θ is given by

$$R_\theta = \begin{pmatrix} cos\,\theta & -sin\,\theta \\ sin\,\theta & cos\,\theta \end{pmatrix}.$$

Hence, R_θ acts as follows:

$$|0\rangle \mapsto cos\,\theta|0\rangle + sin\,\theta|1\rangle,\ |1\rangle \mapsto -sin\,\theta|0\rangle + cos\,\theta|1\rangle.$$

One can easily verify that $R_\theta^\dagger R_\theta = I_2$, hence R_θ is unitary.

Note that in the special case $\theta = 0$ we get the identity transformation of $\mathcal{C}^2$:

$$R_0 = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The NOT transformation interchanges the vectors $|0\rangle$ and $|1\rangle$, is given by the matrix

$$\text{NOT} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

It flips that state of its input,

$$\text{NOT}\,|0\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |1\rangle,$$

and

$$\text{NOT}\,|1\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle.$$

The phase shift gate $Shift$ is defined by

$$Shift\,|0\rangle = |0\rangle, Shift\,|1\rangle = -|1\rangle,$$

so

$$Shift = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Since $\mathrm{NOT}^\dagger \cdot \mathrm{NOT} = I_2$ and $Shift^\dagger \cdot Shift = I_2$, the operators NOT and $Shift$ are also unitary.

The operator $Shift \cdot \mathrm{NOT}$ is also a unitary and we have:

$$Shift \cdot \mathrm{NOT}\ |0\rangle = Shift\,|1\rangle = -|1\rangle,$$
$$Shift \cdot \mathrm{NOT}\ |1\rangle = Shift\,|0\rangle = |0\rangle.$$

Therefore, its associated matrix is

$$R_{3\pi/2} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The *square-root of NOT* (introduced by Deutsch) is the transformation

$$\sqrt{\text{NOT}}\ : \begin{array}{rcl} |0\rangle & \rightarrow & \frac{1}{2}(1+i)|0\rangle + \frac{1}{2}(1-i)|1\rangle, \\ |1\rangle & \rightarrow & \frac{1}{2}(1-i)|0\rangle + \frac{1}{2}(1+i)|1\rangle, \end{array}$$

$$\sqrt{\text{NOT}} = \frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}.$$

$$\sqrt{\text{NOT}} \cdot \sqrt{\text{NOT}} = \text{NOT}\ , \tag{6}$$

and

$$\sqrt{\text{NOT}}^{\dagger} \cdot \sqrt{\text{NOT}} = \frac{1}{4}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}\begin{pmatrix} 1-i & 1+i \\ 1+i & 1-i \end{pmatrix} = I_2.$$

The square-root of NOT is a typical "quantum" gate in the sense that *it is impossible to have a single-input/single-output classical binary logic gate that satisfies* (6). Indeed, any classical binary

$$\sqrt{\text{NOT}}_{\text{classical}}$$

gate is going to output a 0 or a 1 for each possible input 0/1. Assume that we have such a classical square-root of NOT gate acting as a pair of transformations

$$\sqrt{\text{NOT}}_{\text{classical}}(0) = 1,\ \sqrt{\text{NOT}}_{\text{classical}}(1) = 0.$$

Then, two consecutive applications of it will *not* flip the input!

Figure 2: David Deutsch

Finally we consider the Hadamard transformation H is defined by

$$\begin{array}{rcl} |0\rangle & \rightarrow & \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |1\rangle & \rightarrow & \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{array},$$

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

This transformation has a number of important applications. When applied to $|0\rangle$, H creates a superposition state

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

Applied to n bits individually, H generates a superposition of all 2^n possible states.

More Examples

The Controlled-NOT Gate

A useful transformation on $\mathcal{C}^2 \otimes \mathcal{C}^2$ is the "controlled-NOT" gate, $C_{\rm NOT}$ defined as follows:

$$C_{\rm NOT}: \begin{array}{ccc} |00\rangle & \rightarrow & |00\rangle \\ |01\rangle & \rightarrow & |01\rangle \\ |10\rangle & \rightarrow & |11\rangle \\ |11\rangle & \rightarrow & |10\rangle \end{array},$$

$$C_{\rm NOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Given the input state $|ij\rangle$, $i, j \in \{0, 1\}$, the output state produced by C_{NOT} is $|ik\rangle$, where

$$k = i \oplus j(\ \mathrm{mod}\ 2) = \begin{cases} 0 & \text{if 2 divides } i + j, \\ 1 & \text{otherwise.} \end{cases}$$

The first bit is not disturbed (it is a control bit) and the second one interchanges 0 and 1 iff the first bit is 1, which corresponds to the logical exclusive-OR (XOR).

The controlled-NOT gate C_{NOT} can be represented by a circuit of the form specified in the following Figure.

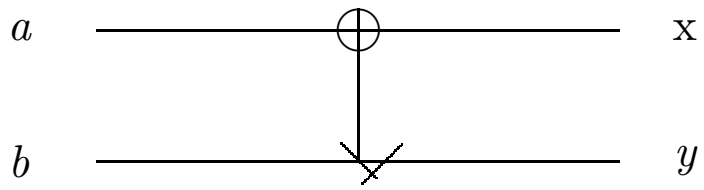


Figure 4: The controlled-NOT gate

The oplus indicates the control bit; the opposite symbol indicates the conditional negation of the second bit. If the input states at a and b are in bases states $|0\rangle$ or $|1\rangle$, then the output state at x is the same as the input state at a, and the output state at y is the exclusive-OR of the two input states.

The transformation C_{NOT} is unitary since

$$C_{\mathrm{NOT}}^{\dagger} C_{\mathrm{NOT}} = I_4$$

and

$$C_{\mathrm{NOT}}^{2} = I_4$$

(the 4×4 identity matrix).

More importantly,

> C_{NOT} *cannot be written as a tensor product of two operators.*

For the proof you assume the contrary, take two unitary matrices A, B and prove that the equality $C_{\mathrm{NOT}} = A \otimes B$ is impossible.

The Controlled-controlled-NOT Gate

The "controlled-controlled-NOT" transformation, CC_{NOT}, operates on three Qbits: it negates the rightmost bit iff the first two are both 1:

$$CC_{\mathrm{NOT}}: \begin{array}{llllll} |000\rangle & \rightarrow & |000\rangle, & |100\rangle & \rightarrow & |100\rangle \\ |001\rangle & \rightarrow & |001\rangle, & |101\rangle & \rightarrow & |101\rangle \\ |010\rangle & \rightarrow & |010\rangle, & |110\rangle & \rightarrow & |111\rangle \\ |011\rangle & \rightarrow & |011\rangle, & |111\rangle & \rightarrow & |110\rangle \end{array},$$

$$CC_{\mathrm{NOT}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

We have $CC_{\mathrm{NOT}}^{\dagger} \cdot CC_{\mathrm{NOT}} = I_8$, hence CC_{NOT} is also unitary.

Deutsch's Problem

The simplest way to illustrate the power of quantum computing is to solve the so-called *Deutsch's problem*. Consider a Boolean function $f : \{0, 1\} \to \{0, 1\}$ and suppose that we have a black box to compute it. We would like to know whether f is constant (that is, $f(0) = f(1)$) or balanced ($f(0) \neq f(1)$). To make this test classically, we need to compute $f(0)$ and $f(1)$ and to compare the results. Is it possible to do it better, i.e with *only one* computation of f? The answer is *affirmative*, and here is a possible solution.

We note that the problem can be stated in an equivalent way by asking to compute the value of $\mathrm{XOR}(f(0), f(1))$ with just one use of the black box computing f.

Suppose that we have a quantum black box to compute f. We will need to use the quantum states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ (which you recall form the basis of $\mathcal{C}^2$).

We assume that we have access only once to the black box computing f.

Consider the transformation U_f which applies to two Qbits, $|x\rangle$ and $|y\rangle$, and produces $|x\rangle|y \oplus f(x)\rangle$. This transformation flips the second Qbit if f acting on the first Qbit is 1, and does nothing if f acting on the first Qbit is 0.

The quantum evolution U_f can be presented equivalently in matrix form as:

$$U_f = \begin{pmatrix} (1-f(0),0) & (f(0),0) & (0,0) & (0,0) \\ (f(0),0) & (1-f(0),0) & (0,0) & (0,0) \\ (0,0) & (0,0) & (1-f(1),0) & (f(1),0) \\ (0,0) & (0,0) & (f(1),0) & (1-f(1),0) \end{pmatrix}.$$

Indeed, the first column of the matrix is obtained from

$$U_f|00\rangle = |0\rangle|0 \oplus f(0)\rangle = |0f(0)\rangle.$$

There are two cases: if $f(0) = 0$, then

$$|0f(0)\rangle = \begin{pmatrix} (1,0) \\ (0,0) \\ (0,0) \\ (0,0) \end{pmatrix} = \begin{pmatrix} (1-f(0),0) \\ (f(0),0) \\ (0,0) \\ (0,0) \end{pmatrix},$$

and if $f(0) = 1$, then

$$|0f(0)\rangle = \begin{pmatrix} (0,0) \\ (1,0) \\ (0,0) \\ (0,0) \end{pmatrix} = \begin{pmatrix} (1-f(0),0) \\ (f(0),0) \\ (0,0) \\ (0,0) \end{pmatrix}.$$

Whatever the values of $f(0)$ and $f(1)$, the matrix U_f is unitary, $U^\dagger U = I_4$, so according to the Evolution Postulate, U_f is a legitimate quantum black box.

Next we are going to use the Hadamard transformation H to generate a superposition of states:

$$H = \begin{pmatrix} (\frac{1}{2},0) & (\frac{1}{2},0) & (\frac{1}{2},0) & (\frac{1}{2},0) \\ (\frac{1}{2},0) & (-\frac{1}{2},0) & (\frac{1}{2},0) & (-\frac{1}{2},0) \\ (\frac{1}{2},0) & (\frac{1}{2},0) & (-\frac{1}{2},0) & (-\frac{1}{2},0) \\ (\frac{1}{2},0) & (-\frac{1}{2},0) & (-\frac{1}{2},0) & (\frac{1}{2},0) \end{pmatrix}.$$

We are now in a position to describe the quantum algorithm solving Deutsch's problem:

1. Start with a closed physical system prepared in the quantum state $|01\rangle$.
2. Evolve the system according to H.
3. Evolve the system according to U_f.
4. Evolve the system according to H.
5. Measure the system.

If $\mathrm{XOR}(f(0), f(1)) = 0$, then the quantum measurement yields the outcome 2; if $\mathrm{XOR}(f(0), f(1)) = 1$, then the quantum measurement yields the outcome 4, hence, computing $\mathrm{XOR}(f(0), f(1))$ with *only one* use of U_f.

To prove the correctness of the quantum algorithm, let us follow step-by-step its evolution.

In Step 1 we start with a closed physical system prepared in the quantum state $|01\rangle$:

$$V = \begin{pmatrix} (0,0) \\ (1,0) \\ (0,0) \\ (0,0) \end{pmatrix}.$$

After Step 2 the system has evolved in the state:

$$HV = \begin{pmatrix} (\frac{1}{2},0) & (\frac{1}{2},0) & (\frac{1}{2},0) & (\frac{1}{2},0) \\ (\frac{1}{2},0) & (-\frac{1}{2},0) & (\frac{1}{2},0) & (-\frac{1}{2},0) \\ (\frac{1}{2},0) & (\frac{1}{2},0) & (-\frac{1}{2},0) & (-\frac{1}{2},0) \\ (\frac{1}{2},0) & (-\frac{1}{2},0) & (-\frac{1}{2},0) & (\frac{1}{2},0) \end{pmatrix} \cdot \begin{pmatrix} (0,0) \\ (1,0) \\ (0,0) \\ (0,0) \end{pmatrix}$$

$$= \begin{pmatrix} (\frac{1}{2},0) \\ (-\frac{1}{2},0) \\ (\frac{1}{2},0) \\ (-\frac{1}{2},0) \end{pmatrix}.$$

After Step 3 the quantum system is in the state

$$U_f HV = \begin{pmatrix} (1-f(0),0) & (f(0),0) & (0,0) & (0,0) \\ (f(0),0) & (1-f(0),0) & (0,0) & (0,0) \\ (0,0) & (0,0) & (1-f(1),0) & (f(1),0) \\ (0,0) & (0,0) & (f(1),0) & (1-f(1),0) \end{pmatrix}$$

$$\cdot \begin{pmatrix} (\frac{1}{2},0) \\ (-\frac{1}{2},0) \\ (\frac{1}{2},0) \\ (-\frac{1}{2},0) \end{pmatrix} = \begin{pmatrix} (\frac{1}{2}-f(0),0) \\ (-\frac{1}{2}+f(0),0) \\ (\frac{1}{2}-f(1),0) \\ (-\frac{1}{2}+f(1),0) \end{pmatrix}.$$

Note that at this stage the state of the system *depends* upon f!

After Step 4, the quantum state of the system has become:

$$HU_f HV = \begin{pmatrix} (\frac{1}{2},0) & (\frac{1}{2},0) & (\frac{1}{2},0) & (\frac{1}{2},0) \\ (\frac{1}{2},0) & (-\frac{1}{2},0) & (\frac{1}{2},0) & (-\frac{1}{2},0) \\ (\frac{1}{2},0) & (\frac{1}{2},0) & (-\frac{1}{2},0) & (-\frac{1}{2},0) \\ (\frac{1}{2},0) & (-\frac{1}{2},0) & (-\frac{1}{2},0) & (\frac{1}{2},0) \end{pmatrix}$$

$$\cdot \begin{pmatrix} (\frac{1}{2}-f(0),0) \\ (-\frac{1}{2}+f(0),0) \\ (\frac{1}{2}-f(1),0) \\ (-\frac{1}{2}+f(1),0) \end{pmatrix} = \begin{pmatrix} (0,0) \\ (1-f(0)-f(1),0) \\ (0,0) \\ (f(1)-f(0),0) \end{pmatrix}.$$

Finally, in Step 5 we *measure* the current state of the system, that is we measure the state HU_fHV, and according to Born's Measurement Postulate we get:

1. outcome 1 with probability $p_1 = 0$,
2. outcome 2 with probability $p_2 = (1 - f(0) - f(1))^2$,
3. outcome 3 with probability $p_3 = 0$,
4. outcome 4 with probability $p_4 = (f(1) - f(0))^2$.

To conclude:

- if $\mathrm{XOR}(f(0), f(1)) = 0$, then $f(0) = f(1)$, so $f(0) + f(1) = 0 (\mathrm{mod}\ 2)$, $f(1) - f(0) = 0$; consequently, $p_2 = 1, p_4 = 0$.
- if $\mathrm{XOR}(f(0), f(1)) = 1$, then $f(0) \neq f(1)$, so $f(0) + f(1) = 1$, $f(1) - f(0) = -1$ or $f(1) - f(0) = 1$; consequently, $p_2 = 0, p_4 = 1$.

A compact mathematically formulation of the above quantum algorithm is the following: Start with U_f and evolve it on a superposition of $|0\rangle$ and $|1\rangle$. Assume first that the second Qbit is initially prepared in the state $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. Then,

$$\begin{aligned} U_f\left(|x\rangle\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\right) &= |x\rangle\frac{1}{\sqrt{2}}(|0 \oplus f(x)\rangle - |1 \oplus f(x)\rangle) \\ &= (-1)^{f(x)}|x\rangle\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \end{aligned}$$

Next take the first Qbit to be $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The black box will produce

$$\begin{aligned} &U_f\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\right) \\ &\quad = \frac{1}{\sqrt{2}}((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle)\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \\ &\quad = \frac{1}{2}(-1)^{f(0)}(|0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle)(|0\rangle - |1\rangle). \end{aligned}$$

Next will perform a measurement that projects the first Qbit onto the basis

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

We will obtain $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ if the function f is balanced and $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ in the opposite case.

So, Deutsch's problem was solved with only one computation of f. The explanation consists in the ability of a quantum computer to be in a blend of states: we can compute $f(0)$ and $f(1)$, but also, and more importantly, we can extract some information about f which tells us whether $f(0)$ is equal or not to $f(1)$.

We finish this issue with the following question: Can any function $f : \{0, 1\} \to \{0, 1\}$ be implemented by a quantum gate array U_f?

The answer is affirmative. Identifying the values 0 and 1 with the kets $|0\rangle$ respectively $|1\rangle$, U_f may be defined as the linear operator $U_f : \mathcal{C}^4 \to \mathcal{C}^4$, which satisfies, for any $x, y \in \{0, 1\}$, the equality

$$U_f|x, y\rangle = |x, y \oplus f(x)\rangle. \tag{7}$$

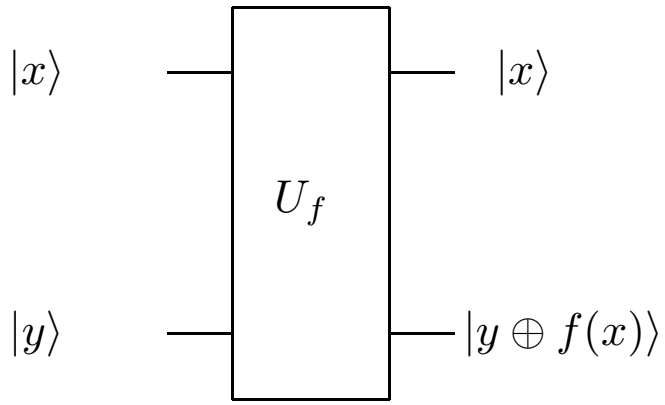


Figure 3: Quantum gate array U_f.

To compute $f(x)$ we apply U_f to $|x0\rangle$. We shall prove that

> *for any function* $f : \{0, 1\} \to \{0, 1\}$, U_f *is a unitary transformation.*

We have

$$U_f U_f|x, y\rangle = U_f|x, y \oplus f(x)\rangle = |x, (y \oplus f(x)) \oplus f(x)\rangle = |x, y\rangle,$$

hence, in view of the equality $U_f U_f = I_2$, it suffices to prove that $U_f^\dagger = U_f$.

The function f can be defined in four ways:

1. $f(0) = f(1) = 0$,
2. $f(0) = 0$, $f(1) = 1$,
3. $f(0) = 1$, $f(1) = 0$, and
4. $f(0) = f(1) = 1$.

We will investigate the matrix U_f in each situation, taking into account the correspondences:

$$0 \to |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; 1 \to |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In the first case, we have $U_f|x, y\rangle = |x, y \oplus 0\rangle = |x, y\rangle$, hence $U_f = I_2 = U_f^\dagger$.

In the second case, $U_f|00\rangle = |00\rangle$, $U_f|01\rangle = |01\rangle$, $U_f|10\rangle = |11\rangle$, $U_f|11\rangle = |10\rangle$, so it follows that

$$U_f = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = U_f^\dagger.$$

A direct computation shows that in the third case, $U_f|00\rangle = |01\rangle$, $U_f|01\rangle = |00\rangle$, $U_f|10\rangle = |10\rangle$ and $U_f|11\rangle = |11\rangle$, therefore,

$$U_f = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = U_f^\dagger.$$

Finally, $U_f|x, y\rangle = |x, y \oplus 1\rangle$, i.e., $U_f|x0\rangle = |x1\rangle$ and $U_f|x1\rangle = |x0\rangle$, hence we have again

$$U_f = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = U_f^\dagger.$$

Quantum states cannot be cloned, as Wooters and Zurek, and Dieks have proved as an application of the linearity of unitary transformations. It is not possible to create the state $(a|0\rangle + b|1\rangle) \otimes (a|0\rangle + b|1\rangle)$ from an *unknown state* $a|0\rangle + b|1\rangle$.

In other words, there is no unitary transformation U such that $U|\varphi 0\rangle = |\varphi\varphi\rangle$ for all quantum states $|\varphi\rangle$.

Indeed, assume the contrary and let $|\varphi\rangle$ and $|\psi\rangle$ be two orthogonal vectors in $\mathbf{C}^2$ and take $|x\rangle = \frac{1}{\sqrt{2}}(|\varphi\rangle + |\psi\rangle)$. Then, $U|\varphi 0\rangle = |\varphi\varphi\rangle$ and $U|\psi 0\rangle = |\psi\psi\rangle$. On the one hand,

$$\begin{aligned} U|x0\rangle &= |xx\rangle \\ &= \frac{1}{\sqrt{2}}(|\varphi\rangle + |\psi\rangle) \otimes \frac{1}{\sqrt{2}}(|\varphi\rangle + |\psi\rangle) \\ &= \frac{1}{2}(|\varphi\varphi\rangle + |\varphi\psi\rangle + |\psi\varphi\rangle + |\psi\psi\rangle). \end{aligned}$$

On the other hand,

$$\begin{aligned} U|x0\rangle &= U(\frac{1}{\sqrt{2}}(|\varphi 0\rangle + |\psi 0\rangle)) \\ &= \frac{1}{\sqrt{2}}(U|\varphi 0\rangle + U|\psi 0\rangle) \\ &= \frac{1}{\sqrt{2}}(|\varphi\varphi\rangle + |\psi\psi\rangle). \end{aligned}$$

Since the vectors φ and ψ are orthogonal, the vectors $|\varphi\varphi\rangle$, $|\varphi\psi\rangle$, $|\psi\varphi\rangle$, $|\psi\psi\rangle$ constitute a basis in $\mathbf{C}^2 \otimes \mathbf{C}^2$ and the vector $|xx\rangle = U|x0\rangle$ has been written in two different ways as a linear combination of this basis vectors, an impossibility.

The no cloning principle states the impossibility of reliably cloning an *unkown* quantum state: it is possible to clone a *known* quantum state. It is possible to obtain n particles in an entangled state $a|00\ldots0\rangle + b|11\ldots\rangle$ from an unknown state $a|0\rangle + b|1\rangle$. Each particles will behave in exactly the same way when measured with respect to the basis $\{|00\ldots0\rangle, |00\ldots01\rangle, \ldots |11\ldots1\rangle\}$, but *not* when measured with respect to other bases. It is not possible to create the n particle state

$$(a|0\rangle + b|1\rangle) \otimes (a|0\rangle + b|1\rangle) \otimes \ldots (a|0\rangle + b|1\rangle)$$

from an unkown state $a|0\rangle + b|1\rangle$.

In a sense, the no cloning principle seems to announce "bad news": we loose one of the most important facilities of classical computation, the unlimited possibility to copy. There are "good news" derived from this principle, for example, the possibility of unconditional secure key generation (see Section 6.2 in Gruska's book). New techniques open possibilites to produce "approximate" copies of qubits: imperfect, but very close to real copies of qubits can be produced with a "quality" no depending upon the qubits to be copied. Of course, there is price to be paid: copies produced in this way are entangled.

The measurement of one or more particles in a quantum system results in a projection of the state of the system prior to measurement onto the subspace of the state space compatible with the measured values. The amplitude of the projection is rescaled to make sure that the resulting state vector has length one. The probability that the result of the measurement is a given value is the sum of the squares of the absolute values of the amplitudes of all components compatible with that value of the measurement.

A simple example of measurement in a two qubit system will illustrate the above points. Let's fix the basis $\{|0\rangle, |1\rangle\}$, and assume that all measurements of individual qubits will be done with respect to this basis. An arbitrary state of a two qubit system can be written as

$$a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle,$$

where a, b, c and d are complex numbers such that

$$|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1.$$

When the first qubit is measured, then the probability that the result is $|0\rangle$ is $|a|^2 + |b|^2$.

Assume now that the measurement gives the first qubit exactly that value, that is, $|0\rangle$. Consequently, the state is projected onto the subspace compatible with the measurement which is the subspace spanned by $|00\rangle$ and $|01\rangle$ and the result of this projection is $a|00\rangle + b|01\rangle$. Renormalizing we get:

$$\frac{1}{\sqrt{|a|^2 + |b|^2}} \cdot (a|00\rangle + b|01\rangle).$$

In general, consider a system containing n qubits ($n \geq 2$). Any state $|x\rangle$ of the system can be expressed as

$$\sum_{i_1,i_2,\ldots,i_n=0,1} c_{i_1 i_2 \ldots i_n} |i_1 i_2 \ldots i_n\rangle,$$

where

$$\sum_{i_1,i_2,\ldots,i_n=0,1} |c_{i_1 i_2 \ldots i_n}|^2 = 1.$$

When the first qubit is measured with respect to the basis $\{|0\rangle, |1\rangle\}$, then the result $|0\rangle$ is obtained with probability

$$P = \sum_{i_2,\ldots i_n=0,1} |0 i_2 \ldots i_n\rangle|^2.^{\text{a}}$$

After rescaling, the new state obtained after the measurement is

$$\frac{1}{\sqrt{\sum_{i_2,\ldots,i_n=0,1} |c_{0 i_2 \ldots i_n}|^2}} \cdot \left(\sum_{i_2,\ldots,i_n=0,1} c_{0 i_2 \ldots i_n} |0 i_2 \ldots i_n\rangle \right).$$

[a] We used the projection onto the space spanned by $\{|0 i_2 \ldots i_n\rangle | i_k \in \{0,1\}, 2 \leq k \leq n\}$.

Similarly, the measurement gives the outcome $|1\rangle$ with the probability

$$1 - P = \sum_{i_2,\ldots,i_n=0,1} |c_{1i_2\ldots i_n}|^2,$$

and the state changes correspondingly.

What is the price of measurement? According to Landauer

> *If it [measurement] is simply information transfer, that is done all the time inside the computer, and can be done with arbitrarily little dissipation.*

There are many speculations about the "collapse of the wave function (state)" due to an irreversible interaction of the microphysical quantum system with the macroscopic measurement apparatus. Some authors (Greenberg and YaSin or Herzog, Kwiat Weinfuter and Zeilinger) have argued that it is, in fact, possible to reconstruct the state of the physical system before the measurement, that is, to "reverse the collapse of the wave function" if the process of measurement is reversible. After "reconstruction" no information about the measurement is left.

The act of measurement gives another perspective about entangled particles. Particles are not entangled if the measurement of one has no effect on the other. For instance, the state

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

is entangled since the probability that the first bit is measured to be $|0\rangle$ is $1/2$ *if* the second bit has not been measured. However, if the second bit *had been* measured, then the probability that the first bit is measured as $|0\rangle$ is different from $1/2$, it is either 1 or 0, depending on whether the second bit was measured as $|0\rangle$ or $|1\rangle$, respectively. Hence, the probability of measuring the first bit has been changed by the measurement of the second bit.

In contrast, the state

$$\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle) = |0\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

is not entangled. Reason: any measure of the first qubit will produce the result $|0\rangle$ independently whether a measurement is performed or not on the second qubit, and the second qubit has probability $1/2$ to be measured to $|0\rangle$ regardless of whether the first qubit was measured or not.

In a sense, entangled states can be equivalently presented in mathematical terms (they cannot be represented as a tensor product of two states) or in physical terms (the measurement on one affects the other); however, the physical meaning is reacher than the mathematical formalism.

An important consequence of the existence of entangled states is the fact that if a quantum memory register exists in an entangled state, one can change the state of one part of the register simply by measuring another part of it. This is a unique feature of quantum physics[a] which has no parallel in classical physics. *Entanglement is one of the most important features which distinguishes Quantum from conventional Computing.*

[a] Which is crucial in many quantum algorithms, teleportation, information transmission, etc.

How to produce entangled quantum states?

One possibility is to create a source which, by construction, is such that the quantum states emerging already have the indistinguishability feature. For example, consider the decay of a spin-0 particle into two spin-1/2 particles under conservation of the internal angular momentum. The two spins of the emerging particles have to be opposite, so the emerging quantum state is

$$|\psi\rangle_{12} = \frac{1}{\sqrt{2}}(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2),$$

where $|\uparrow\rangle_1$ means particle 1 with spin up.

The above state is rotationally invariant, so the two spins are anti-parallel along whichever direction we choose to measure.

The EPR Conundrum and Bell's Theorem

According to the philosophical view called realism, *reality* exists and has definite properties irrespective of whether they are observed by some agent. Motivated by this view point, Einstein, Podolsky and Rosen suggested a classical argument to "show" that quantum mechanics is incomplete.

EPR assumed:
(a) the non-existence of action-at-a-distance,
(b) that some of the statistical predictions of quantum mechanics are correct, and
(c) a reasonable criterion defining the existence of "an element of physical reality".

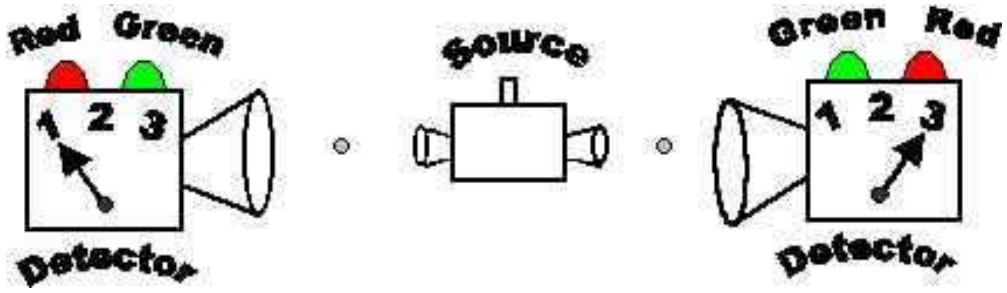
If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physica quantity.

They considered a system of two spatially separated but quantum mechanically correlated particles. A "mysterious" feature appears: By counterfactual reasoning, quantum mechanical experiments yield outcomes which cannot be predicted by quantum theory; hence the quantum mechanical description of the system is incomplete!

One possibility to complete the quantum mechanical description is to postulate additional "hidden-variables" in the hope that completeness, determinism and causality will be thus restored. But then, another conundrum occurs: Using basically the same postulates as those of EPR, Bell showed that no deterministic local hidden-variables theory can reproduce all statistical predictions of quantum mechanics.

Mermin's EPR Device

Mermin's EPR device has three "completely unconnected" parts (there are no relevant connections, neither mechanical nor electromagnetic), two detectors (D1) and (D2) and a source (S) emitting particles. The source is placed between the detectors: whenever a button is pushed on (S), shortly thereafter two particles emerge, moving off toward detectors (D1) and (D2). Each detector has a switch that can be set in one of three possible positions – labelled 1,2,3 – and a bulb that can flash a red (R) or a green (G) light. The purpose of lights is to "communicate" information to the observer. Each detector flashes either red or green whenever a particle reaches it. Because of the lack of any relevant connections between any parts of the device, the link between the emission of particles by (S), i.e. as a result of pressing a button, and the subsequent flashing of detectors, can only be provided by the passage of particles from (S) to (D1) and (D2). Additional tools can be used to check and confirm the lack of any communication.



Mermin's experiment

The device is repeatedly operated as follows:

1. the switch of either detector (D1) and (D2) is set randomly to 1 or 2 or 3, i.e. the settings or states 11, 12, 13, 21, 22, 23, 31, 32, 33 are equally likely,
2. pushing a button on (S) determines the emission toward both (D1) and (D2),
3. sometime later, (D1) and (D2) flash one of their lights, G or R,
4. every run is recorded in the form $ijXY$, meaning that (D1) was set to state i and flashed X and (D2) was set to j and flashed Y.

For example, the record $31GR$ means "(D1) was set to 3 and flashed G and (D2) was set to 1 and flashed R".

Long recorded runs show the following pattern:

(a) For records starting with ii, i.e. 11, 22, 33, both (D1) and (D2) flash the same colours, RR, GG, with equal frequency; RG and GR are never flashed.

(b) For records starting with $ij, i \neq j$, i.e. 12, 13, 21, 23, 31, 32, both (D1) and (D2) flash the same colour only 1/4 of the time (RR and GG come with equal frequencies); the other 3/4 of the time, they flash different colours (RG, GR), occurring again with equal frequencies.

Of course, the above patterns are statistical, that is they are subject to usual fluctuations expected in every statistical prediction: patterns are more and more "visible" as the number of runs becomes larger and larger.

The conundrum posed by the existence of Mermin's device reveals as soon as we notice that the seemingly simplest physical explanation of the pattern (a) is incompatible with pattern (b). Indeed, as (D1) and (D2) are unconnected there is no way for one detector to "know", at any time, the state of the other detector or which colour the other is flashing. Consequently, it seems plausible to assume that the colour flashed by detectors is determined only by some property, or group of properties of particles, say speed, size, shape, etc. What properties determine the colour does not really matter; only the fact that each particle carries a "program" which determines which colour a detector will flash in some state is important.

So, we are led to the following two hypotheses:

H1 *Particles are classified into eight categories:*

$$GGG, GGR, GRG, GRR, RGG, RGR, RRG, RRR.^{a}$$

H2 *Two particles produced in a given run carry identical programs.*

According to H1–H2, if particles produced in a run are of type RGR, then both detectors will flash R in states 1 and 3; they will flash G if both are in state 2. Detectors flash the same colours when being in the same states because *particles carry the same programs.*

[a] A particle of type XYZ will cause a detector in state 1 to flash X; a detector in state 2 will flash Y and a detector in state 3 will flash Z.

It is clear that from H1–H2 it follows that *programs carried by particles do not depend in any way on the specific states of detectors*: they are properties of particles not of detectors. Consequently, both particles carry the same program whether or not detectors (D1) and (D2) are in the same states. The emitting source (S) has no knowledge about the states of (D1) and (D2) and there is no communication among any parts of the device.

We are ready to argue that

[L] For each type of particle, *in runs of type* (b) *both detectors flash the same colour at least one third of the time.*

If both particles are of types GGG or RRR, then detectors will flash the same colour all the time. For particles carrying programs containing one colour appearing once and the other colour appearing twice, only in two cases out of six possible combinations both detectors will flash the same light. For example, for particles of type RGR, both detectors will flash R if (D1) is in state 1 and (D2) is in state 3 and vice versa. In all remaining cases detectors will flash different lights. The argument remains the same for all combinations as the conclusion was solely based on the fact that one colour appears once and the other twice. So, the lights are the same one third of the time.

The conundrum reveals as a significant difference appears between the data dictated by particle programs (colours agree at least one third of the time) and the quantum mechanical prediction (colours agree only one quarter of the time):

under H1–H2, *the observed pattern (b) is incompatible with* [L].

Mermin's GHZ Device

Mermin's GHZ device is based on Greenberg, Horne and Zeilinger's version of EPR experiment. The device has a source and three widely separated detectors (A), (B), (C), each of which has only two switch settings, 1 and 2. Any detector, when triggered, flashes red (R) or green (G). Again, detectors are supposed to be far away from the source and there are no connections between the source and detectors (except those induced by a group of particles flying from the source to each detector).

The experiment runs as follows. Each detector is in a randomly chosen state (1 or 2) and then by pressing a button at the source a trio of particles are released towards detectors; each particle will reach a detector and, consequently, each detector will flash a light, green or red.

There are eight possible states, but for the argument we need to take into consideration only those for which the number of 1's is odd, i.e. 111, 122, 212, 221.

(a) If one detector is set to 1 (and the others to 2), then an *odd* number of red lights always flash, i.e. RRR, RGG, GRG, GGR, and they are equally likely.

(b) If all detectors are set to 1, then an *odd* number of red lights is *never* flashed: GRR, RGR, RRG, GGG.

It is immediate that in case (a) knowing the colour flashed by two detectors, say (A) and (B), *determines uniquely* the colour flashed by the third detector, (C). The explanation can come only because particles are emitted by the same source (there are no connections between detectors). A similar conclusion as in the case of EPR device reveals: *particles carry programs instructing their detectors what colour to flash.* Any particle carries a program of the form XY telling its detector to flash colour X if in state 1 and colour Y if in state 2. There are four types of programs: GG, GR, RG, RR. A run in which programs carried by the trio of particles are of types (RG, GR, GG) will result in RRG if the states were 122, in GGG if the states were 212, and in GRG if the states were 221. This is an *illegal* set of programs as the number of R's is not odd (in RRG, for example).

A *legal* set of programs is (RG, GR, GR) as it produces RRR, GGR, GRG on 122, 212, 221. There are eight legal programs,

$$(RR, RR, RR), (RR, GG, GG), (GG, RR, GG),$$

$$(GG, GG, RR), (RG, GR, GR), (RG, RG, RG),$$

$$(GR, GR, RG), (GR, RG, GR)$$

out of 64 possible programs.

The conundrum reveals again as none of the above programs respects (b), i.e. it is compatible with the case 111. *A single* 111 *run suffices to prove inconsistency!*

Particle programs require an odd number of R*'s to be flashed on* 111*, but quantum mechanics prohibits this in every* 111 *run.*

Bell's Theorem

Bell showed, using basically the same postulates as those of EPR, that no deterministic local hidden-variables theory can reproduce all statistical predictions of quantum mechanics. The setting is the following. We consider two physical systems; on one two types of measurements are made (A, B), and on the other one two other types (C, D). The results are binary, so they will be denoted by "+" and "−". We will repeat these measurements to ensure statistically relevant results. *Correlations* appear when measurements give the same outcome, that is, "++" and "−−". The basic result is that in almost all cases, more "++" and "−−" (and less "+−" and "−−") coincidences are recorded than one can explain by any local classical analysis.

Let $p(x|i)$ be the probability that, by taking the measure $i \in \{A, B\}$ on the first system, the outcome will be $x \in \{+, -\}$; $p(x|ij)$ is the probability that by taking the measure $i \in \{A, B\}$ on the first system *and* the measure $j \in \{C, D\}$ on the second, the outcome of the first system *alone* will be x; $p(xy|ij)$ is the probability that by taking the measure i on the first system and measure j on the second system, the outcomes will be respectively, $x \in \{+, -\}$ and $y \in \{+, -\}$; finally, $p(x|ijy)$ is the probability that when taking the measures $i \in \{A, B\}$ on the first system and $j \in \{C, D\}$ on the second one, and having outcome y on the second, the outcome of the first will be x.

The main result can be stated as follows:

> *If the outcomes of the experiments on both systems are* independent, *that is*
>
> $$p(xy|ij) = p(x|i) \cdot p(y|j),$$
>
> *then the lack of correlation in one of the two types of measures cannot exceed the lack of correlation in the remaining types, that is, the following* quadrangular inequality *holds true:*

$$\begin{aligned} p(+-|AC) + p(-+|AC) & \\ \leq \quad & p(+-|AD) + p(-+|AD) \\ + \quad & p(+-|BD) + p(-+|BD) \\ + \quad & p(+-|BC) + p(-+|BC). \end{aligned} \tag{8}$$

It is remarkable that this inequality can be obtained with just an elementary manipulation of binary variables. To see this, let's denote $p(+|A)$ by a, $p(-|A)$ by $1-a$ (due to the bivalence nature of measurements we have $p(+|A)+p(-|A)=1$), and so on. Using the independence hypothesis, that is,

$$p(+-|AC) = p(+|A) \cdot p(-|C) = a(1-c),$$

and the like, the inequality (8) can be re-written as

$$\begin{aligned} a\,(1-c) + (1-a)\,c \quad &\leq \quad a\,(1-d) + (1-a)\,d + b\,(1-d) \\ &\quad + (1-b)\,d + b\,(1-c) + (1-b)\,c, \end{aligned}$$

or, equivalently,

$$ab + bd + bc \leq ac + b + d,$$

where $a, b, c, d \in [0, 1]$.

To finish we consider the following three cases:

- if $b \leq a$, then $c(b-a) \leq 0$, so $ad+bd+c(b-a) \leq ad+bd$, and (8) follows as $ad \leq d$ and $bd \leq b$;
- if $d \leq c$, then $a(d-c)+bd+bc \leq b+d$, so (8) follows;
- if $a \leq b$ and $c \leq d$, then either $b \leq d$ and in this case $d(a+b)+c(b-a) \leq b+d$, or $d \leq b$ and in this case $a(d-c)+b(d+c) \leq b+d$, and in each case we deduce (8).

The probabilistic hypothesis of independence can actually be decomposed in the conjunction of two hypotheses with more physical significance:

Separability: The statistical outcomes performed on one system are *independent of the outcomes* performed on the other system:

$$p(x|ijy) = p(x|ij) \text{ and } p(y|ijx) = p(y|ij).$$

Locality: The statistical outcomes performed an experiment on one system are *independent of the types of experiments* performed on the other system:

$$p(x|ij) = p(x|i) \text{ and } p(y|ij) = p(y|j).$$

Separability says that the spatio-temporal separation between the two systems makes them reducible to individual parts, the "whole" is no more than the "sum of parts"; locality forbids any instantaneous interaction.

Separability and locality implies independence as

$$p(xy|ij) = p(xy|ijy) \cdot p(y|ij) = p(x|ij) \cdot p(y|ij) = p(x|i) \cdot p(y|j).$$

Consequently, *if the outcomes of the experiments on both systems are* separable *and* local, *then the lack of correlation in one of the two types of measures cannot exceed the lack of correlation in the remaining types.*

Probabilities can be interpreted as truth-values of elementary propositions, so the above analysis can be reformulated in the language of "classical logic". Indeed, let's write A for $p(+|A)$ and $\neg A$ for $p(-|A)$, and similarly for B, C. Further on, let's notice that the elementary operations with probabilities can be reformulated as logical operations, namely, conjunction $\wedge$ will correspond to product, disjunction $\vee$ to sum, and implication $\rightarrow$ to $\leq$.

A "logical" version of the quadrangular inequality can be deduced:

If the conjunction is distributive with respect to disjunction for all propositions $A, \neg A, B, \neg B, C, \neg C$, *that is,*

$$\alpha \wedge (\beta \vee \gamma) \rightarrow (\alpha \wedge \beta) \vee (\alpha \wedge \gamma),$$

then the following quadrangular implication *holds true:*

$$\begin{aligned} (A \wedge \neg C) \vee (\neg A \wedge C) \quad &\rightarrow \quad (A \wedge \neg D) \vee (\neg A \wedge D) \\ &\vee \quad (D \wedge \neg B) \vee (\neg D \wedge B) \\ &\vee \quad (B \wedge \neg C) \vee (\neg B \wedge C). \end{aligned}$$

First, use the following weak form of distributivity

$$\alpha \wedge (\neg\beta \vee \beta) \rightarrow (\alpha \wedge \neg\beta) \vee (\alpha \wedge \beta),$$

for $\alpha = X \wedge \neg Y$, and $\beta = Z$:

$$(X \wedge \neg Y) \wedge (\neg Z \vee Z) \rightarrow (X \wedge \neg Y \wedge \neg Z) \vee (X \wedge \neg Y \wedge Z),$$

so by the law of excluded middle we get:

$$(X \wedge \neg Y) \rightarrow (X \wedge \neg Y \wedge \neg Z) \vee (X \wedge \neg Y \wedge Z).$$

Weakening the conclusion we get:

$$(X \wedge \neg Y) \rightarrow (X \wedge \neg Z) \vee (Z \wedge \neg Y). \tag{9}$$

Using (9) for the triples $(X, Y, Z) = (A, C, D), (D, C, B)$ we get

$$(A \wedge \neg C) \rightarrow (A \wedge \neg D) \vee (D \wedge \neg C),$$

and

$$(D \wedge \neg C) \rightarrow (D \wedge \neg B) \vee (B \wedge \neg C),$$

which imply

$$(A \wedge \neg C) \rightarrow (A \wedge \neg D) \vee (D \wedge \neg B) \vee (B \wedge \neg C).$$

Similarly, we obtain the implication

$$(\neg A \wedge C) \rightarrow (\neg A \wedge D) \vee (\neg D \wedge B) \vee (\neg B \wedge C),$$

which concludes the argument.

Both quadrangular inequality and implication have been experimentally falsified, hence no theory satisfying their hypotheses can be physically correct. So, *locality* and *separability* cannot be simultaneously adopted. Quantum mechanics has chosen to drop separability. The failure of independence affects Reichenbach's *causality* principle: two correlated (non independent) events have a common cause, that there exists an event in their "past" with respect to which they are independent.

So, we arrive at the idea of *synchronicity* that has important implication for Quantum Computation:

> *there exist events which are correlated in a way which is neither casual nor causal.*

Finally, the failure of *distributivity* – the "mark" of quantum logic, has been proved to be more pervasive than the universe of quantum mechanics statements: it is excluded from any logic aiming to describe the physical world. Is any hope to rescue classical logic, which seems to be so brutally excluded ...

A Probabilistic Automaton Simulating Mermin's EPR Device

The states of the automaton are all combinations of states of detectors (D1) and (D2), $Q = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$, the input alphabet models the lights, red and green, $\Sigma = \{G, R\}$, the output alphabet captures all combinations of lights flashed by (D1) and (D2), $O = \{GG, GR, RG, RR\}$, and the output function $f : Q \to O$, modeling all combinations of green/red lights flashed by (D1) and (D2) in all their possible states, is probabilistically defined by:

$$
\begin{aligned}
f(ii) &= XX, \text{ with probability } 1/2, \text{ for } i = 1, 2, 3, \\
&\quad X \in \{G, R\}, \\
f(ii) &= XY, \text{ with probability } 0, \text{ for } i = 1, 2, 3, \\
&\quad X, Y \in \{G, R\}, X \neq Y, \\
f(ij) &= XX, \text{ with probability } 1/8, \text{ for } i, j = 1, 2, 3, \\
&\quad i \neq j, X \in \{G, R\}, \\
f(ij) &= XY, \text{ with probability } 3/8, \text{ for } i, j = 1, 2, 3, \\
&\quad i \neq j, X, Y \in \{G, R\}, \\
&\quad X \neq Y.
\end{aligned}
$$

For example, $f(11) = RR$ with probability $1/2$, $f(11) = GR$ with probability 0, $f(11) = RG$ with probability 0, $f(11) = RR$ with probability $1/2$, $f(12) = GG$ with probability $1/8$, $f(12) = GR$ with probability $3/8$, $f(12) = RG$ with probability $3/8$, $f(12) = RR$ with probability $1/8$, etc.

The automaton transition $\delta : Q \times \Sigma \to Q$ is *not specified.* In fact, varying all transition functions δ we get a class of Mermin EPR automata:

$$\mathcal{M}$$
$$(EPR) = (Q, \Sigma, O, \delta, (p_{ij}^{XY}, i, j = 1, 2, 3, X, Y \in \{G, R\})),$$

where
$p_{ii}^{XX} = 1/2,\ p_{ii}^{XY} = 0,\ X \neq Y,\ p_{ij}^{XX} = 1/8,\ p_{ij}^{XY} = 3/8,\ X \neq Y.$

Are there two identical, spatially separated, probabilistic automata with identical initial states, whose direct product "simulates" a Mermin's EPR automaton? More formally, are there two probabilistic automata

$$\mathcal{M}_i = (\{1,2,3\}, \{G,R\}, \{G,R\}, \delta_i, (\alpha^X_{i,j}, j = 1,2,3, X \in \{G,R\}))$$

such that their direct product $\mathcal{M}_1 \otimes \mathcal{M}_2$ is isomorphic to a Mermin's automaton $\mathcal{M}$ (EPR), i.e.,

$\delta(ij, X) = \delta_1(i, X)\delta_2(j, X)$, and $p^{XY}_{ij} = \alpha^X_{1,i}\alpha^Y_{2,j}$, for all $j = 1,2,3, X, Y \in \{G,R\}$?

The answer is *negative*. In fact, a stronger result is true:

> *no single state of any Mermin's EPR probabilistic automaton* $\mathcal{M}$ (EPR) *can be simulated by the product of the corresponding states of any probabilistic automata* $\mathcal{M}_i$.

Indeed, $\alpha^G_{i,j} = 1 - \alpha^R_{i,j}$. For a state ii we get the following contradictory relations:

$$\alpha^G_{1,i}\ \alpha^G_{2,i} = (1 - \alpha^G_{1,i})(1 - \alpha^G_{2,i}) = 1/2,$$

$$\alpha^G_{1,i}\ (1 - \alpha^G_{2,i}) = (1 - \alpha^G_{1,i})\ \alpha^G_{2,i} = 0.$$

For a state kl with $k \neq l$ we, again, get two contradictory relations:

$$\alpha^G_{1,k}\ \alpha^G_{2,l} = (1 - \alpha^G_{1,k})(1 - \alpha^G_{2,l}) = 1/8,$$

$$\alpha^G_{1,k}\ (1 - \alpha^G_{2,l}) = (1 - \alpha^G_{1,k})\ \alpha^G_{2,l} = 3/8.$$

Every Mermin's EPR probabilistic automaton $\mathcal{M}$ (EPR) has strong correlations preventing it from being decomposed as a direct product of two independent probabilistic automata, no matter what transitions and output functions.

Let's turn our attention to Mermin's GHZ device and to this aim consider a probabilistic automaton simulating Mermin's GHZ device. The states of the Mermin's GHZ automaton are all combinations of states of detectors (A), (B) and (C),

$Q = \{111, 112, 121, 122, 211, 212, 221, 222\}$,

the input alphabet models the lights, red and green, $\Sigma = \{G, R\}$, the output alphabet captures all combinations of lights flashed by (A), (B) and (C),

$O = \{GGG, GGR, GRG, GRR, RGG, RGR, RRG, RRR\}$,

and the output function $f : Q \to O$, modeling all combinations of green/red lights flashed by (A), (B) and (C), is determined by the following conditions. (Note that the following conditions do not determine uniquely the output function.)

$$
\begin{aligned}
f(ijk) &= XYZ, \text{ with probability } 1/4, \text{ for} \\
&\quad ijk \in \{122, 212, 221\}, \\
&\quad XYZ \in \{RRR, RGG, GRG, GGR\}, \\
f(ijk) &= XYZ, \text{ with probability } 0, \text{ for} \\
&\quad ijk \in \{122, 212, 221\}, \\
&\quad XYZ \in \{GRR, RGR, RRG, GGG\}, \\
f(111) &= XYZ, \text{ with probability } 0, \text{ for} \\
&\quad XYZ \in \{RRR, RGG, GRG, GGR\}, \\
f(111) &= XYZ, \text{ with probability } 1/4, \text{ for} \\
&\quad XYZ \in \{GRR, RGR, RRG, GGG\}.
\end{aligned}
$$

Again, the transition function $\delta : Q \times \Sigma \rightarrow Q$ is not specified. We get a class of Mermin GHZ automata

$$\mathcal{M}\ (GHZ) = (Q, \Sigma, O, \delta, (p_{ijk}^{XYZ}, i, j, \\ k = 1, 2, X, Y, Z \in \{G, R\})),$$

where

$p_{ijk}^{XYZ} = 1/4$, for $ijk \in \{122, 212, 221\}$, $XYZ \in \{RRR, RGG, GRG, GGR\}$ or $i = j = k = 1$, $XYZ \in \{GRR, RGR, RRG, GGG\}$,

and

$p_{ijk}^{XYZ} = 0$, for $ijk \in \{122, 212, 221\}$, $XYZ \in \{GRR, RGR, RRG, GGG\}$ or $i = j = k = 1$, $XYZ \in \{RRR, RGG, GRG, GGR\}$.

Is there any Mermin's GHZ automaton which can be decomposed into three identical, spatially separated, probabilistic automata with identical initial values? Rephrased, are there three probabilistic automata

$$\mathcal{M}_i = (\{1,2\}, \{G,R\}, \{G,R\}, \delta_i, (\alpha^X_{i,j}, j = 1,2, X \in \{G,R\}))$$

such that their direct product $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ is isomorphic to a Mermin's automaton $\mathcal{M}$ (GHZ):

$\delta(ijk, XYZ) = \delta_1(i,X)\delta_2(j,Y)\delta_3(k,Z)$ and $p^{XYZ}_{ijk} = \alpha^X_{1,i}\alpha^Y_{2,j}\alpha^Z_{3,k}$, for all $j = 1,2,\ X,Y \in \{G,R\}$?

The answer is again *negative:*

> *no single state of any Mermin's GHZ probabilistic automaton* $\mathcal{M}$ (GHZ) *can be simulated by the product of the corresponding states of any probabilistic automata* $\mathcal{M}_i$.

We have $\alpha^G_{i,j} = 1 - \alpha^R_{i,j}$. Take the output $XYZ = GGR$. As $p^{GGR}_{111} = 0$ we deduce that

$$\alpha^G_{1,1}\alpha^G_{i2,1}(1 - \alpha^G_{3,1}) = 0,$$

which contradicts the system of equalities

$$p^{GGR}_{122} = p^{GGR}_{212} = p^{GGR}_{221} = 1/4,$$

and the same conclusion can be derived for any output.

Again, due to strong correlations, every Mermin's GHZ probabilistic automaton $\mathcal{M}$ (EPR) cannot be decomposed as a direct product of three independent probabilistic automata, no matter what transitions and output functions.

We continue with an analysis using Meal automata.

First we deal with Mermin EPR device. To this aim we discuss a configuration in which two identical deterministic Mealy automata. (recall that in a Mealy automaton the output function depends both on the current state and input letter.) $\mathcal{M}_1$ and $\mathcal{M}_2$ with unknown but identical initial states are detected in (D1) and (D2), respectively.

More precisely, let us assume that each automaton $\mathcal{M}_j$, $j = 1, 2$, has three states $Q = \{1, 2, 3\}$, the input alphabet $\Sigma = \{1, 2, 3\}$, the output alphabet $O = \{G, R\}$, as well as a(n) (irreversible, i.e., many-to-one) transition function $\delta_j(q, i) = i$ and output function $\lambda_j(q, i) = G$, if $q = i$ and $\lambda_j(q, i) = R$, otherwise; $q \in Q$ and $i \in \Sigma$. Let us further assume that there is an equidistribution of initial states, i.e., each one occurs with equal probability 1/3.

We can construct a joint output function by the Cartesian product $\lambda : Q \times \Sigma \to O \times O$, $\lambda(q,i) = (\lambda_1(q,i), \lambda_2(q,i))$.

Since both $\mathcal{M}_1$ and $\mathcal{M}_2$ are in an identical initial value, there are just three allowed categories GRR, RGR, RRG out of the conceivable ones $GGG, GGR, GRG, GRR, RGG, RGR, RRG, RRR$.

A straightforward combinatorial argument shows that with these assumptions one obtains the following probabilities:

$$
\begin{aligned}
\lambda(i,i) &= GG, \text{ with probability } 1/3, \text{ for } i = 1,2,3, \\
\lambda(i,i) &= RR, \text{ with probability } 2/3, \text{ for } i = 1,2,3, \\
\lambda(i,i) &= XY, \text{ with probability } 0, \text{ for } i = 1,2,3, \\
&\quad X, Y \in \{G,R\}, X \neq Y, \\
\lambda(i,j) &= GG, \text{ with probability } 0, \text{ for } i,j = 1,2,3, i \neq j, \\
\lambda(i,j) &= GR, \text{ with probability } 1/3, \text{ for } i,j = 1,2,3, \\
&\quad i \neq j, \\
\lambda(i,j) &= RG, \text{ with probability } 1/3, \text{ for } i,j = 1,2,3, \\
&\quad i \neq j, \\
\lambda(i,j) &= RR, \text{ with probability } 1/3, \text{ for } i,j = 1,2,3, \\
&\quad i \neq j.
\end{aligned}
$$

The automata flash the same colour (red) 1/3 of the time and different colours 2/3 of the time. This is not exactly the classical case as discussed by Mermin, but it comes close to it in terms of classicality and locality of the automata arrangement. To understand why, let us define the notion of *correlation function* in the automaton context. Assume again two output symbols, say R and G, and three input symbols, say $1, 2$ and 3.

Associate the numbers $n_t(i, \mathcal{M}_j) = +1$ and $n_t(i, \mathcal{M}_j) = -1$ with the outcomes R and G of the experiment with input i at discrete time t, respectively. In analogy to physical correlation functions we can define a correlation function C as the weighted average over the product of the numbers associated with the outcomes of the first and second automata $\mathcal{M}_1, \mathcal{M}_2$, i.e.,

$$C(i,j) = \tfrac{1}{N} \textstyle\sum_{t=1}^{N} n_t(i, \mathcal{M}_1) \cdot n_t(j, \mathcal{M}_2).$$

We always get $-1 \leq C(i,j) \leq +1$. In the above case, for identical inputs, $C(i,i) = 1$, $i = 1,2,3$. For nonidentical input $i \neq j$, $C(ij) = -1/3$. The "Bell inequality" is considered a measure for classicality and locality; in particular

$$|C(1,2) - C(1,3)| \leq 1 + C(2,3). \tag{10}$$

is always satisfied for classical systems. The automaton correlation functions always satisfy this inequality and the others obtained by permuting the inputs. This is an indication (although no sufficient condition) that the corresponding classical system behaves locally in the sense used in physics. That is, no causal influence such as a light signal originating from a measurement on one particle can influence the measurement on the other particle and vice versa. This comes as no surprise, because the way the two-automaton setup was conceived, both automata are causally separated in a classical sense.

These results are independent of the particular transition function δ involved, provided it is not a permutation (one-to-one).

An automaton realization which comes close to Mermin's treatment of the GHZ experiment can be given by three identical automata $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ with identical initial value, given by the following table ($q \in Q$, $i \in \Sigma$, $o \in O$):

$q/i, o$	1	2	1	2
1	1	1	R	R
2	1	1	R	G

Here, in configurations like 122, there always occurs an odd number of R's, whereas for 111, only a single result RRR emerges, which has an odd number of R's and is distinct from the quantum mechanical result containing an even number of R's.

Again, the argument is independent of the transition function as long as it is not a permutation.

Quantum teleportation

It is possible to transmit qubits without sending qubits!

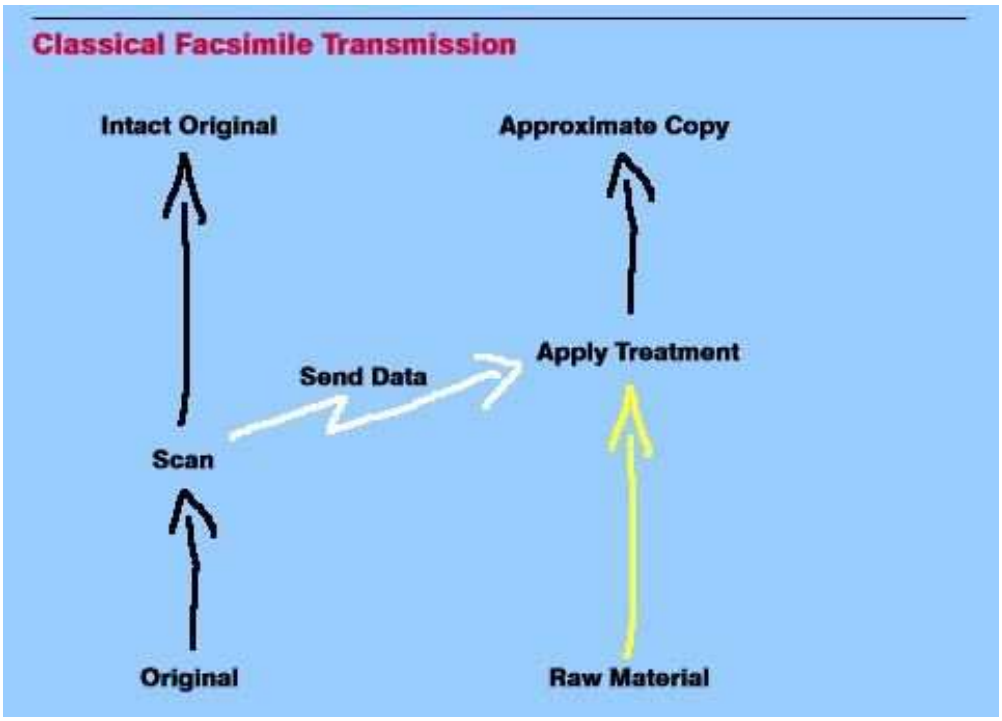
What does this mean? It's pun? According to Bennett[a]

"It's a mean by which you can take apart an unknown quantum state into classical information and purely quantum information, send them through two separate channels, put them back together, and get back the original quantum state".

Teleportation, as it is commonly understood, is a fictional procedure of transferring an object from one location to another location in a three stage process: a) dissociation, b) information transmission, c) reconstitution. The point is that, in contrast with fax transmission–where the original object remains intact at the initial location, only an approximate replica is constructed at destination,[b] in teleportation the original object is destroyed after enough information about it has been extracted, the object is not traversing in any way the space between locations, but it is reconstructed, as an exact replica, at the destination.

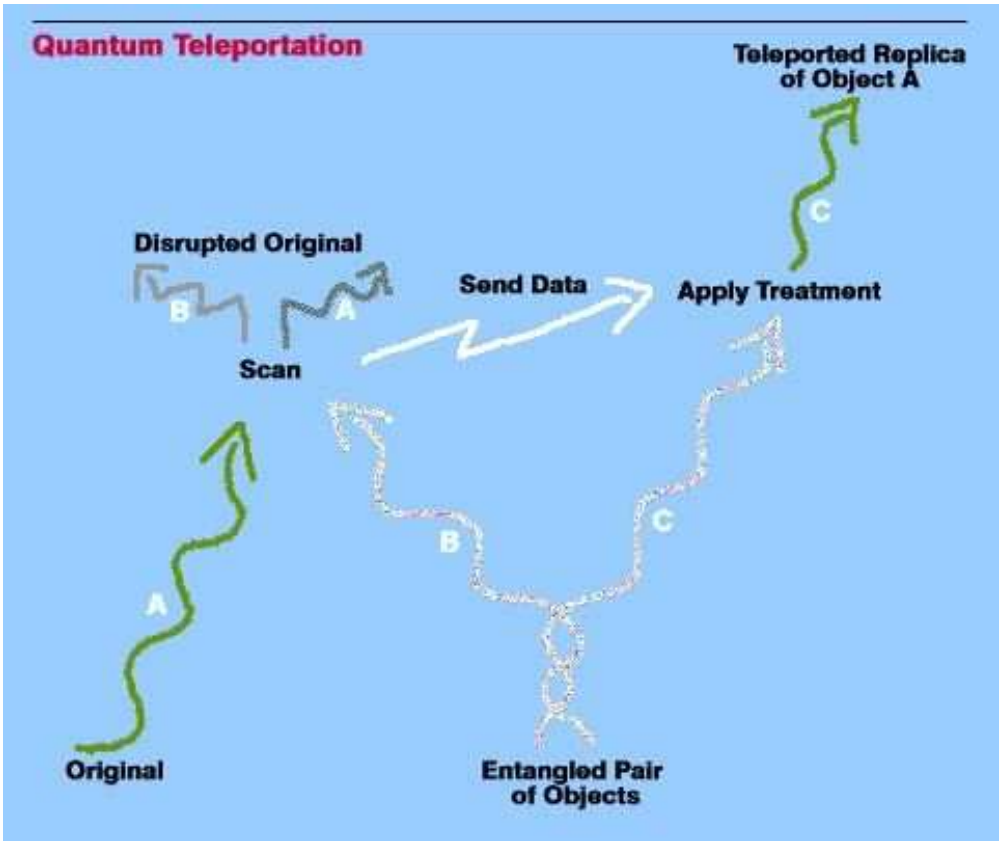
[a] A co-author of a 1993 paper that proposed quantum teleportation.

[b] At the end, two "identical" versions of the original object result.



Fax transmission

Quantum teleportation allows for the transmission of quantum information to a distant location. The objective is to transmit the quantum state of a particle using classical bits and reconstruct the state at the receiver.



Quantum teleportation

Locality

Locality interaction is

- mediated by another entity (particle, field),
- propagates no faster than light,
- its strength drops off with distance.

All known forces in the universe (electromagnetic, gravitational, strong/weak nuclear) are *local*. So, what's left? The *collapse of the state vector*. Nothing explains, mediates or determines the exact mechanism of the collapse. In particular, the collapse involves *no forces* of any kind.

Let's assume that Alice wishes to communicate with Bob a single qubit in an unknown state $\varphi = a|0\rangle + b|1\rangle$; she wants to make the transmission through classical channels. Alice cannot know with certainty the state as any measurement she may perform may change it; she cannot clone it because of the no cloning result! So, it seems that the only way to send Bob the qubit is to send him the *physical qubit*, or to swap the state into another quantum system and then send Bob that system.

Alice and Bob use an entangled pair

$$\psi_0 = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Alice controls the first half of the pair and Bob controls the second one. The input state is

$$\begin{aligned}
\varphi \otimes \psi_0 &= (a|0\rangle + b|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
&= \frac{1}{\sqrt{2}}(a|0\rangle \otimes |00\rangle + a|0\rangle \\
&\quad \otimes |11\rangle + b|1\rangle \otimes |00\rangle + b|1\rangle \otimes |11\rangle) \\
&= \frac{1}{\sqrt{2}}(a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle).
\end{aligned}$$

Alice now applies the transformation $(H \otimes I \otimes I) \circ (C_{not} \otimes I)$ to this state. The third bit is left unchanged; only the first two bits belong to Alice and the rightmost one belongs to Bob.

Applying now $H \otimes I \otimes I$, we have:

$$\begin{aligned}
& (H \otimes I \otimes I) \circ (C_{not} \otimes I)(\varphi \otimes \psi_0) \\
= \quad & \frac{1}{\sqrt{2}} H \otimes I \otimes I(a|000\rangle + a|011\rangle \\
& + b|110\rangle + b|101\rangle) \\
= \quad & \frac{1}{\sqrt{2}}(aH|0\rangle \otimes (I \otimes I)|00\rangle \\
& + aH|0\rangle \otimes (I \otimes I)|11\rangle + \\
& + bH|1\rangle \otimes (I \otimes I)|10\rangle + bH|1\rangle \\
& \otimes (I \otimes I)|01\rangle) \\
= \quad & \frac{1}{\sqrt{2}}(a\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |00\rangle + a\frac{1}{\sqrt{2}}(|0\rangle \\
& + |1\rangle) \otimes |11\rangle + b\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \otimes |10\rangle \\
& + b\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \otimes |01\rangle) \\
= \quad & \frac{1}{2}(a(|000\rangle + |100\rangle + |011\rangle + |111\rangle) + b(|010\rangle \\
& - |110\rangle + |001\rangle - |101\rangle)).
\end{aligned}$$

This state may be re-written by regrouping terms:

$$\begin{aligned}
& (H \otimes I \otimes I) \circ (C_{not} \otimes I)(\varphi \otimes \psi_0) \\
= \quad & \frac{1}{2}(|00\rangle(a|0\rangle + b|1\rangle) \\
& + |01\rangle(a|1\rangle + b|0\rangle) + |10\rangle(a|0\rangle - b|1\rangle) \\
& + |11\rangle(a|1\rangle - b|0\rangle)).
\end{aligned}$$

Alice then measures her two qubits, obtaining four possible results: $|00\rangle$, $|01\rangle$, $|10\rangle$, or $|11\rangle$ with equal probability 1/4. Depending on the result of the measurement, the quantum state of Bob's qubit is projected to $a|0\rangle + b|1\rangle$, $a|1\rangle + b|0\rangle$, $a|0\rangle - b|1\rangle$, $a|1\rangle - b|0\rangle$, respectively. Alice sends the result of her measurement as two classical bits to Bob. He will know what has happened, and can apply the decoding transformation $T \in \{I, X, Y, Z\}$ to fix his qubit.

Received bits	State	Transformation	Result Result
00	$a\|0\rangle + b\|1\rangle$	I	$a\|0\rangle + b\|1\rangle$
01	$a\|1\rangle + b\|0\rangle$	X	$a\|0\rangle + b\|1\rangle$
10	$a\|0\rangle - b\|1\rangle$	Z	$a\|0\rangle + b\|1\rangle$
11	$a\|1\rangle - b\|0\rangle$	Y	$a\|0\rangle + b\|1\rangle$

The final output state is $\varphi = a|0\rangle + b|1\rangle$, which, as desired, is the unknown qubit that Alice wanted to send.

1. The above scheme teleports the "quantum state" not the object.
2. We cannot use the scheme for teleporting an electron, for example; rather we can teleport the "spin" orientation of one electron.
3. The scheme is limited by the classical component.
4. According to S. Braunstein, the current technology would need 100 million centuries to transmit a human body (described down to atomic structure) via a single channel!
5. So, *why teleport a quantum state?* One reason is that this type of communication may be used inside a quantum computer or between quantum computers.

Recently, important teleportation experiments have been performed in Vienna (A. Zeilinger), Rome (F. De Martini) and Caltech (J. Kimble). There is a lot of controversy about the nature of quantum teleportation and what criteria should be met by a successful experiment. The following criteria for evaluating a quantum teleportation procedure have been proposed:

- How well can it teleport any arbitrary quantum state it is intended to teleport? (fidelity of teleportation)
- How often does it succeed to teleport, when it is given an input state within the set of states it is designed to teleport? (efficiency of teleportation)
- If given a state the scheme is not intended to teleport, how well does it reject such a state? (cross-talk rejection efficiency)

Let us close this discussion with another controversial statement of the same Bennett:

"I think it's quite clear that anything approximating teleportation of complex living beings, even bacteria, is so far away technologically that it's not really worth thinking about it."

Quantum Cryptography

While classical cryptography employs various mathematical techniques to restrict eavesdroppers from learning the contents of encrypted messages, in quantum mechanics the information is protected by the laws of physics. The Heisenberg uncertainty principle and quantum entanglement can be exploited in a system of secure communication, often referred to as "quantum cryptography".

There are three main types of quantum cryptosystems:

- Cryptosystems with encoding based on two non-commuting observables proposed by Wiesner (1970), and by Bennett and Brassard (1984),
- Cryptosystems with encoding built upon quantum entanglement and the Bell Theorem proposed by Ekert (1990),
- Cryptosystems with encoding based on two non-orthogonal state vectors proposed by Bennett (1992).

The system includes a transmitter and a receiver.

A sender may use the transmitter to send photons in one of four polarisations: 0, 45, 90, or 135 degrees.

A recipient at the other end uses the receiver to measure the polarisation.

According to the laws of quantum mechanics, the receiver can distinguish between rectilinear polarisations (0 and 90), or it can quickly be reconfigured to discriminate between diagonal polarisations (45 and 135); *it can never, however, distinguish both types.*

The key distribution requires several steps:

- The sender sends photons with one of the four polarisations which are chosen at random.
- For each incoming photon, the receiver chooses at random the type of measurement: either the rectilinear type or the diagonal type.
- The receiver records the results of the measurements but keeps them secret.
- Subsequently the receiver publicly announces the type of measurement (but not the results).
- The sender tells the receiver which measurements were of the correct type.
- The two parties (the sender and the receiver) keep all cases in which the receiver measurements were of the correct type. These cases are then translated into bits (1's and 0's) and thereby become the key.

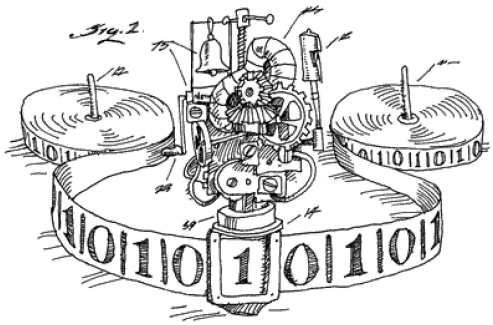
An eavesdropper is bound to introduce errors to this transmission because she does not know in advance the type of polarisation of each photon and quantum mechanics does not allow her to acquire sharp values of two non-commuting observables (here rectilinear and diagonal polarisations).

The two legitimate users of the quantum channel test for eavesdropping by revealing a random subset of the key bits and checking (in public) the error rate. Although they cannot prevent eavesdropping, they will never be fooled by an eavesdropper because any, however subtle and sophisticated, effort to tap the channel will be detected. Finally, they can try to set up the key distribution again.

For the first time in 1992 a photon polarization measurement scheme has been used to make a working quantum key distribution system in a laboratory at the IBM Thomas J. Watson Research Center, which transmits over the admittedly modest length of 30 cm at a rate of 10 bits/second.

Is it Possible to Break Turing's Barrier?

Turing's halting problem (THP), i.e. the problem to decide whether an arbitrary Turing machine (TM)



halts on an arbitrary input, is arguably the most (in)famous unsolvable (by any TM) mathematical problem.

The essence of the proof is the impossibility in answering in a *finite* time the *infinite* set of questions

"does $\mathtt{T}(x)$ stop in t steps", for $t = 1, 2, \ldots$

It is essential that `HALT` *is a* TM, *so* Q *is itself a* TM.

So, what about trying to prove that `HALT` is "computable by some other type of machine"?

Three natural ideas come to mind:

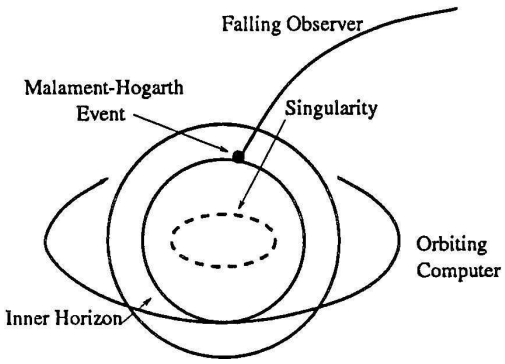
- relativistic machines
- quantum machines
- biological schemes

Relativistic Computing

Accelerate machines, which execute their n-th instruction in 2^{-n} seconds, may not be impossible as the same physical theory which limits the speed of information processing by the velocity of light maintains that time is relative to the observer.

So, if a satellite revolving with instantaneous tangential velocity $c(1-e^{-2t})^{\frac{1}{2}}$ (c is the speed of light, t is the earth time scale) and local time scale T, the time interval $dT = e^{-t}dt$, then one second in the satellite's time scale corresponds to an eternity on earth as $\int_0^\infty e^{-t}dt = 1$.

In 2002 Etesi and Németi observed that in Malament-Hogarth space-times, due to infinite time contraction, it may be possible for a computer to receive, in *finite* time, the answer to a *yes-or-no question* from an *infinite* computation.



A Quantum Strategy

In 2002 Calude and Pavlov have proposed the following attack on the infinite Merchant Problem.

We are given $\theta = 2^{-n}$ and we assume that we work with a quantum "device" with sensitivity $\varepsilon = 2^{-m}$.

- First, we *compute classically* a time $T = T_{\theta,\varepsilon}$,
- Then, we run the "device" on a random input for the time T.

The quantum "device" may or may not produce a click.

- If we get a click, then the system has false coins (in the finite case the stack containing false coins can be located).
- If we don't get a click, then with probability greater than $1 - \theta$ all coins are true.

An essential part of the method is the requirement that the time limit T is *computable* in a classical way.

A Biological Strategy

In 2003 Calude and Păun have proposed two biological hypotheses which have as consequences the possibility of building "accelerated P systems" capable of solving the HP.

The idea came from Russell and Weyl (early 30s) who observed that a process that performs its first step in one unit of (global) time, the second step in 1/2 unit of (global) time, and, in general, each subsequent step in half the (global) time of the step before, is able to complete an infinity of steps in just two global units of time since

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots = 2.$$

Grounded on suggestions coming from cell and brain biology we assume that acceleration is a part of the hardware (not a quality of the environment) and it is realised either by *decreasing the size of "reactors"* (thus making possible that reactants find each other and react in a shorter time), or by *training*, by *speeding-up the communication channels*

The general scenario is suggested in Figure 4: we have two scales of time, an *external, global* one, of the "user" of the accelerated device (the black box in the figure), and the *internal, local* time of the device. The problem is formulated in global time, at some moment t, and introduced into the accelerated device, which is able to perform an 'inner' infinite computation in a finite number, T, of external time units, when the "user" gets the answer to the problem.

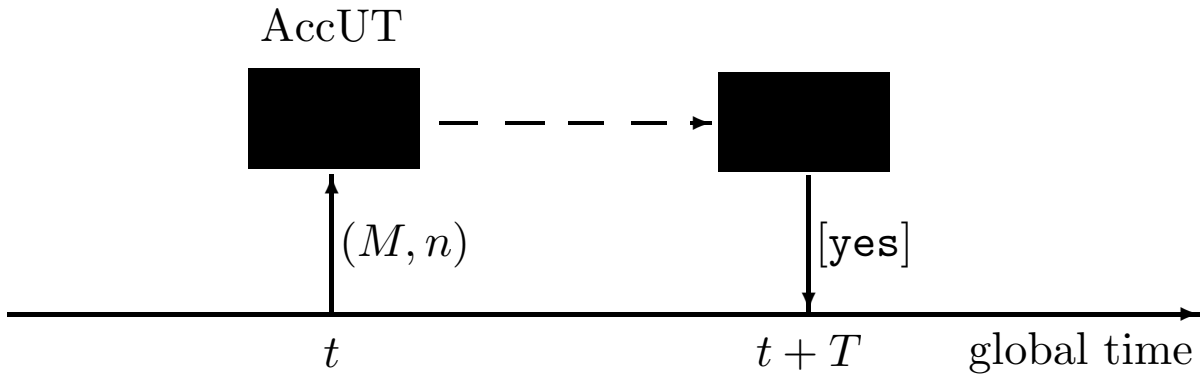


Figure 4: The interplay between local and global time used for solving the Halting Problem by means of an accelerated device