Lecture Notes
on
Physics of Classical and Quantum Information

Abstract

Notes on the course held at NTU by prof. Mile Gu. This notes are based on Nielsen and Chuang [2010] as well as the lecture notes from the aforesaid Gu [2018]. Extended to include the content of the course Fundamentals of Quantum Information held at TU Delft by Leo di Carlo and the course Quantum Communication and Cryptography teached by Stephanie Wehner.

Adrià Labay
NTU & TU Delft – July 31, 2020

Fuchs and Peres [2000]

> *Contrary to those desires, quantum theory does not describe physical reality. What it does is provide an algorithm for computing probabilities for the macroscopic events ("detector clicks") that are the consequences of our experimental interventions. This strict definition of the scope of quantum theory is the only interpretation ever needed, whether by experimenters or theorists.*

Preface

Fuchs [2002]

In all cases, a quantum state is specifically and only a mathematical symbol for capturing a set of beliefs or gambling commitments

Ever since the “discovery” of Quantum Mechanics (QM), the scientist community was shocked by the effects and strange phenomena emerging from this theory. Almost a century has passed and we still struggle to understand the surprising features of this theory, one of the most evolving of our present days, specially in the context of quantum information. This arises from the limits of classical computation that we face today. After all, information is encoded in a physical system, whatever it is and in the form we want, so the study of information and computation should be linked to the study of the underlying physical processes. This point of view is enriched in the well known statement “It from Bit” first pronounced by John Wheeler suggesting ”the idea that every item of the physical world has at bottom — at a very deep bottom, in most instances — an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and this is a participatory universe”.

Today, we still don’t have a unified view on what is a quantum state and what information does it encode. But this is a question that has been there since its origins, Einstein was one of the first who put into question the completeness of quantum mechanics [Einstein et al., 1935]. John Bell later proved that the hidden variable theory proposed by Einstein was not possible or it would otherwise violate local realism [Bell, 1964]. Nevertheless, even if we now take QM as a complete theory of reality, we haven’t been able to discover what a quantum state is. We should content ourselves with its probabilistic nature.

QM formalism is squeezed into 5 postulates (e.g. see Nielsen and Chuang [2010]) that cover the rules of a very big game. From these postulates, the scientist community has been able to move forward with the development of Quantum Field Theory with all its consequences and the creation of a new field in physics: Quantum Information. Many problems that are encountered in classical computation were brought to the quantum regime like cloning, teleportation, factoring, discrimination... “There is a feeling that the advent of quantum information theory heralds a new way of doing physics and supports the view that information should play a more central role in our world picture” says Fuchs [2002]. This is certainly a reality, for instance the European Union set a flagship in 2016 for the next 10 years with a founding of €1 billion* to investigate in the development of certain applications.

We seek to continue this trend with more and more people joining the world wide community of quantum scientists and engineers.

*https://qt.eu/about/

Contents

1 It from bit **5**
1.1 Experiements and computation . . . 5
1.2 Universal computer: Turing machines . . . 5
1.3 Limits of computation: halting problem . . . 6
1.4 Rice's theorem . . . 7
1.5 Emergence vs. reductionism . . . 7

2 Computational models **9**
2.1 Circuit model . . . 9
2.2 Game of life . . . 9
2.3 Ising model . . . 10

3 Introduction to quantum mechanics **13**
3.1 Observables . . . 13
3.2 Density operator . . . 14
3.3 Weirdness of quantum world: quantum interferometer . . . 15
3.4 Entangled states . . . 16
3.5 EPR theory and Bell inequality . . . 18
3.6 No cloning theorem . . . 21

4 Quantum State Discrimination **22**
4.1 On the notion of distance . . . 22
4.2 Quantum Hypothesis Testing . . . 24
4.3 Optimality conditions . . . 25
4.4 Ambiguous vs. unambiguous state discrimination . . . 25
4.5 Two-state discrimination . . . 26

5 Quantum computing **28**
5.1 P vs. NP . . . 28
5.2 Classical vs. Quantum computation . . . 29
5.3 Universal Quantum Gates . . . 29

6 Quantum algorithms **31**
6.1 Teleportation . . . 31
6.2 Dense coding . . . 31
6.3 Quantum parallelism . . . 32
6.4 Deutsch-Josza algorithm . . . 32
6.5 Bernstein-Varizani problem . . . 34
6.6 Simon's problem . . . 34
6.7 Trace estimation . . . 36
6.8 Grover search algorithm . . . 37
6.9 Quantum Fourier Transform (QFT) . . . 38
6.10 Quantum phase estimation . . . 39
6.11 Quantum period finding . . . 41
6.12 Shor's algorithm . . . 42

7 Quantum information **44**
7.1 Shannon entropy . . . 44
7.2 Von Neumann entropy . . . 45
7.3 Thermodynamic entropy . . . 46

8 Quantum cryptography **49**
8.1 Conditions of security 49
8.2 The quantum approach, how to make Eve ignorant? 50
8.3 Classical cryptography: RSA algorithm 50
8.4 BB84 50
8.5 Ekert91 51

9 Quantum Error Correction **52**
9.1 Indirect measurement 52
9.2 Bit flip code 53
9.3 Phase flip code 55
9.4 9 qubit Shor code 55
9.5 General quantum error correction protocol 56
9.6 Fault-tolerant quantum computation 57
9.7 Threshold theorem and concatenated codes 57
9.8 Entanglement distillation 58

1 It from bit

John Archibald Wheeler, 1990

> *'It from bit' symbolizes the idea that every item of the physical world has at bottom—a very deep bottom, in most instances—an immaterial source and explanation; that which we call reality arises in the last analysis from the posing of yes–no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and that this is a participatory universe.*

In this chapter, we begin this journey with a discussion on the ultimate limits of computation, and show how this can lead us to new understanding about universal principles of our reality. The standard reductionist rationale is that since all physical systems are built from conglomerations of fundamental particles, the principles governing these such particles will also apply to all such systems.

The plausible existence of universal systems suggests a complementary approach to the understanding of universal principles. If all observable qualities of any physical process may be simulated by a single system, then the limitations on that system will also be universal limitations that allow us to make generic statements about what we can observe within our universe. Indeed, there exists many tasks that universal computers cannot perform. Any computation is facilitated by a physical process, and any physical process can be thought of as a computation thus by knowing the computational limits we can have a knowledge on the physical limits of that system.

1.1 Experiements and computation

There is a one-by-one relationship between a computer simulation and an actual physic experiment: the initial configuration of the system correspond to the input bits, the program execution to the experiment itself and the results we obtain are the output bits.

Every physical experiment is a computation, and every computation is facilitated by some physical experiment. The class of all possible experiments define the class of all *computable functions*. A universal computer is a system capable of computing all functions that can be computed using any physically realisable system.

This motivates us to only consider models of computation that are *physically reasonable*. Any algorithm specified within such a model should admit a physical implementation that is experimentally possible, at least in principle. One postulate on the set of computations we would consider to be physically reasonable are:

1. Only a finite subsystem of the entire computer is “in motio” at any time.
2. The motion depends only on the state of a finite subsystem.
3. The amount of information used to specify this motion is finite.

These conditions for reasonable physical processes is not a mathematical definition and is completed dictated by our current knowledge of physics.

1.2 Universal computer: Turing machines

A universal computer is that capable of computing any function that may be computed on a reasonable model of computation. Such systems not only exists, but can be remarkably simple: Alan Turing proposed the Turing machine in 1936, a simple device which he claimed to be capable of performing any algorithmic process.

We can think of a Turing machine as a one dimensional tape divided in finite cells with possible states Σ. Over one of them, let's say at the k-th position, there's a head which is the brain of the machine. This brain has a possible set of states Q and can only see one cell at a time. Then a

Turing machine is just a function that maps the current cell state and the head state into a new cell and brain state and the next cell to look at (left or right cell). Mathematically, this reads

$$\begin{aligned} T : \Sigma \times Q &\longrightarrow \{\leftarrow, \rightarrow\} \times \Sigma \times Q \\ (s, q) &\longrightarrow (\leftrightarrow, s', q') \end{aligned}$$

So by imposing the rules over all the possible values of $\Sigma \times Q$ the machine is programmed and ready to work, we have created an algorithm the machine strictly follows.[*]

From an operational perspective, an algorithm is a set of instructions that a person or indeed, a sufficient sophisticated machine, can follow, without additional insight. Thus, such a process can be regarded as a physically realisable experiment, and corresponds to some reasonable model of computation. This leads to the Church-Turing thesis:

Thesis 1 (Church-Turing)**.** *Any function that can be computed by a reasonable model of computation can be computed by an universal Turing machine.*

The Church-Turing thesis is not a theorem, but rather a universal principle whose validity is based on observation. This implies the following principle

Principle 1 (Church-Turing)**.** *Every function that can be computed by a physically reasonable process can be computed by a Turing Machine.*

And coming back to the physical world this turns to

Principle 2 (Deutsch-Church-Turing)**.** *Every physically reasonable process can be exactly simulated by a Universal Computing Device.*

But, all universal computer are limited by the constrains of the physics and vice versa, so they must obey the *second law of thermodynamics* ($dS > 0$), the *theory of relativity*[†]...

1.3 Limits of computation: halting problem

Since any algorithm may be implemented by a Turing machine, any decision problem that no Turing machine can solve would not be solvable by any reasonable physical process (assuming the truth of the Church-Turing thesis).

The Halting problem is a classical example of a non-computable problem. No Turing machine can take an arbitrary Turing machine T and some number x as input, and output with certainty whether or not T will halt on input x. There exists no algorithm, and hence no physical process, that would allow us to determine whether a given algorithm will ever end or be trapped in an infinite loop. These type of machine is often called *Oracle*[‡].

Theorem 1. *There exists no Turing machine that can decide whether or not a general Turing Machine will halt in finite time.*

Proof. Let's proof this by contradiction, we will write a program that does the following:

```
Require: $h(T, x)$: halting function
Require: $x$: integer number
  $halt \longleftarrow h(h, x)$
  if $halt$ then
    loop_forever()
  else
    halt
  end if
```

We assume at first that there exist such a machine $h(T, x)$ that takes as input a Turing machine T and some general input x. On top of it we create another Turing machine that does the following: if T halts on input x then it loops forever and if it loops forever it halts.

[*] If you want to try to create your own Turing machine visit https://turingmachinesimulator.com/.

[†] Later on, we will see how this principles limit the transmission of information.

[‡] From the Greek mithology, people that transmit messages from gods to humans.

Now suppose you feed this machine with h itself then two things can happen: if h halts on input h then it loops forever and if h does not halt on input h then it halt. We end up with a contradiction in both cases but because our machine is well defined except for the Oracle machine we conclude that it can't exist. □

This non-computability theorem has a consequence on the physical world:

Theorem 2. *There exists no physically realisable process that can decide whether or not a Turing machine will halt in finite time.*

1.4 Rice's theorem

There's a generalisation to the halting problem which is Rice's theorem that says:

Theorem 3 (Rice)**.** *Any non-trivial black box property of a Turing machine is non-computable.*

Let's define what we understand by non-trivial black box property. First of all, a *property* is just a function that maps each Turing machine into $\{0, 1\}$, so it tells us if that certain machine has or doesn't have this property. Then, a black box property depends only on the input-output relations but not on the machine itself, thus if $T(x) = T'(x)$ for all x we say that $P(T) = P(T')$. Finally, non-trivial means that there exist other machines such that $T(x) \neq T''(x)$, P is not constant.

For example, say that we want to know if a machine T doubles its input for all x ($T(x) = 2x$), this is also a non-trivial black box property because it only depends on the input-output and there exist other machines that don't do this so the problem is non-computable (just as the halting problem).

As always, this has a physical consequence that is

Principle 3. *(Rice) There exists no reasonable physical process that can determine any given non-trivial black-box property of a Universal Computer.*

Rice's theorem, together with the Church-Turing thesis, prohibits the existence of any physical device that could predict the long term behavior of any other device of sufficient complexity.

1.5 Emergence vs. reductionism

When we look at the physical universe around us, we often observe some sort of 'macroscopic order'. When we analyse the flow of water, or the dynamics of a glacier, we do not need to compute the exact motion of every atom. The trick here is that when we observe the macroscopic world, we generally neglect the microscopic details.

We call these equations macroscopic laws. A *reductionist* view is that all macroscopic laws are logically implied by microscopic laws. Since a metal bar is made out of atoms, then the microscopic laws governing the dynamics of these atoms would allow you to systematically derive the laws that govern how the bar behaves under stress. Here 'systematically derive' implies formal, mathematical implication. That is, feed the microscopic laws that governed each atom and their interactions into a computer, along with formal definitions of the macroscopic observables, and it would eventually be able to output any law that governs those macroscopic observables. The derivation of macroscopic laws requires no additional 'insight' or 'creativity'.

The idea of *emergence* is completely opposite to reductionist, it is the belief that not all macroscopic laws of physics can be reduced to laws governing their microscopic constituents. Even if we are given all the fundamental laws of the universe that govern on all fundamental particles – we may still not be able to derive certain macroscopic phenomena with additional assumptions.

For example, the fundamental interaction of a Universal Computer (say a turing machine or the game of life as we will see later on) are known however there are macroscopic properties about this system that are non-computable. In fact, emergence is a consequence of Rice's thereom because if we can encode a physical system into a universal computer and the knowledge of an observable property O of the physical system S reveals the value of some black box property on T, then O is non-computable.

There is a recipe for demostrating emergence:

1. Take a macroscopic physical system
2. Demonstrate that its microscopic components interact in a way that is complex enough to encode a Turing machine
3. Show that a macroscopic property on this physical system is a black-box property of the underlying Turing machine.

Questions

1. In the standard Turing machine, the head of the tape can move only 1 step left or 1 step right per time-step. Consider a modified Turing machine where the head of the tape can point to any point on the potentially infinitely long tape.

   This is not a reasonable model of computation. If the head of the tape can point to any point on the potentially infinitely long tape, the number of instructions for this Turing machine should be infinite, which is not reasonable.

2. Which of the following properties are non-trivial black-box properties of a Turing machine?

   (a) The property that a given Turing machine has a tape.

   This is a trivial property which doesn't depend on the input-output of the machine and every one of them has.

   (b) The property that a given Turing machine always output 5 on input x.

   This is a non-trivial black-box properties which only depends on the input-output relations on a process, so this problem is not computable.

   (c) The property that a given Turing machine halt before 30 time-steps.

   This is not a black-box property. It depends on the internal structure of a Turing machine (e.g. the time-step), we just have to wait 30 time-steps and see if it halts or not.

3. Consider a restricted Turing machine T that operates on a tape of length N, where N is finite. Is the question "Does T halt on input" computable? You may assume that the Turing machine has a finite number of internal states.

   Computable. As the tape has finite length N, the number of the possible states of this restricted Turing machine is then finite (we use S to represent the number). As long as the time-steps taken when input x into Turing machine T is larger than S, T will not halt. The reason is because the machine, in the worst case has gone over all the possible states $s \in S$ so after s_S it must start repeating another state s_1 and because the rules are well defined after s_1 it will follow s_2 and so on until s_S and repeat again so it never halts. Otherwise, T will halt on input x before some certain time-steps. Thus, this question is computable.

4. Is the question, "Given Turing machines T and T', does T and T' halt on the same inputs?" for general T and T' computable?

   Non-computable. If the question "T and T' halt on the same input" is computable, we can assume a specific case where the Turing machine T' actually "does nothing" (identity). Then this question will become the same as the halting problem: "Does T halt on input x", which is non-computable. As a result, the original question is non-computable.

5. A Game of Life has been initialized in a way such that there are no living cells outside a 30×30 grid. Consider the question: "Will there ever be alive cells outside this 30×30 grid". Is this question computable?

   This is the same problem as in question number 3 because there is a finite number of possibilities, thus after $2^{30\times 30}$ time-steps we can know the answer to this question.

2 Computational models

Universal systems are in fact, not very rare at all. In this section, we will review the circuit model and cellular automata, both of which bear a closer relation to existing computers than the Turing machine.

2.1 Circuit model

An algorithm that acts on at most n bits of information can be defined by a binary function $f_n : \{0,1\}^n \to \{0,1\}^n$. Thus, for a model of computation to be universal, it is sufficient to show that model can evaluate all binary functions, f_n.

Any binary function may be decomposed into logical operations that act on at most two bits. In fact, a subset of such operations, referred to as elementary logic gates can be concatenated to evaluate any f_n. One such universal gate set involves the operations (see fig. 1): FANOUT, SWAP, AND and NOT.

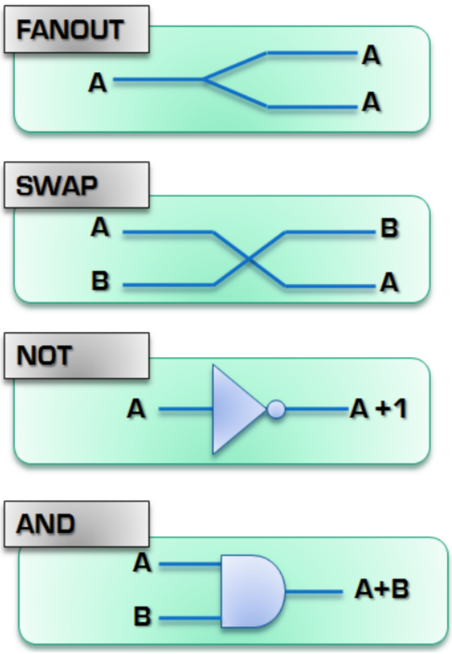

Figure 1: Primary logic gates

The two models relate to each other through the notion of uniform circuit families. Given any particular algorithm defined by a Turing machine T, we may define a family of circuits $\{C_n\}$ where the action of C_n coincides with the action T on input of size n. The set of circuits, when taken together, equate to the action of T.

There are of course more gates than the ones presented like XOR, XAND... but it turns out that those constitute a universal set, that is all the other gates can be built out of this primitive ones, which reduces to the following thesis

Thesis 2. *Any physical system that can synthesis the following elementary gates is capable of computation.*

2.2 Game of life

The universality of the Game of the Life results from its capacity to implement each of the operations necessary to construct an arbitrary logic circuit. In particular, one can demonstrate the existence of gliders, stables configurations of cells that propagate in some fixed sketched direction. The proof involves designating specific parts of the grid as wires, and encoding 1(0) as the existence (absence) of gliders on those particular wires. A series of rather ingenious constructions allow one create gliders and collide these gliders in specific fashions so that their interaction results the exact logic interactions required for universality.

While the circuit model appeared rather artificial since it required a careful concatenations of specific interactions in an exact order, the Game of Life appears more natural. The update rules of the Game of Life are isotropic, independent of location in the grid. The update rule may even be interpreted as a primitive model of evolving biological communities (and hence where the model got its name from). The universality of such a deceptively simple system suggests that universality may not be a rare trait.

The special thing this game has is that it is unpredictable following Rice's theorem, the problem of computing if a given cell will be alive in the future is totally uncomputable because the long term behaviour of the system can't be predicted.

2.3 Ising model

Square Ising lattices describe a classical system of spins arranged at the vertices of a d-dimensional rectangular grid. The state of each spin is described by a single value (0 or 1) and interacts only with its $2d$ neighbours.

Mathematically, we index each spin of the $2d$ square Ising lattice by a vector of integers $\boldsymbol{x}_{ij} = (i, j)$, such that $s_{ij} \in \{0, 1\}$ denotes the state of the spin at location x_{ij}. Interactions on this lattice are described by the Hamiltonian $\mathcal{H}$, a function that maps each configuration of the lattice, A, to a real number. In the setting of magnetism, $\mathcal{H}(A)$ would correspond to the potential energy of a lattice in configuration A.

The general Ising model with an external field has a Hamiltonian of the form

$$\mathcal{H}(\mathcal{S}) = \sum_{i,j} M_{ij} s_{ij} + \sum_{i,j} \sum_{\substack{k \neq i \\ l \neq j}} c_{ij}^{kl} s_{ij} s_{kl} \tag{2.1}$$

where c_{ij}^{kl} is the interaction between spins and M_{ij} the external field applied at position $\boldsymbol{x}_{ij}$. For the square Ising model, c_{ij}^{kl} for all spins $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{kl}$ that are not directly adjacent to each other, this is $(k, l) = \{(i + 1, j), (i - 1, j), (i, j + 1), (i, j - 1)\}$.

In statistical mechanics, the probability that a system exists in a state A is directly proportional to $\exp[-\mathcal{H}(A)/(k_b T)]$, where k_b is the Boltzmann constant, and T the temperature of the system. Thus, the lattice tends to be in configurations where $\mathcal{H}(A)$ is small. The configurations that minimize $\mathcal{H}$ are referred to as ground states, and represents the possible states of the system at zero temperature. Observe that it is always possible to label the spin states such that one of the ground states is 0. Thus, we assert that 0 is a ground state of $\mathcal{H}$ with $\mathcal{H}(0) = 0$ without loss of generality.

Our goal is to encode a graph of spins defining the values of M and c for each spin and vertex so that when it is cooled down to its ground state the system acts as a defined logical gate.

Designer Ising blocks We make use of Designer Ising blocks, bounded 2-dimensional blocks of spins with an associated Hamiltonian whose ground state encodes a desired logical operation f. Input is encoded in bits on one boundary of the block, while output bits on the boundary opposite. Formally, consider an arbitrary binary function f with m inputs and n outputs; we define a designer Ising block as follows. Take a $C \times D$ block of spins, where $C, D > \max(m, n)$, governed by a Hamiltonian $\mathcal{H}_f$ with ground state set G_f. We designate m input spins, $\boldsymbol{s} = (s_1, \dots, s_m)$ from the first row to encode the input and n output spins, $\boldsymbol{r} = (r_1, \dots, r_n)$ from the last row as output.

We say a configuration of the lattice, s, satisfies $\{\boldsymbol{s}, \boldsymbol{r}\}$ if the input and output spins are in states $\boldsymbol{s}$ and $\boldsymbol{r}$ respectively. Suppose that

(a) There exists $\boldsymbol{s} \in G_f$ that satisfies $\{\boldsymbol{s}, \cdot\}$ for each of the 2^m possible inputs of f.

(b) Every $\boldsymbol{s} \in G_f$ satisfies $\{\boldsymbol{s}, \boldsymbol{r} = f(\boldsymbol{s})\}$.

then we can set the ground state of the Ising block to encode the action of f on any desired input by appropriately biasing the input spins by external fields. Such an encoding is, in fact, universal.

Theorem 4. *For any binary function f, we can construct a designer Ising block such that the conditions (a) and (b) hold.*

Proof. The proof of this theorem is done by constructing a set of universal gates as Ising blocks (see fig. 2), from them we can automatically build more complex circuits. □

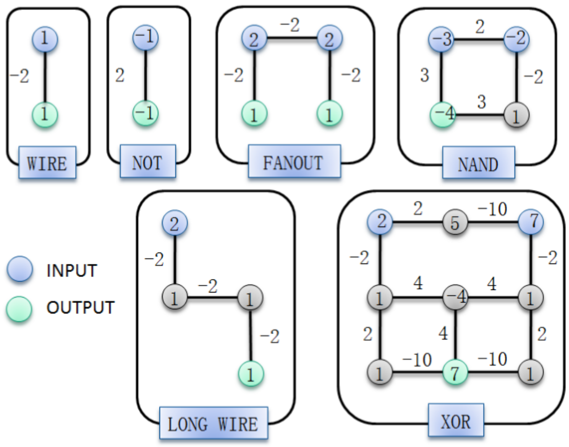

Figure 2: Ising blocks for a universal set of elementary gates.

Let's take the NOT gate as an example, the Hamiltonian of the system is

$$\mathcal{H}(s_i, s_o) = -s_i - s_o + 2s_i s_o$$

where s_i and s_o are the input and output bits respectively. Then, the four possibilities are

s_i	s_o	$\mathcal{H}$
0	0	0
0	1	-1
1	0	-1
1	1	0

We have two ground states $\{(0,1),(1,0)\}$, if the input is 1 then the output is 0 and vice versa. This is exactly the expected behaviour of a NOT gate which negates the input bit. If we do the same for the other gates in fig. 2 we'd obtain that the ground states correspond exactly to the behaviour of the corresponding gate.

Questions

1. Can you write down two sets of elementary gates that are universal?

   The two sets are {NOT, AND, XOR, OR} and {NAND, FANOUT, SWAP}. It can be proved that by just using the second set, you can build all the gates in the first one.

2. The ground state of an Ising lattice is the configuration which has lowest energy. Answer the following questions:

   (a) Consider two Hamiltonians $\mathcal{H}$ and $\mathcal{H}' = \mathcal{H} + C$, where C is some constant. Does this guarantee that $\mathcal{H}$ and $\mathcal{H}'$ have the same ground state set?

   Yes, they still have the same ground state because all the energies are displaced by the same amount.

   (b) Consider a system with two spins b_1 and b_2, with the Hamiltonian $\mathcal{H}(b_1, b_2) = b_1 + b_2 - 2b_1 b_2$. What is its ground state?

   Let's construct the lookup table (see below). The ground states are $\{(0,0),(1,1)\}$ in fact this Hamiltonian correspond to a wire (see fig. 2).

s_i	s_o	$\mathcal{H}$
0	0	0
0	1	1
1	0	1
1	1	0

3. Can you write down the Hamiltonian for a n-bit wire, acting on a chain of N spins, such that at ground state, they are all identically 0, or identically 1?

   The complete Hamiltonian for this Ising block will be

   $$\mathcal{H}(b_1, \ldots, b_N) = \sum_{i=1}^{N-1} \mathcal{H}(b_i, b_{i+1})$$

   where $H(b_i, b_{i+1}) = b_i + b_{i+1} - 2b_i b_{i+1}$ is the Hamiltonian for one single wire block.

3 Introduction to quantum mechanics

The bit is the fundamental concept of classical computation and classical information. Quantum computation and quantum information are built upon an analogous concept, the quantum bit, or qubit for short. What then is a qubit? Just as a classical bit has a state – either 0 or 1 – a qubit also has a state. Two possible states for a qubit are the states $|0\rangle$ and $|1\rangle$, which as you might guess correspond to the states 0 and 1 for a classical bit. The difference between bits and qubits is that a qubit can be in a state other than $|0\rangle$ or $|1\rangle$. It is also possible to form linear combinations of states, often called superposition:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{3.1}$$

When we measure a qubit we get either the result 0, with probability $|\alpha|^2$, or the result 1, with probability $|\beta|^2$. Naturally, $|\alpha|^2+|\beta|^2 = 1$, since the probabilities must sum to one. Geometrically, we can interpret this as the condition that the qubit's state be normalised to length 1. Thus, in general a qubit's state is a unit vector in a two-dimensional complex vector space. The basis $\{|0\rangle\,,|1\rangle\}$ is known as the computational basis.

We may write the most general qubit in the form

$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|1\rangle \tag{3.2}$$

where $\theta, \varphi \in \mathbb{R}$ define a point in a $3d$ sphere of radius 1. This sphere is often called *Bloch sphere.*

3.1 Observables

A measurement is made by an observable $\mathcal{A}$ that has an associated Hermitian matrix or *operator* A which acts on the quantum states of some Hilbert space. In the computational basis, the operator reads $A = \sum_{j,k=1}^{n} a_{jk}\,|j\rangle\langle k|$ where $|j\rangle\langle k|$ is the outer product.

The hermiticity property of all observables allows a spectral decomposition as a sum $A = \sum_\lambda a_\lambda \hat{P}_\lambda$ being $\{a_\lambda\}$ the eigenvalues of A and $\hat{P}_\lambda$ the projector onto the eigenspace spanned by the eigenvectors corresponding to a_λ obeying the orthogonality and completeness relations

$$\hat{P}_\lambda\hat{P}_\mu = \hat{P}_\lambda\delta_{\lambda\mu} \tag{3.3a}$$

$$\sum_\lambda \hat{P}_\lambda = \mathbb{I}_d \tag{3.3b}$$

If A is a physical observable, then $\{a_\lambda\}$ are the physical values that we can observe after measuring a state $|\psi\rangle$. The probability that the result a_λ is obtained given that the state measured was $|\psi\rangle$ is

$$p(a_\lambda|\psi) = \langle\psi|\hat{P}_\lambda|\psi\rangle \tag{3.4}$$

and it holds that $\sum_\lambda p(a_\lambda|\psi) = 1$ on account of eq. (3.3b).Also, the state after the measuremnent becomes

$$|\psi'\rangle = \frac{\hat{P}_\lambda\,|\psi\rangle}{\sqrt{p(a_\lambda|\psi)}} \tag{3.5}$$

The measure is completely defined once the operators $\{\hat{P}_\lambda\}$ are given, then for each $\hat{P}_\lambda$ we associate the hypothesis that the value of the physical property $\mathcal{A}$ observed is a_λ. This measure is called projective or *von Neumann measure* because the elements are orthogonal projectors (eq. (3.3a)). The number of projectors is limited by the dimension of the space, otherwise the orthogonality condition wouldn't be satisfied. For this reason, we define a generalised measurement or Positive-Operator Value Measure (POVM) as a set of positive operators $\{\Pi_j\}_{j=1}^n$, with n not necessarily equal to d, satisfying the completeness and positivity conditions

$$\sum_{j=1}^{n} \Pi_j = \mathbb{I}_d \tag{3.6a}$$

$$\Pi_j \geq 0 \qquad \forall j \tag{3.6b}$$

An observable of $\mathcal{H}_2$ is expressed, in the most general form, as a complex linear combination of the identity matrix $\mathbb{I}_2$ and the *Pauli matrices* $\{\sigma_i\}_{i=1}^3$[*] which span the space of 2×2 Hermitian matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.7}$$

The elements of the computational basis are by convention the eigenvectors of the third Pauli matrix such that $\sigma_3 |0\rangle = + |0\rangle$ and $\sigma_3 |1\rangle = - |1\rangle$.

The Pauli matrices satisfy the product relations

$$\sigma_i \sigma_j = \delta_{ij} + i\epsilon_{ijk}\sigma_k \tag{3.8}$$

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \tag{3.9}$$

$$\{\sigma_i, \sigma_j\} = 0 \tag{3.10}$$

Moreover, the 3 Pauli matrices are the generators of rotations in $SU(2)$, the symmetry group of qubits. A rotation of a qubit $|\psi\rangle$ along the direction $\boldsymbol{n}$ by an angle θ is performed by the unitary operator

$$U_{\theta,\boldsymbol{n}} = \exp\left(-i\frac{\theta}{2}\boldsymbol{n} \cdot \boldsymbol{\sigma}\right) = \cos\frac{\theta}{2}\mathbb{I} - i \sin\frac{\theta}{2}\boldsymbol{n} \cdot \boldsymbol{\sigma} \tag{3.11}$$

Specifically, a rotation of $\theta = \pi$ along the y axis is known as the Hadamard gate that have the expression

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.12}$$

Essentially, it maps the Z basis to the X basis: $H |0\rangle = |+\rangle$ and $H |1\rangle = |-\rangle$.

Other impotant gates which are quite very used in the construction of quantum circuits are the $\pi/2$ and $\pi/4$ gate represented by the letters S and T respectively. They provide rotations along the Z axis of the specified angle,

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \quad , \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \tag{3.13}$$

Note that $S = T^2$.

3.2 Density operator

The density operator language provides a convenient way for describing quantum systems whose state is not completely known. More precisely, suppose a quantum system is in one of a number of states $|\psi_k\rangle$ with respective probabilities p_k. We shall call $\{p_k, |\psi_k\rangle\}$ an ensemble of pure states. The density operator for the system is defined by the equation

$$\rho \equiv \sum_k p_k |\psi_k\rangle\langle\psi_k| \tag{3.14}$$

The density operator is often known as the *density matrix.* It turns out that all the postulates of quantum mechanics can be reformulated in terms of the density operator language.

Suppose that the evolution of a closed quantum system is described by the unitary operator U. If the system was initially in the state ψ_k with probability p_k then after the evolution has occurred the system will be in the state $U |\psi_k\rangle$ with probability p_k. Thus, the evolution of the density operator is described by the equation

$$\rho_t = \sum_k p_k U |\psi_k\rangle \langle\psi_k| U^\dagger = U\rho_0 U^\dagger \tag{3.15}$$

[*] Sometimes also expressed as $\{\sigma_x, \sigma_y, \sigma_z\}$.

Under this formulation, the probability of finding the state $|\phi\rangle$ is[*]

$$p(\phi) = \mathrm{Tr}(\rho\, |\phi\rangle\langle\phi|) = \langle\phi|\rho|\phi\rangle \tag{3.16}$$

and the expectation value of an observable A is

$$\langle A\rangle = \mathrm{Tr}(A\rho) \tag{3.17}$$

The density operator contains all the information about a physical system needed to deduce its output statistics when measured in any basis. This implies that to physical systems with the same density operator are completely equivalent and indistinguishable.

The properties of the density operator are:

- Hermitian: $\rho = \rho^\dagger$.
- Unit trace: $\mathrm{Tr}(\rho) = 1$, all the probabilities must add to 1.
- Non-negative: $\langle\phi|\rho|\phi\rangle \geq 0\ \forall\phi$, all eigenvalues are real and non-negative.

Classical states Suppose X is a classical random variable described by the probability distribution $P(X)$ where p_x denotes the probability of finding symbol x. We can express a quantum source preparing the states $|x\rangle$ as

$$\rho_X = \sum_x p_x\, |x\rangle\langle x| \tag{3.18}$$

where the states $|x\rangle$ are mutually orthogonal and form a basis, thus the density matrix is diagonal in this basis. In this situation, we say that ρ is classical with respect to X.

3.3 Weirdness of quantum world: quantum interferometer

Niels Bohr

> *Everything we call real is made of things that cannot be regarded as real. If quantum mechanics hasn't profoundly shocked you, you haven't understood it yet.*

Let's illustrate some of the surprising properties of the quantum world using the historical example of the quantum bomb detector.[†]

Suppose that we have a system like the one in fig. 3, where laser light goes through a $50-50$ beamsplitter (half to the top mirror and half to the bottom mirror) reflects in those mirrors and joins again in the second beam splitter. Classically, light after each reflection gains a phase of π so at the end we won't see light on the upper arm because it interferes destructively; otherwise, all the light will be on the lower arm because both path interfere constructively. This happens when we send a continuous beam of photons but what if we send single photons through the interferometer?

Experimentally, it has been found that when one sends a single photon the same interference pattern is obtained: if the laser is in the upper arm then the photon will exit the system on th lower arm and vice versa. This makes no sense classically because it means that the photons has interfered with itself, it has travelled through both arms!! The photon inside the interferometer is in a superposition of the upper and lower state. This explanation, although it may sound strange, has been the only reasonable explanation of this phenomena.

Suppose we put a detector in one of the arms inside the interferometer such that if the photon goes through it we can know. Following our previous statement, it must always "click" because it crosses both path but it turns out that this is not what actually happens. When we situate a detector, the interference pattern is lost, the photon is no longer in a superposition of both states because we are observing it. Now, half of the times goes through the upper arm and half of the times through the lower arm. Heisenberg resumed this fact in his principle

[*] Remember that the trace is invariant under cyclic permutations, $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$, and that $\mathrm{Tr}(A\otimes B) = \mathrm{Tr}(A)\cdot\mathrm{Tr}(B)$.

[†] Of course, there are other examples like the double slit experiment but I'm really tired of this experiment.

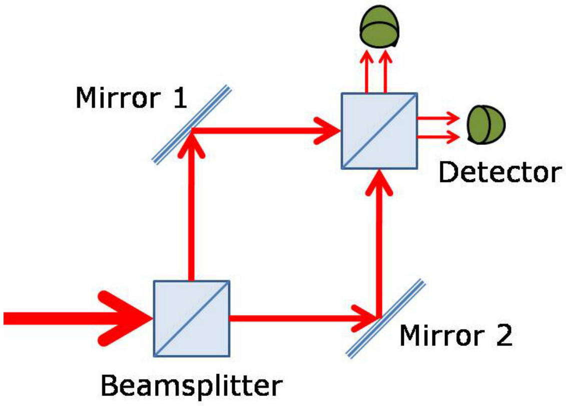

Figure 3: Quantum interferometer

Principle 4 (Heisenberg uncertainty principle)**.** *The act of measuring a quantum state disturbs the state.*

Let's look at the consequences this principle has on quantum computer, the fact that by observing a state we change it and the whole superposition is gone means that if we want to build a quantum computer it must be isolated from the whole world because any tiny perturbation on a quantum state may change the expected result.

Thesis 3 (Quantum computing paradox)**.** *A quantum computer only does what it is meant to do if we don't know what it is doing.*

3.4 Entangled states

In general, a physical quantum state is an element of a Hilbert space $\mathcal{H}$ with dimension $d = \dim\mathcal{H}$. If $\{|k\rangle\}_{k=0,\dots,d}$ is a basis of this space then, any state $|\psi\rangle$ in this space can be constructed as a lineal superposition of this states with d complex amplitudes $\{c_k = r_k e^{i\phi_k}\}$

$$|\psi\rangle = c_0 |0\rangle + \cdots + c_d |d\rangle = \sum_{k=0}^{d} c_k |k\rangle \tag{3.19}$$

We'd say that we have $2d$ degrees of freedom at first sight but that's not true because we have a constrain which is $\langle\psi|\psi\rangle = 1$ and we know that any two physical systems which differ only by a global phase factor are physically identical so there are in fact only $2(d-1)$ degrees of freedom.

The way to mix to states is by the tensor product defined as $|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = |\psi_1\rangle |\psi_2\rangle = |\psi_1\psi_2\rangle$. In vector form, if $|\psi_1\rangle = \sum c_k |k\rangle$ and $|\psi_2\rangle = d_k |k\rangle$ then

$$|\Psi\rangle = |\psi_1\rangle |\psi_2\rangle = \sum_{k,l=0}^{d} c_k d_l |k\rangle |l\rangle = \sum_{k,l=0}^{d} \psi_{kl} |kl\rangle \tag{3.20}$$

Apart from the usual properties (associative, lineal, scalar multiplication...) the most important is the inner product because each ket interacts only with the one of the same kind

$$\langle\Psi|\Psi\rangle = (\langle\psi_1| \langle\psi_2|)(|\psi_1\rangle |\psi_2\rangle) = \langle\psi_1|\psi_1\rangle \langle\psi_2|\psi_2\rangle \tag{3.21}$$

The question now is how those new states behave under measurements. Suppose we have two systems A and B with a joined state $|\Psi\rangle = |\psi_A\rangle |\psi_B\rangle$ with U_A and U_B being observable of the two Hilbert spaces (not that they do not have to be of the same dimension) then $U_A \otimes U_B$ is an observable of the total system where

$$U_A \otimes U_B = \left(\sum_{i,j=0}^{d_A} a_{ij} |i\rangle \langle j| \right) \left(\sum_{k,l=0}^{d_B} b_{kl} |k\rangle \langle l| \right) = \sum_{i,j=0}^{d_A} \sum_{k,l=0}^{d_B} a_{ij} b_{kl} |i,k\rangle \langle j,l| \tag{3.22}$$

where the new observable has a total dimension of $d = d_A d_B$. Following from the previous property, the expectation value is also taken with respect to the states corresponding to the same system. If you only want to observe one of the states is as simple as choosing as your observable for that system the identity, like $\mathbb{I} \otimes U_B$.

This is a way of constructing mixed states between systems but note that not all states can be decomposed into its subsystems, $|\Phi\rangle \neq |\phi_1\rangle |\phi_2\rangle$, this is the origin of entanglement.

Principle 5. *A quantum state $|\Psi\rangle$ on a n mixed system $\mathcal{H}^{\otimes n}$ is said to be entangled if there does exist no $|\psi_1\rangle, \ldots, |\psi_n\rangle$ such that $|\Psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$.*

An example of entangled states are the Bell states

$$|\phi_+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{3.23a}$$

$$|\phi_-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{3.23b}$$

$$|\psi_+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{3.23c}$$

$$|\psi_-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{3.23d}$$

the last one of these is also called the *singlet* as it represents a two state system with spin 0, invariant under rotations. The way to construct them is by using entangling gates like the C-NOT (controlled not) gate

$$C_X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.24}$$

which negates the second qubit if the first one is set to $|1\rangle$, so if we apply C_X on $|+\rangle |0\rangle$ the resulting state is exactly $|\phi_+\rangle$.

Other two qubit quantum gates are the control phase gate C_Z and the control swap gate S which operate on $|jk\rangle$ as $C_Z |jk\rangle = (-1)^{jk} |jk\rangle$ and $S |jk\rangle = |kj\rangle$. In matrix form their expression is

$$C_Z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad , \quad S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.25}$$

Partial trace The partial trace Tr_B is a mapping from the density matrices ρ^{AB} on a composite space $\mathcal{H}_A \otimes \mathcal{H}_B$ onto density matrices ρ^A on $\mathcal{H}_A$. It is defined as the linear extension of the mapping

$$\mathrm{Tr}_B : \quad M_A \otimes M_B \longrightarrow M_A \,\mathrm{Tr}(M_B) \tag{3.26}$$

for any matrix M_A on $\mathcal{H}_A$ and M_B on $\mathcal{H}_B$ and Tr indicates the normal trace.

Let $\{|a_i\rangle\}$ be a basis of $\mathcal{H}_A$ and $\{|b_i\rangle\}$ be a basis of $\mathcal{H}_B$. Any density matrix ρ^{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ can then be decomposed as $\rho_{AB} = \sum_{ijkl} c_{ijkl} |a_i\rangle\langle a_j| \otimes |b_k\rangle\langle b_l|$ and the partial trace reads

$$\rho^A = \mathrm{Tr}_B\, \rho^{AB} = \sum_{ijkl} c_{ijkl} |a_i\rangle\langle a_j| \langle b_l|b_k\rangle \tag{3.27}$$

which is the density matrix of $\mathcal{H}_A$. Note, that $\mathrm{Tr}(|b_k\rangle\langle b_l|) = \langle b_l|b_k\rangle$ is in general complex number (note the flip of indices).

Take into account that $\rho^{AB} \neq \rho^A \otimes \rho^B$ for correlated systems (entagled), equality only occurs when all the subsystems are uncorrelated.

For example, consider a general density matrix for two qubits $\rho^{AB} = \sum_{i,j=0}^{1} \sum_{k,l=0}^{1} \rho_{i,j}^{k,l} \, |i,k\rangle\langle j,l|$ with a total dimension of 4×4. According to the previous expression the partial trace is calculated as

$$\rho^A = \mathrm{Tr}_B \, \rho^{AB} = (\rho_{00}^{00} + \rho_{01}^{01}) \, |0\rangle\langle 0| + (\rho_{10}^{00} + \rho_{11}^{01}) \, |0\rangle\langle 1| + (\rho_{00}^{10} + \rho_{01}^{11}) \, |1\rangle\langle 0| + (\rho_{10}^{10} + \rho_{11}^{11}) \, |1\rangle\langle 1| \tag{3.28}$$

On the other hand, the partial trace with respect to A is

$$\rho^B = \mathrm{Tr}_A \, \rho^{AB} = (\rho_{00}^{00} + \rho_{10}^{10}) \, |0\rangle\langle 0| + (\rho_{01}^{00} + \rho_{11}^{10}) \, |0\rangle\langle 1| + (\rho_{00}^{01} + \rho_{10}^{11}) \, |1\rangle\langle 0| + (\rho_{01}^{01} + \rho_{11}^{11}) \, |1\rangle\langle 1| \tag{3.29}$$

In matrix form, more user friendly, these results read

$$\rho^{AB} = \begin{pmatrix} \rho_{00}^{00} & \rho_{00}^{01} & \rho_{00}^{10} & \rho_{00}^{11} \\ \rho_{01}^{00} & \rho_{01}^{01} & \rho_{01}^{10} & \rho_{01}^{11} \\ \rho_{10}^{00} & \rho_{10}^{01} & \rho_{10}^{10} & \rho_{10}^{11} \\ \rho_{11}^{00} & \rho_{11}^{01} & \rho_{11}^{10} & \rho_{11}^{11} \end{pmatrix} \longrightarrow \begin{cases} \rho^A & = \begin{pmatrix} \rho_{00}^{00} + \rho_{01}^{01} & \rho_{00}^{10} + \rho_{01}^{11} \\ \rho_{10}^{00} + \rho_{11}^{01} & \rho_{10}^{10} + \rho_{11}^{11} \end{pmatrix} \\ \rho^B & = \begin{pmatrix} \rho_{00}^{00} + \rho_{10}^{10} & \rho_{00}^{01} + \rho_{10}^{11} \\ \rho_{01}^{00} + \rho_{11}^{10} & \rho_{01}^{01} + \rho_{11}^{11} \end{pmatrix} \end{cases} \tag{3.30}$$

Measuring entangled states As said, the measurement on one quantum state will only apply to the ket corresponding to the observable used but the other will remain the same. Now suppose we start with the *singlet* state ψ_- and we want to measure the first qubit with $U_1 = |0\rangle\langle 0|$ so $U_T = |0\rangle\langle 0| \otimes \mathbb{1}_B$, obviously the first qubit will collapse to the $|0\rangle$ state but due to the entanglement the second qubit will collapse too (in this case to $|1\rangle$). Think of it, we have collapsed the second qubit without interacting with it, only by measuring the first one.

This is called the *locality problem* because from special relativity we know that

Principle 6. *No physical object can travel faster than light in anu inertial frame.*

But quantum mechanics at no point puts any restriction on the maximum distance at which two entangled systems can collapse thus allowing *instantaneous* collapsing of states. Can this two theories agree?

The answer is yes, but first let's consider the consequences of instant or faster-than-light communication. If this existed then causality will be broken and you could send messages to the past because any signal sent will be received before it was actually sent. To overcome this issue the Einstein-Podolsky-Rosen was born.

3.5 EPR theory and Bell inequality

The EPR theory came because Einstein (& friends) weren't really happy how quantum mechanics could break its beautiful theory of relativity. Einstein, Podolsky and Rosen (EPR) came up with a temporarily solution with what they termed "elements of reality". Their belief was that any such element of reality must be represented in any complete physical theory. The goal of the argument was to show that quantum mechanics is not a complete physical theory, by identifying elements of reality that were not included in quantum mechanics.

The way they attempted to do this was by introducing what they claimed was a sufficient condition for a physical property to be an element of reality, namely, that it be possible to predict with certainty the value that property will have, immediately before measurement. In simpler words, that any entangled state will have prior to any measurement all the possible outcomes in the so called *hidden variables.*[*]

The key to this experimental invalidation is a result known as Bell's inequality. To obtain Bell's inequality, we're going to do a thought experiment, which we will analyse using our common sense notions of how the world works – the sort of notions Einstein and his collaborators thought nature ought to obey. After we have done the common sense analysis, we will perform a quantum mechanical analysis which we can show is not consistent with the common sense analysis. Nature can then be asked, by means of a real experiment, to decide between our common sense notions of how the world works, and quantum mechanics.

[*] Spoiler, nature itself invalidates this point of view while agreeing with quantum mechanics.

We start with some kind of referee who prepares two particles and sends one of them to Alice and one of them to Bob. Alice will randomly choose A_1 and A_2 to measure her particle while Bob will also randomly choose between B_1 and B_2. For simplicity, the result of all four observable will be either 1 or -1 and they are made at the same time (in causally disconnected frames of reference). Let's say that they lose 1 coin if they agree when choosing A_1 and B_2 but win 1 coin if they agree in the other cases, the quantity to be computed is $z = a_1b_1 + a_2b_1 + a_2b_2 - a_1b_2$.

EPR approach This is the same as saying the classical approach. Considering z, it can be expressed as $z = (a_1+a_2)b_1 + (a_2-a_1)b_2$ but because $a_1, a_2 = \pm 1$ either $a_1 + a_2 = 0$ or $a_1 - a_2 = 0$ so in both cases we conclude that $z = \pm 2$. If $p(a_1, a_2, b_1, b_2)$ is the probability that, before the measurements are performed, the system is in the state a_1, a_2, b_1, b_2 then the expectation value for z is

$$\langle z \rangle = \sum_{a_1,a_2,b_1,b_2} p(a_1, a_2, b_1, b_2)(a_1b_1 + a_2b_1 + a_2b_2 - a_1b_2) \leq 2 \tag{3.31}$$

in fact $-2 < \langle z \rangle < 2$. But we can also express this expectation value like

$$\langle z \rangle = \langle A_1B_1 \rangle + \langle A_2B_1 \rangle + \langle A_2B_2 \rangle - \langle A_1B_2 \rangle \leq 2 \tag{3.32}$$

This is the so called *Bell inequality.* By repeating the experiment many times, Alice and Bob can determine each quantity on the left hand side of Bell inequality.

Quantum approach Suppose the two particles are entangled in the singlet state $|\psi_-\rangle$ and the observables of Alice and Bob are chosen to be

$$\begin{aligned} A_1 &= Z_1 \quad , \quad B_1 = \frac{1}{\sqrt{2}}(Z_2 + X_2) \\ A_2 &= X_1 \quad , \quad B_2 = \frac{1}{\sqrt{2}}(Z_2 - X_2) \end{aligned} \tag{3.33}$$

The expectation value of the 4 separate quantities is easy to calculate, here we present the calculus for the first one $\langle A_1B_1 \rangle$:

$$\begin{aligned} \langle A_1B_1 \rangle &= \frac{1}{\sqrt{2}} \langle Z_1 \otimes (Z_2 + X_2) \rangle = \frac{1}{\sqrt{2}}(\langle Z_1Z_2 \rangle + \langle Z_1X_2 \rangle) \\ &= \frac{1}{2} [(\langle Z_1 \rangle_0 \langle Z_2 \rangle_1 - \langle Z_1 \rangle_1 \langle Z_2 \rangle_0) + (\langle Z_1 \rangle_0 \langle X_2 \rangle_1 - \langle Z_1 \rangle_1 \langle X_2 \rangle_0)] \\ &= \frac{1}{2} [(-1 - (-1)) + (\langle X_2 \rangle_1 + \langle X_2 \rangle_0)] \\ &= \frac{\sqrt{2}}{2} \left(\frac{\langle 0| + \langle 1|}{\sqrt{2}} \right) X_2 \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right) \\ &= \frac{1}{\sqrt{2}} \langle + | X_2 | + \rangle = \frac{1}{\sqrt{2}} \end{aligned}$$

With similar calculations for the other values, the resulting expectations values are $\langle A_2B_1 \rangle = \langle A_2B_2 \rangle = 1/\sqrt{2}$ and $\langle A_1B_2 \rangle = -1/\sqrt{2}$. Thus, the expectation value of the quantity z is

$$\langle z \rangle = \frac{3}{\sqrt{2}} + \frac{1}{\sqrt{2}} = 2\sqrt{2} \tag{3.34}$$

Which is greater than what any classical theory would predict.

Now it's time to do ask nature which of the two inequalities is satisfied and it turns out that *Bell inequality* ($\langle z \rangle \leq 2$) is not obeyed by nature implying that some of the assumptions made are not correct, either

1. <u>Realism</u>: the assumption that the physical properties A_1, A_2, B_1 nad B_2 have definite values a_1, a_2, b_1, b_2 which exist independent of observation, or

2. Locality: the assumption that Alice performing her measurement does not influence the result of Bob's measurement.

Indeed, Bell's inequality by quantum mechanics is not a proof of its non locality. Quantum theory is essentially local. Bell's discovery was that any realistic theory that could mimic quantum mechanics would necessarily be non local [Fuchs and Peres, 2000].

Locality vs. realism These two assumptions made are the key point of Bell inequality so one of them must be wrong, physicist tend to say that the world is not *locally realistic*. Either we have to assume that there are no hidden variables that govern the evolution of each particle from its birth (free will) or that the collapsing of the wave function must be done instantaneously.

How can we overcome this issue? Well, in quantum information theories the way of not violating the theory of relativity is by saying that *information can't travel faster than light*. Think of two observers which share an entangled state but they are very far away from each other so light takes some observable amount of time to reach them. Suppose Alice observe her qubit and sees a +1 then Alice knows that Bob's qubit will be at the state −1, now she sends a message to Bob telling him the state of his qubit. This message has to be send using light so only at a maximum velocity of c Bob can know that Alice has collapsed the wavefunction, allowing Bob to measure a definite state for his qubit without breaking causality [Sakurai and Napolitano, 2017].

Let's try to prove this formally, we want to see that it's not possible for Bob to gain any information after Alice has collapsed her qubit. Suppose they are sharing the state $|\Psi\rangle = \sum_{ij} c_{ij} |i\rangle |j\rangle$, after Alice has measured $|k\rangle$ the new state is $|\psi_k\rangle = \sum_{ij} c_{ij} \langle k|i\rangle |j\rangle / \sqrt{p_k}$ so the density matrix of Bob is

$$\rho_B = \sum_k |\psi_k\rangle\langle\psi_k| = \sum_{i,j,i',j'} c_{ij}c^*_{i'j'} \langle i'| \left(\sum_k |k\rangle\langle k|\right) |i\rangle |j\rangle\langle j'| = \sum_{j,j'}\sum_i c_{ij}c_{ij'} |j\rangle\langle j'|$$

Note that the dependence on $|k\rangle$ has been removed from the density matrix thus Bob has no knowledge on what Alice has measured before. From this it is argued that, statistically, Bob cannot tell the difference between what Alice did and a random measurement (or whether she did anything at all).

Principle 7. *Without knowledge of Alice's measurement result, Bob can't gain any information about which basis Alice has measured in.*

In our example, Alice and Bob share the Bell state $|\psi_-\rangle$ so the density matrix of the system is

$$\rho_{AB} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.35}$$

where we see that both systems are clearly entangled but when we calculate the local density matrix (what each of them see) we obtain

$$\rho^A = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \rho^B = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.36}$$

Locally, Alice and Bob see completely random measurements outcomes regardless of the measurement basis. So, to conclude this section let's state that

Theorem 5 (No-communication theorem)**.** *It is impossible to communicate information from one space-like event to another.*

The no-communication theorem states that it is not possible to transmit classical bits of information by means of carefully prepared mixed or pure states, whether entangled or not. The theorem disallows all communication, not just faster-than-light communication, by means of shared quantum states.

3.6 No cloning theorem

Quantum cloning is a process that takes an arbitrary, unknown quantum state and makes an exact copy without altering the original state in any way. The process of quantum cloning is described by:

$$U_c \left|\psi\right\rangle \left|i\right\rangle = \left|\psi\right\rangle \left|\psi\right\rangle \tag{3.37}$$

where U_c is the actual cloning operation, $\left|\psi\right\rangle$ is the state to be cloned, and $\left|i\right\rangle$ is the initial state of the copy.

The question is, is there a unitary operator U_c capable of cloning a general quantum state? Well, sometimes. For an actual cloning, it is only possible to build a cloning gate for orthogonal states.

Proof. Consider two states $\left|\psi\right\rangle$ and $\left|\phi\right\rangle$ together with the copy state $\left|i\right\rangle$. If this general operator exist then: $\left|\Psi\right\rangle = U_c \left|\psi\right\rangle \left|i\right\rangle = \left|\psi\right\rangle \left|\psi\right\rangle$ and $\left|\Phi\right\rangle = U_c \left|\phi\right\rangle \left|i\right\rangle = \left|\phi\right\rangle \left|\phi\right\rangle$. Taking the inner product $\langle\Phi|\Psi\rangle$ we have on one hand $(\langle\phi|\psi\rangle)^2$ but on the other hand $\left\langle i\right| \left\langle\phi\right| U_c^\dagger U \left|\psi\right\rangle \left|i\right\rangle = \langle\phi|\psi\rangle$ so $\langle\phi|\psi\rangle = (\langle\phi|\psi\rangle)^2$. This has only two solutions: either $\langle\phi|\psi\rangle = 1 \Rightarrow \left|\psi\right\rangle = \left|\phi\right\rangle$, the machine only works with the one and only state $\left|\psi\right\rangle$; or $\langle\phi|\psi\rangle = 0$, so cloned states must be orthogonal. □

Notice that this result still allows cloning but it states that it's impossible to construct a general cloning machine that works for any state.

Of course, there are more sophisticate methods that work by doing an approximate copy of the input state up to certain accuracy.

4 Quantum State Discrimination

Fundamental properties of quantum mechanics make it impossible to perfectly distinguish non-orthogonal quantum states. Note that, if state discrimination was perfect, it would imply that quantum cloning could be done perfectly or that quantum entanglement would lead to instantaneous communication [Gisin, 1998]. For instance, the BB84 quantum key distribution scheme [Bennett and Brassard, 2014] is based in sending photons polarised in two non-orthogonal basis, usually the Z and X basis*. Because of the non-orthogonality and the impossibility to distinguish perfectly quantum states, an eavesdropper that intercepts the message without any knowledge on the basis the photons were encoded on, won't be able to read the full sentence. Indeed, if she receives a photon polarised in the X direction but she measures, unconsciously, in the Z direction there will be a 50% chance of mistake. Ultimately, one is forced to make a guess and it is the necessity of this guess that makes quantum mechanics intrinsically indeterministic.

Another example is quantum cloning, if that could be done perfectly then we would be able to generate n copies of two non-orthogonal quantum states $|\psi\rangle$ and $|\phi\rangle$. Since they are not orthogonal, we can't perfectly discriminate a single pair of them. However, if n copies are considered, then the overlap of the composite system goes as $|\langle\psi|\phi\rangle|^n$ which tends to 0 as the number of copies increases. Therefore, because a general quantum state cannot be cloned, state discrimination cannot be done perfectly.

The question now is, how can we *best* discriminate different quantum states? We can't certainly predict the result of a measurement, however, the foundations of quantum mechanics gives us with accuracy the probabilities of those outcomes. These follow some classical probability distribution and with the help of classical information theory we could find ways to distinguish them. The idea is to vary over the measurements that we make on a system to find the one that makes the classical distinguishability the best it can be [Fuchs, 1996].

4.1 On the notion of distance

First of all, it is not possible to go to the Hilbert space, put a ruler between quantum states and decide from this whether they are the same or not, just because a posterior measurement might change its nature. In any case, we can define a pseudo-distance between two general states ρ and ρ' as

$$D(\rho,\rho') = \|\rho-\rho'\|_1 \tag{4.1}$$

which is the so called *trace distance* [Nielsen and Chuang, 2010], denoting by $\|A\|_1$ the *trace norm* (or norm one)

$$\|A\|_1 = \operatorname{tr}\sqrt{AA^\dagger} = \sum_\lambda |a_\lambda| \tag{4.2}$$

where $\{a_\lambda\}$ are the eigenvalues of A.

Properties

- Non-negative: $D(\rho,\rho') \geq 0$, with equality only when $\rho=\rho'$.
- Symmetric: $D(\rho,\rho') = D(\rho',\rho)$.
- Triangle inequality: $D(\rho,\rho') \leq D(\rho,\sigma) + D(\sigma,\rho')$.
- Convexity: $D(\sum_i p_i\rho_i,\sigma) \leq \sum_i p_i D(\rho_i,\sigma_i)$.

Two quantum states are said to be close to each other if the trace distance is near zero. If the states are qubits, with state vector $\boldsymbol{r}$ and $\boldsymbol{r}'$ respectively, the expression (4.1) reduces to

$$D(\rho_{\boldsymbol{r}},\rho_{\boldsymbol{r}'}) = \frac{\|\boldsymbol{r}-\boldsymbol{r}'\|_2}{2} \tag{4.3}$$

where $\|\boldsymbol{a}\|_2 = \sum_k |a_k|^2$ is the usual vector norm (or norm two). Notice that this pseudo-distance is half the ordinary distance between two points inside a sphere.

*Defined as the eigenvalues of the Pauli matrices σ_z and σ_x respectively.

A second definition of pseudo-distance is the *fidelity* F defined for two $\rho, \rho' \in \mathcal{H}$ as

$$F(\rho,\rho') = \left[\operatorname{tr}\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}\right]^2 \tag{4.4}$$

That has a much simpler form if $\rho' = |\phi\rangle\langle\phi|$,

$$F(\rho, |\phi\rangle\langle\phi|) = \langle\phi|\rho|\phi\rangle \tag{4.5}$$

and even simpler if $\rho = |\psi\rangle\langle\psi|$,

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2 \tag{4.6}$$

The previous expression give us an intuitive view of F, it measures the overlap of two states ρ and ρ', in other words, how much information of ρ' is contained in ρ (and vice-versa).

Although it is not clear from the definition, the fidelity is symmetric and is restricted to be in the range $[0,1]$. If $F(\rho,\rho') = 1$ we can say that $\rho = \rho'$ (up to a global phase) but the interpretation of $F(\rho,\rho') = 0$ is not clear, only in the case where both states are pure then $F = 0$ implies that the states are orthogonal.

Properties

- Symmetric: $F(\rho,\rho') = F(\rho',\rho)$.
- Multiplicative under tensor product: $F(\rho\otimes\rho', \sigma\otimes\sigma') = F(\rho,\sigma)F(\rho',\sigma')$.
- Invariant under unitary operation: $F(U\rho U^\dagger, U\rho' U^\dagger) = F(\rho,\rho')$.

The fidelity is sometimes usefull to impose an upper and lower bound on the trace distance as it holds that

$$1 - F(\rho,\rho') \leq D(\rho,\rho') \leq \sqrt{1 - F^2(\rho,\rho')} \tag{4.7}$$

Even though the trace distance is very useful from the mathematical perspective, it doesn't provide us with a clear intuition of the quantity of information, this is when the concept of Shannon entropy comes in. Consider a set of events $X = (p_k, x_k)_{k=1}^n$, then the minimal amount of bits that are needed to encode X is given by the Shannon entropy [Nielsen and Chuang, 2010]

$$H(X) = -\sum_{k=1}^{n} p_k \log_2 p_k \tag{4.8}$$

At the quantum level, this definition is replaced by the Von-Neuman entropy $S(\rho)$. Again, consider an ensemble of pure states $\Xi = \{(\xi_k, |\psi_k\rangle)\}$ with density matrix ρ_k, then the Von-Neuman entropy is *Von-Neuman entropy*

$$S(\rho) \equiv -\operatorname{tr}[\rho\log_2\rho] = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{4.9}$$

The second identity is true when $p = \{p_1, \ldots, p_n\}$ are the eigenvalues of ρ but they need not be equal to $\{\xi_1, \ldots, \xi_n$, this will be the case only when the states in the ensemble Ξ are mutually orthogonal.

Then, the amount of information shared between X and Y is

$$I(X:Y) = H(X) + H(Y) - H(X,Y) \tag{4.10}$$

where $H(X,Y) = -\sum_{x,y} p(x,y)\log_2 p(x,y)$.

The most interesting result is that the mutual information I has an upper bound that restricts the amount of information that two quantum states can share. This result was first proved by Holevo [1973b], although a simple proof has been done by Fuchs and Caves [1994]. Here, we will just state the main result in the form of a theorem:

Theorem 6 (Holevo bound)**.** *Suppose Alice prepares a state ρ_X where $X = 1, \dots, n$ with probability ξ_X. Bob performs a measurement described by POVM elements $\{\Pi_y\}$ on that state with measurement outcome Y. The Holevo bound states that for any such measurement Bob may do*

$$I(X:Y) \leq S(\rho) - \sum_x \xi_x S(\rho_x) \tag{4.11}$$

where $\rho = \sum_x \xi_x \rho_x$.

The right hand side of eq. (4.11) is usually called the Holevo χ quantity and it is easily proved [see Nielsen and Chuang, 2010, sec 11.3.2] to be strictly less than $H(X)$, the mutual information is strictly smaller than the initial information, i.e. $I(X:Y) < H(X)$. This is a milestone result because it tells us that no matter which measure we do on the system, we will never be capable of completely determining X from the results Y, the information after the measurement is always reduced. However, in this thesis, we will try to get close to this limiting value.

In fact, all the subsequent work could be done using the mutual information instead of the success probability as a measure to distinguish two state. Fuchs [1996], in his thesis, does this with 5 functions that measure in some way the distinguishability, including the mutual information. This is just to remark that there is nothing special in using one or another, just that some are simpler than others from the mathematical point of view.

4.2 Quantum Hypothesis Testing

Consider that Alice prepares one state of some ensemble $\Xi = \{\xi_k, \rho_k\}_{k=1}^n$, all living in a d-dimensional Hilbert space $\mathcal{H}_d$. The probability that Alice chooses ρ_k is ξ_k, with $\sum_k \xi_k = 1$. After that, this state is send to Bob who is asked to distinguish it among the states inside the set Ξ. In Bob's hands, the system is described by the mixed state

$$\rho = \sum_{k=1}^n \xi_k \rho_k \tag{4.12}$$

Bob can perform any measurement on the state, the most general form of such measurement is a POVM measure $\mathcal{M} = \{\Pi_j\}_{j=1}^m$ satisfying eqs. (3.6a) and (3.6b). Note that, m is not in general equal to n (the number of states), but it can be greater or smaller. This number is related to the number of hypothesis that can be made. For instance, if $m > n$ then we can assign to a combination of multiple outcomes of our measurement the same $\rho_k \in \Xi$. On the other hand, if $m < n$ then there will be some states for which we will have to make a guess, unless we know that they occur with 0 probability.

If the Π_j are orthogonal projectors ($\Pi_i \Pi_j = \delta_{ij} \Pi_i$), then $\mathcal{M}$ is a von Neumann measure and $m \leq n$, but they do not have to be. As an example, the operators $\Pi_j = \mathbb{I}/d$ associated to the no-measurement strategy are clearly not projectors.

Actually, it is found by Davies [1978] that the number of POVM elements m needed to distinguish n pure states is bounded between $n \leq m \leq n^2$ for linearly independent states. The number of POVM operators can be any inside this range but the process becomes an arduous task if the optimisation needs to be made also on the number of hypothesis. For simplicity, in our problem, we will fix $m = n$, i.e. the number of hypothesis is the same as the number of states, where j is the proposition that the state was ρ_j. We can do this because, even if m was greater than the number of states, we could group the operators from our hypothesis to build only n operators verifying eqs. (3.6a) and (3.6b).

The probability of outcome j (Π_j) conditional that the given state was ρ_k is

$$p(j \mid k) = p(\mathcal{M} = \Pi_j \mid \Xi = \rho_k) = \mathrm{tr}(\Pi_j \rho_k) \tag{4.13}$$

Therefore, the state k will be successfully identified whenever the hypothesis Π_k is selected which happens with probability $p(k \mid k)$. Putting all together, it follows that the probability of correctly guessing the state is

$$P_s = \sum_{k=1}^n \xi_k \,\mathrm{tr}(\Pi_k \rho_k) \tag{4.14}$$

and because the states Ξ are not mutually orthogonal, there will be non-zero probability of failure (measure of an incorrect state): $0 \leq P_s \leq 1$. The expression for the error probability is just $P_e = 1 - P_s$.

4.3 Optimality conditions

In general, a measure $\mathcal{M}$ will give us some success probability (4.14) which will be suboptimal. We seek to find the POVM that maximises the success probability. It has been found by Holevo [1973a] that the optimal operators must satisfy the conditions

$$\Pi_j(\xi_j\rho_j - \xi_k\rho_k)\Pi_k = 0 \qquad \forall j,k = 1,\dots,m \tag{4.15}$$

$$\Gamma - \xi_k\rho_k \geq 0 \qquad \forall k = 1,\dots,N \tag{4.16}$$

with the definition of the so called Lagrange operator

$$\Gamma = \sum_{k=1}^{n} \xi_k \Pi_k \rho_k \tag{4.17}$$

which places the role of a Lagrange multiplier taking account of the constraint (3.6a). It can be shown from the first condition that the Lagrange operator is hermitian. Take the sum over j and k in eq. (4.15), because $\Pi_j = \Pi_j^\dagger$ and $\rho_k = \rho_k^\dagger$, we are left with $\Gamma^\dagger - \Gamma = 0$ proving the hermiticity of the Lagrange operator. Indeed, eqs. (4.15) and (4.16) are not independent but the first can be derived from the second.

In fact, the first condition (4.15) can also be written, by summing over j, in terms of the Lagrange operator Γ as

$$(\Gamma - \xi_k\rho_k)\Pi_k = 0 \qquad \forall k = 1,\dots,m \tag{4.18}$$

which gives us a way to determine the operators Π_k once Γ is known. Indeed, both Π_k and $\Gamma - \xi_k\rho_k$ are positive operators, and thus eq. (4.18) can hold only if they are orthogonal, that is Π_k lays entirely within the kernel of $\Gamma - \xi_k\rho_k$ [Weir et al., 2017].

Equation (4.16) gives a necessary and sufficient condition for an optimal measurement, while eq. (4.15) gives only a necessary condition. In our posterior work, we will seek to find such measurement whose Lagrange operator Γ (4.17) is hermitian and for all the initial states in the ensemble Ξ we have that the second Holevo condition is verified.

4.4 Ambiguous vs. unambiguous state discrimination

In the problem of quantum state discrimination, there are two major techniques: minimum error discrimination (MED) and unambiguous state discrimination (USD). The former approach, also named ambiguous state discrimination, consist on minimising the probability of guessing a wrong result P_e, which can sometimes be achieved by not making any measurement at all and randomly guessing the result. In contrast, the latter has no error, if hypothesis Π_j is obtained we are 100% sure of that the state was ρ_j, yet we allow the possibility of an inconclusive result by introducing an extra operator $\Pi_?$. The two tasks are equally valid, the use of one or another only depends on the requirements of the problem. For example, in situations where we can't be wrong we should use USD instead of MED. In fact, there is a correspondence between both as it is possible to take a MED to a USD [Bagan et al., 2012].

Unambiguous state discrimination forces the operators to satisfy $\mathrm{tr}(\Pi_j\rho_k) = 0$ if $j \neq k$, this is not possible in general since both Π_j and ρ_k are positive operators. Only when the states $\{\rho_k\}_{k=1}^{n}$ have disjoint kernels, $\ker(\rho_k) \cap \ker(\rho_l) = \emptyset \; \forall k \neq l$, USD would be possible [Raynal, 2006; Rudolph et al., 2003]. Therefore, in what follows, we will be working in the context of MED which now proceed to explain in more detail.

In Section 4.2, the form of the success probability was deduced. It follows that the probability of error is $P_e = 1 - P_s$, so finding the minimum P_e is the same as maximising the success probability as a function of the measure,

$$P_s = \max_{\mathcal{M}} \sum_{k=1}^{n} \xi_k \, \mathrm{tr}(\Pi_k\rho_k) \tag{4.19}$$

under the conditions eqs. (3.6a) and (3.6b). Putting the sum inside the trace, we identify the operator to be maximised as Γ and from eq. (4.16) we can rewrite the problem as that of finding

$$P_s = \min_{\Gamma} \operatorname{tr} \Gamma \tag{4.20}$$

subject to the constraints $\Gamma - \xi_k \rho_k \geq 0$.

The meaning of this is that, for an arbitrary set of positive observables $\{\Pi_j\}$ that add up to the identity, we can construct the corresponding Lagrange operator. However, only the one which is optimal according to the relation eq. (4.16) will give the maximum probability. Even though two measures $\mathcal{M}$ and $\mathcal{M}'$, with Γ and Γ' respectively, are found to be optimal, the success probability will still be the same [Helstrom, 1969].

Equation (4.19) may look like a *tour de force* to the reader, having to maximise over all the possible measures. It happens that this is as complicated as it seems, very few analytical solutions are found while most of the results in quantum discrimination problems are found using numerical methods like Semi-Definite Programming. The analytical solutions are only well known for the case of discrimination between two states or for geometrically uniform states. For example, the case of two states was first found by Helstrom who provided an exact value for the success probability [Helstrom, 1969] which we will reproduce in the following section. Then, Bae and Kwek [2015]; Barnett [2001] showed a minimum-error discrimination strategy between multiply symmetric states with a deeper study of the so called three mirror-symmetric states [Andersson et al., 2002; Chou, 2004; Ha and Kwon, 2013]. For a general number of states, there are unambiguous strategies found by Chefles and Barnett [1998] when the states are linearly independent and for minimum error discrimination, it is found that the discrimination between n qubit states can be divided into patches of only 4 qubits with a known optimal solution [Weir et al., 2017]. Also, Deconinck and Terhal [2010] provide a geometrical representation of the optimal measure in the Bloch sphere.

4.5 Two-state discrimination

It is instructive to work out the solution to the simplest problem in QSD following the process explained previously. We will evaluate the maximum success probability for the case of two general states $\Xi = \{(\xi_1, \rho_1), (\xi_2, \rho_2)\}$ and then give some simplified versions for when the states are pure, qubits*... The measure is made up of only two positive operators $\{\Pi_1, \Pi_2\}$ that satisfy $\Pi_1 + \Pi_2 = \mathbb{I}$. The maximum guess probability is given by eq. (4.20) where the Lagrange operator is

$$\Gamma = \xi_1 \Pi_1 \rho_1 + \xi_2 \Pi_2 \rho_2$$

but using the completeness relation, the dependence in one of the operators can be removed. Write

$$\begin{cases} \Gamma_+ & = \xi_2 \rho_2 + \Pi_1 X \\ \Gamma_- & = \xi_1 \rho_1 - \Pi_2 X \end{cases}$$

defining

$$X = \xi_1 \rho_1 - \xi_2 \rho_2 \tag{4.21}$$

Although the process can be done with Γ_+ or Γ_-, it is convenient to symmetrise those expressions and write the Lagrange operator for the problem as

$$\Gamma = \frac{1}{2}(\Gamma_+ + \Gamma_-) = \frac{1}{2}\left(\rho + \Lambda X\right) \tag{4.22}$$

where $\rho = \xi_1 \rho_1 + \xi_2 \rho_2$ and $\Lambda = \Pi_1 - \Pi_2$. The original POVM operators are related to Λ by $\Pi_1 = (\mathbb{I} + \Pi)/2$ and $\Pi_2 = (\mathbb{I} - \Pi)/2$; while $\Pi_1, \Pi_2 \geq 0$ the condition over Λ is that $-\mathbb{I} \leq \Lambda \leq \mathbb{I}$.

Putting all together, the success probability becomes

$$P_s = \max_{\Pi} \operatorname{tr} \Gamma = \frac{1}{2}\left(1 + \max_{\Lambda} \operatorname{tr} \Lambda X\right) \tag{4.23}$$

*The following is not restricted to two dimensional spaces but is general to any two level system in a $\mathcal{H}_d$.

From the definition of X, because ρ_1 and ρ_2 are positive, its eigenvalues can be divided into positive and negative parts. Denoting by X_+ (X_-) the subspace spanned by the eigenspace of positive (negative) eigenvalues and λ_+ (λ_-, in absolute value) its sum, by the spectral theorem X reads $X = \lambda_+ X_+ - \lambda_- X_-$ where $X_+, X_- \geq 0$. Thus, the optimal measurement Λ is the one that projects the positive subspace to itself and flips the sign of the negative part, i.e. $\Lambda = X_+ - X_-$. Finally, the success probability is [Bae and Kwek, 2015]

$$P_s = \frac{1}{2}(1 + \lambda_+ + \lambda_-) = \frac{1}{2} + \frac{1}{2}\|X\|_1 \tag{4.24}$$

and the POVM consists on

$$\mathcal{M} = \{\Pi_1 = X_+,\ \Pi_2 = X_-\} \tag{4.25}$$

where we have used that $X_+ + X_- = \mathbb{I}$. Equation (4.24) is known as the Helstrom bound and establishes the best success probability to discriminate two mixed states [Helstrom, 1969] which depends only on the trace distance between the two.

It is easily checked that this measure is indeed optimal by constructing the Lagrange operator from eq. (4.22) using the measure found in eq. (4.25), it follows that

$$\Gamma = \frac{1}{2}[\rho + (X_+ - X_-)X] = \frac{1}{2}\rho + \frac{1}{2}(\lambda_+ X_+ + \lambda_- X_-) \tag{4.26}$$

Then, for the two states in Ξ the Holevo condition reads

$$\begin{aligned}\Gamma - \xi_1\rho_1 &= \frac{1}{2}(-\xi_1\rho_1 + \xi_2\rho_2) + \frac{1}{2}\Lambda X = -\frac{1}{2}X + \frac{1}{2}\Lambda X = \lambda_- X_- \geq 0 \\ \Gamma - \xi_2\rho_2 &= \frac{1}{2}(\xi_1\rho_1 - \xi_2\rho_2) + \frac{1}{2}\Lambda X = \frac{1}{2}X + \frac{1}{2}\Lambda X = \lambda_+ X_+ \geq 0\end{aligned}$$

Since the Holevo conditions are satisfied, we can be sure that the measure (4.25) is optimal.

We should be careful with the previous result (4.25) since there may be cases where all eigenvalues are positive or negative if the *a priori* probabilities are different. Then, one of the eigenspaces will be the full space, in fact, it will correspond to the hypothesis of the state with maximum probability. The result is telling us not to waste any effort at all in measuring because we have enough information beforehand to achieve the maximum success probability by just guessing the state with maximum probability.

Of course, eq. (4.24) is much simplified when specific cases are considered. For example, if the states have *a priori* equal probabilities $\xi_1 = \xi_2 = 1/2$, the success probability is

$$P_s = \frac{1}{2} + \frac{1}{4}\|\rho_1 - \rho_2\|_1 \tag{4.27}$$

For qubits with state vector $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ respectively

$$P_s = \frac{1}{2} + \frac{1}{4}|p_1 - p_2 + \|\xi_1\boldsymbol{r}_1 - \xi_2\boldsymbol{r}_2\|_2| + \frac{1}{4}|p_1 - p_2 - \|\xi_1\boldsymbol{r}_1 - \xi_2\boldsymbol{r}_2\|_2| \tag{4.28}$$

which, for the case of equal *a priori* probabilities, reduces to

$$P_s = \frac{1}{2} + \frac{1}{2}\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|_2 \tag{4.29}$$

Finally, if ρ_1 and ρ_2 are pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ then [Barnett and Croke, 2009]

$$P_s = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\xi_1\xi_2|\langle\psi_1|\psi_2\rangle|^2} \tag{4.30}$$

All of the above expressions contain a constant term, which doesn't depend at all of the states, and another that depends on the difference between them. Thus, whenever they are the same, i.e. X is the 0 matrix, the probability of success will be just 1/2 which is to just pick one of the two possible hypothesis at random. In any other situation, the probability will increase, up to its maximum value.

5 Quantum computing

Now that we know the basic formalism of quantum mechanics let's return to the very first question we have asked: what problems are computable? From the Church-Turing principle, every physically reasonable process can be computed by a Turing machine. Yet, from a pragmatic perspective, there is little difference between such non-computable problems and those that would take more than the length of several human lifetimes, or perhaps the lifetime of the universe, to solve.

5.1 P vs. NP

Gregory Chaitin

> *Computer scientists widely believe that $P \neq NP$, but no proof is known. One could say that a lot of quasi-empirical evidence points to P not being equal to NP. Should $P \neq NP$ be adopted as an axiom, then? In effect, this is what the computer science community has done.*

We can classify each of the problems experienced in various scientific disciplines or every day life as either easy (polynomial, P) or hard (non polynomial, NP). To classify each problem we must analyse how the time scales with the length of the input, for example:

P Multiplication is considered to be an easy problem because the time scales proportional to the length of the input.

NP Factoring is considered as a difficult problem as it scales exponentially fast with the length. This is the reason why most cryptography algorithms are base on factorisation of large numbers that would take the length of the universe to compute.

Let's modify a new principle to take this into account.

Principle 8 (Strong Church-Turing)**.** *Any reasonable model of computation can be efficiently simulated on a Turing machine.*

Should this thesis be true, then any task that has no efficient solution on a Turing machine will have no efficient solution in any reasonable model of computation, and is thus guaranteed to be hard.

For this we define the complexity class P as the set of problems that can be solved efficiently by a Turing machine. Problems in P are guaranteed to be easy, while problems proven to lie outside P could be hard, conditioned on the truth of the Strong Church-Turing principle (SCT). Yet, even this is too optimistic, the reduction of an arbitrary model of computation to a Turing machine is highly non-trivial. There exists many physically reasonable models of computation whose computational power remains largely unknown.

For example, it remains an open question whether the addition of a random number generator to an Turing machine bestows it with greater computational power. Polynomial identity testing (determining whether a given polynomial is identically 0) has no known efficient solution on Turing machines, and yet can be solved in polynomial time on its probabilistic extension. This motivates the introduction of the complexity class BPP (bounded-error, probabilistic, polynomial time), the class of problems that can be efficiently solved by probabilistic Turing machines.

Principle 9 (Probabilistic Strong Church-Turing)**.** *Any reasonable model of computation can be efficiently simulated on a probabilistic Turing machine.*

With the introduction of quantum computing a new class of problems have emerged into a new complexity class called BQP (bounded quantum polynomial) which represents the class of problems that have efficient solutions when quantum processes are also permitted. For example, factoring in a quantum computer can be efficiently solved in order $\mathcal{O}(n^3)$ using Shor's algorithm.

Principle 10 (Quantum Strong Church-Turing)**.** *Any reasonable model of computation can be efficiently simulated on a quantum computer.*

5.2 Classical vs. Quantum computation

In classical models of computation, the intrinsic state of a physical system is synonymous with the observable properties of a system. A string of n bits is the abstraction of a physical system with 2^n possible configurations. In particular, the amount of information that may be extracted by measurement from the system coincides with the amount of information needed to precisely define the state of the system (namely n bits).

In contrast, the amount of information needed to precisely define a quantum system far exceeds that of which can be extracted by measurement. Mathematical representation of the fundamental unit of quantum information requires more than a single bit while any measurement can only extract 1 single qubit. To properly define a state on a system on n qubits, such state requires $2^n - 2$ independent parameters (note the normalisation condition and the global phase). This leads to the important observation that the amount of information stored within a quantum system scales exponentially with respect to both (a) the size of the system, and (b) amount of information retrievable by measurement.

(a) There must be states that cannot be expressed of its individual constituents, *entangled states*, this suggests the potential exponential speed-up of the quantum processing. Consider comparing the action of flipping the first bit on a bit-string of length n and its quantum analogue: in the classical system, this action transforms exactly a single bit of information while in a quantum system 2^{n-1} flips need to be made.

   Proof. Think of it, a general n-qubit can be represented as $|\psi\rangle = \frac{1}{2^{n/2}}\sum_{k_1,\ldots,k_n=0}^{1}|k_1\cdots k_n\rangle$. If the first qubit is flipped then we will have $|\psi\rangle = \frac{1}{2^{n/2}}\sum_{k_1,\ldots,k_n=0}^{1}\left|\bar{k}_1\cdots k_n\right\rangle$ which means that $2^n/2$ qubits have changed. □

(b) While an exponential amount of information is being manipulated during a general quantum process, only a linear amount of bits may be extracted by measurement.

A classical computer operates using boolean functions, for example on n bits $f_n : \{0,1\}^n \to \{0,1\}^n$. A computation is specified by a uniform family of such functions $\{f_n\}$. Analogously, any quantum process on n qubits is specified by a mapping $U_n \in SU(2^n)$ between two arbitrary quantum states of the system. A quantum computation is specified by a uniform family of such operators $\{U_n\}$. To link the two formalisms observe that any string of classical bits $b_1b_2\ldots$ can be encoded within the basis state $|b_1b_2\ldots\rangle$. Also, any f_n may be recast as the action of some unitary operator on a suitable quantum system.

Any model of computation that allows the synthesis of an arbitrary unitary operator U_n is universal. In fact, it is sufficient to synthesise a unitary U_ϵ that is a close approximation to the desired U.

Theorem 7. *A model of quantum computation is universal if for any fixed $\epsilon > 0$ and any desired unitary U it is capable of synthesising a U_ϵ such that $\|U_\epsilon - U\| < \epsilon$.*

For each unitary U we can associate a real number $M_G(U,\epsilon)$ (gate complexity) that is the minimum number of gates from set G required to synthesise U to accuracy ϵ. If $M_G(U_n,\epsilon) < n^d$ for some power d implies that $\{U_n\}$ has an efficient solution.

Finally note that, from the nature of quantum mechanics, all unitary operators are reversible: by knowing the outputs we can exactly determine its inputs. In contrast, classical computation is not reversible just take a look at the action of the AND gate: if the output is 1 then we know for sure that the inputs were 1 and 1 but if the output is zero there are 3 possibilities for the inputs. The implementation of the AND gate in quantum computing is made through the *Toffoli gate* with 3 inputs and 3 outputs: $(a,b,c) \to (a,b,c \oplus ab)$. For example, if c is set to 1 then the output $\forall a,b$ is $(a,b,1\oplus ab) = (a,b,\neg(ab))$ which is the classical NAND gate.

5.3 Universal Quantum Gates

In Section 2 we saw what classical gates constitute a set of universal gates, this is, any arbitrary function can be build upon them. Our desire is to find what quantum gates are universal for quantum computation. First of all, let's announce the following theorem that will help us reduce the possibilities [see Nielsen and Chuang, 2010, Section 4.5].

Theorem 8. *Any unitary operation U acting on n qubits can be decomposed into $\tilde{U}_1, \dots, \tilde{U}_m$ gates acting non-trivially in two-dimensional subsystems.*

The previous theorem is telling us that, the effect of U on a qubit $|\psi\rangle \in \mathcal{H}_n$ is the same as the application of $\tilde{U}_1\tilde{U}_2 \cdots \tilde{U}_m$ on the same state $|\psi\rangle$. Therefore, any universal gate can be expressed as a product of singe or two qubit gates. Single qubit gates are all rotations in the Bloch sphere and two qubit gates can be a controlled operation on a single qubit like the control-not gate eq. (3.24) or swap gate (3.25).

The standard set of universal quantum operations contain $\{H, T, C_{NOT}\}$. We already saw that H performed a rotation of $\pi/2$ about the y-axis and T a rotation of $\pi/4$ around the z-axis. The rotations around the x-axis are accomplished with HTH that rotates the state $\pi/4$ around this axis. Therefore, with just the two gates we can perform all rotations and thus, al unitary operations on single qubits. The C_{NOT} is important as it enables to construct entangled states but any other two-qubit gate can be constructed from it by performing some rotation. For instance, we can obtain the controlled-phase gate by $(\mathbb{I} \otimes H)C_{NOT}(\mathbb{I} \otimes H)$.

Obviously, taking the original gate to a sequence of this gates can be a tough work in most situations. In general, a circuit on n qubits requires $\mathcal{O}(n^2 4^n)$ elementary universal gates.

How to construct other gates? It is important to realise that any gate can be expressed in terms of the universal gates $\{H, T, C_{NOT}\}$. Essentially, we need to find a way to simulate the Pauli gates $\{X, Y, Z\}$ with just these gates, since then any unitary can be expressed as a complex combination of Pauli matrices and in turn, as a complex combination of a universal set of gates.

First of all, it is easily seen that the Z gate can be realised as

$$Z = T^4 \tag{5.1}$$

Then, using that $X = HZH$ we can express X as

$$X = HT^4H \tag{5.2}$$

Finally, for Y we use (3.8) which gives

$$Y = iXZ = iHT^4HT^4 \tag{5.3}$$

equal up to a global phase which play no role in the final computation.

The previous represent rotations of π degrees with respect to the z, x and y axis respectively but we would like to realise a more general kind of rotations as defined in (3.11). This can be done by noting that

$$THTH = R_{\boldsymbol{n}}(\theta) \qquad , \qquad \theta = 2\arccos\cos[2](\pi/8) \approx 0.1744\cdots \times 2\pi \tag{5.4}$$

where $\boldsymbol{n} = (\cos(\pi/8), \sin(\pi/8), \cos(\pi/8))$. The key part is that θ is an irrational multiple of 2π, therefore by applying k times the previous gate we can realise a rotation about $\boldsymbol{n}$ up to the accuracy that we want. That is, suppose that we want to rotate an angle β then, using that $R_{\boldsymbol{n}}(\theta)^k = R_{\boldsymbol{n}}(k\theta)$, we require $|\beta - k\theta| < \epsilon$.

Other combinations of H and T generate rotations along different axis and those, in combination with the Pauli matrices span all the space of rotations.

Another important issue are the two qubit gates but all of them can be expressed as a product of C_{NOT} and single qubit gates. For instance, the C_Z and $SWAP$ gates ((3.25)) are realised using

$$C_Z = (\mathbb{I} \otimes H)C_{NOT}^{ij}(\mathbb{I} \otimes H) \tag{5.5}$$

$$SWAP = C_{NOT}^{ij}C_{NOT}^{ji}C_{NOT}^{ij} \tag{5.6}$$

where i is the control qubit and j the target qubit in C_{NOT}^{ij}.

6 Quantum algorithms

In this section we will see some implementations of quantum algorithms, showing the quantum circuit and the state after each step in the algorithm.

6.1 Teleportation

With the aid of entanglement, our aim is to send a qubit state $|\phi\rangle = \alpha |0\rangle + \beta |1\rangle$ from Alice to Bob. To do so we will demand that Alice and Bob share an entangled state, for instance $|\Phi^+\rangle$. Then, the steps are

1. Alice has in her possession the state $|\phi\rangle$ and part of the state $|\Phi^+\rangle$. She applies a CNOT to her part of the state, followed by a Hadamard gate on the first qubit. The state after the gates is

$$\frac{1}{2}|00\rangle (a_0 |0\rangle + a_1 |1\rangle) + \frac{1}{2}|01\rangle (a_0 |1\rangle + a_1 |0\rangle) + \frac{1}{2}|10\rangle (a_0 |0\rangle - a_1 |1\rangle) + \frac{1}{2}|11\rangle (a_0 |1\rangle - a_1 |0\rangle)$$

2. Alice measures in the computational basis and collapses the state to one of the four possibilities show above.

3. Alice shares with Bob the result of the measurement, who then applies a rectifying gate to completely recover the state:

00	$\mathbb{I}$
01	X
10	Z
11	Y

One may wonder if this algorithm breaks the law of relativity as the collapse of the state happens *instantaneously*. However, the answer is no, because even the state has collapsed, the information that allows Bob to recover the state is send trough a classical channel and thus, at most, travels at the speed of light.

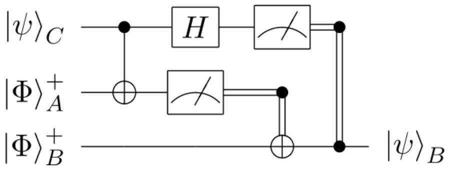

Figure 4: Quantum teleportation circuit.

6.2 Dense coding

We want to show that it is possible to send 2 bits of information using just 1 qubit. To do so, we will start as in the teleportation algorithm. Alice and Bob share an entangled state $|\Phi^+\rangle$, then she applies some gates in her state in order to encode the two bits:

$$\begin{aligned}
00 &: \left|\Phi^+\right\rangle \\
01 &: \left|\Psi^+\right\rangle = (\sigma_x \otimes \mathbb{I}) \left|\Phi^+\right\rangle \\
10 &: \left|\Psi^-\right\rangle = (\sigma_y \otimes \mathbb{I}) \left|\Phi^+\right\rangle \\
11 &: \left|\Phi^-\right\rangle = (\sigma_z \otimes \mathbb{I}) \left|\Phi^+\right\rangle
\end{aligned}$$

These four states are orthogonal so we can distinguish them perfectly, the only downside is that either we must bring the qubits together and apply a measurement in the Bell basis. In this way, Bob finally knows the resulting bits.

6.3 Quantum parallelism

Quantum parallelism is a fundamental feature of many quantum algorithms, it allows quantum computers to evaluate a function $f(x)$ for many different values of x simultaneously.

Suppose $f(x) : \{0, 1\} \to \{0, 1\}$ is a classical one bit function, a convenient way of computing this function is to consider a two qubit computer at state $|a, b\rangle$ to transform it into $|a, b \oplus f(a)\rangle$ after some U_f. The first bit (register) is called the *data* register and the second the *target* register. More formally, $U_f\,|a, b\rangle = |a, b \otimes f(a)\rangle$ so when $b = 0$ the second qubit gives exactly the value of $f(a)$ over the first one.

But think of it, this is in general so the first qubit can be in fact a superposition like $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and if b is set to $|0\rangle$ then

$$U_f\,|+\rangle\,|0\rangle = \frac{1}{\sqrt{2}}\left(U_f\,|0, 0\rangle + U_f\,|1, 0\rangle\right) = \frac{1}{\sqrt{2}}\left(|0, f(0)\rangle + |1, f(1)\rangle\right) \tag{6.1}$$

This is a remarkable state! The different terms contain information about both $f(0)$ and $f(1)$; it is almost as if we have evaluated $f(x)$ for two values of x simultaneously, a feature known as "quantum parallelism".

Let's see the general case of n qubits, then the data register can be expressed as

$$|X_n\rangle = \frac{1}{2^{n/2}}\sum_x |x\rangle \in \mathcal{H}^{\otimes n} \tag{6.2}$$

where the sum is over all the possible permutations of n bits. A simple way to obtain this state is by applying a Hadamard transform on each qubit from the initial state $|0\rangle^{\otimes n}$ which produces an equal superposition of all computational basis states. With only two qubits this is easily seen, starting with $|00\rangle$ we apply $H \otimes H$ obtaining

$$|X_2\rangle = (H \otimes H)(|0\rangle \otimes |0\rangle) = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) = \frac{1}{2}\left(|00\rangle + |01\rangle + |10\rangle + |11\rangle\right) = \frac{1}{2}\sum_{i,j=0}^{1} |ij\rangle \tag{6.3}$$

Preparing the $n{+}1$ state $|0\rangle^{\otimes n}\,|0\rangle$, after the Hadamard transform has been applied, the quantum gate U_f can be computed giving the state

$$\frac{1}{2^{n/2}}\sum_x |x\rangle\,|f(x)\rangle \tag{6.4}$$

In some sense, quantum parallelism enables all possible values of the function f to be evaluated simultaneously, even though we apparently only evaluated f once.

However, this is not still usefull because a measurement of one of the values will produce a collapsing of all states into one loosing the rest of the information, which is what a classical computer can already do. Quantum computation requires something more than just quantum parallelism to be useful; it requires the ability to extract information about more than one value of $f(x)$ from superposition states.

6.4 Deutsch-Josza algorithm

Bob gives Alice a function $f(x)$ that maps a n bit number onto $\{0, 1\}$ and promises Alice that it is either constant (same value for all x) or balanced (0 for exactly half of the inputs). Alice is asked to decide whether the function f is constant or balanced with certainty and with the minimum number of evaluations.

Classically, Alice may have to evaluate $f(x)$ at least $2^n/2 + 1$ times to know exactly the answer since she may get $2^n/2$ 0s before finally getting a 1. This is the best a deterministic classical algorithm can do.

From the quantum information perspective we can make use of quantum parallelism, let's see step by step what happens. Alice start with the state

$$|\psi_0\rangle = |0\rangle^{\otimes n}\,|1\rangle \tag{6.5}$$

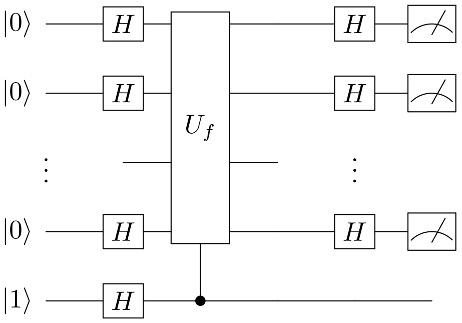

Figure 5: Quantum circuit for the Deutsch-Josza algorithm

and applies a Hadamard transform on all $n+1$ qubits to obtain

$$|\psi_1\rangle = \frac{1}{2^{n/2}} \sum_x |x\rangle |-\rangle \tag{6.6}$$

Next, the function f is evaluated using $U_f : |a, b\rangle \to |a, b \oplus f(a)\rangle$. On a single state $|x\rangle |-\rangle$ this gives $|x\rangle (|f(x)\rangle - |\bar{f}(x)\rangle)/\sqrt{2}$, both $f(x)$ and $\bar{f}(x)$ are evaluated. Consider $f(x) = 0$ then the resulting state is in fact $|x\rangle |-\rangle$ while if $f(x) = 1$ then the state is $-|x\rangle |-\rangle$ so the action of the operator U_f can be expressed as $U_f |x\rangle |-\rangle = (-1)^{f(x)} |x\rangle |-\rangle$. Thus, after the application of the function the state is

$$|\psi_2\rangle = \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} (-1)^{f(k)} |k\rangle |-\rangle \tag{6.7}$$

Finally, wit the given state, let's calculate the probability of finding $|+\rangle^{\otimes n}$ in this superposition:

$$\begin{aligned}
p(+\cdots+) &= |\langle +\cdots+|\psi_2\rangle|^2 \\
&= \left| \frac{1}{2^n/2} \sum_{k_1,\ldots,k_n=0}^{1} (-1)^{f(x)} \langle +\cdots+|k_1 k_2 \cdots k_n\rangle \right|^2 \\
&= \left| \frac{1}{2^n/2} \sum_{k_1,\ldots,k_n=0}^{1} (-1)^{f(x)} \prod_{m=1}^{n} \langle +|k_m\rangle \right|^2 \\
&= \left| \frac{1}{2^n/2} \sum_{k_1,\ldots,k_n=0}^{1} (-1)^{f(x)} \frac{1}{2^{n/2}} \right|^2 \\
&= \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} (-1)^{f(k)} \right|^2
\end{aligned}$$

Now, if $f(x)$ is constant then $p(+\cdots+) = 1$ in either case ($f(x) = 0$ or $f(x) = 1$) but if $f(x)$ is balanced then the terms in the sum will cancel between each other and $p(+\cdots+) = 0$.

Alice declares $f(x)$ to be constant if she measures all n qubits to be in $|+\rangle$, otherwise she declares $f(x)$ to be balanced with only 1 evaluation of the function f compared to the exponential requirement of the classical algorithm.

If our quantum computer only allows measurements in the Z basis (as most do) we need an extra step in our algorithm as shown in fig. 5. After the application of the unitary we perform a Hadamard gate in all the states of the first register which will take them again to the Z basis. After this step, the state will be

- If the function is constant, the factor $(-1)^{f(x)}$ is a global phase and the state after the $H^{\otimes n}$ is just the initial state plus this global phase

$$|\psi_2, \mathrm{c}\rangle = (-1)^{f(x)} |0 \ldots 0\rangle |1\rangle \tag{6.8}$$

- If the function is balanced, the phase factor $(-1)^{f(x)}$ will create a superposition a complicated superposition which after the application of the Hadamard gate look like

$$|\psi_2, \mathrm{b}\rangle = 2^{-n} \sum_{x=0}^{2^n-1} (-1)^{f(x)} \sum_{y=0}^{2^n-1} (-1)^{x\cdot y} |y\rangle |1\rangle = 2^{-n} \sum_{y=0}^{2^n-1} \left[\sum_{x=0}^{2^n-1} (-1)^{f(x)+x\cdot y} \right] |y\rangle |1\rangle \qquad (6.9)$$

  The important thing here is that this state is orthogonal to the state $|0\dots 0\rangle$,

$$\langle 0\dots 0|\psi_2, \mathrm{b}\rangle = 2^{-n} \sum_{x=0}^{2^n-1} (-1)^{f(x)} = 0$$

  because the sum takes half of the times the value 1 and half of the times the value -1, so in total it averages to 0.

Therefore, if the results of the measurement are all $+1$ (corresponding to $|0\rangle$) we are certain that the function is constant but if a -1 is obtained (theoretically, half of them should be -1) then the function is balanced.

6.5 Bernstein-Varizani problem

It's a restricted version of the Deutsch–Jozsa algorithm where instead of distinguishing between two different classes of functions, it tries to learn a string encoded in a function.

Suppose we are given a function $f : \{0,1\}^n \to \{0,1\}^n$ and a secret string $s \in \{0,1\}^n$ such that $f(x) = x \cdot s \mod 2$. Our goal is to determine s with the minimum number of operations.

A classical analysis implies that we should call the function at least n times, each time with an sequence that is linearly independent from the previous ones. for instance, the simplest way is to evaluate the function in the sequences $0\dots01, 0\dots10$,...,$10\dots0$ obtaining the value of one bit in s directly after each evaluation. This is the same as determining the components of the vector by multiplying by the elements of the canonical base.

Obviously, quantum parallelism allows us to find s with just one call to f. The quantum circuit is the same as in the Deutsch-Jozsa algorithm up the obtaining $|\psi_2\rangle$. At this point the state looks like

$$|\psi_2\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} (-1)^{x\cdot s} |x\rangle |-\rangle = \left[\bigoplus_{j=1}^{n} \frac{|0\rangle + (-1)^{s_j} |1\rangle}{\sqrt{2}} \right] \otimes |-\rangle \qquad (6.10)$$

where we replaced $f(x)$ by our concrete definition.

Finally, we can either measure the first n qubits in the X basis or apply a Hadamard gate first and then measure in the Z basis. We will opt for the latter method, removing the last qubit in $|\psi_2\rangle$ as it is no longer important for us, the state of the first n becomes

$$|\psi_3\rangle = |s_n\rangle |s_{n-1}\rangle \cdots |s_1\rangle \qquad (6.11)$$

This is already telling us what s is, we just need to measure each of the individual states in the Z basis and reconstruct the string.

The quantum circuit that performs this operation is the same as in the Deutsch-Josza (see fig. 5) replacing the unitary U_f to that implementing the function in this problem.

6.6 Simon's problem

This is again a variation of the Deutsch-Jozsa problem, in this case the function is such that for some unknown string $s \in \{0,1\}^n$ we have $f(x) = f(y) \Leftrightarrow x = y + s \mod 2$. Our goals is to determine s.

If you want to solve the problem classically, you need to find two different inputs x and y for which $f(x) = f(y)$. There is not necessarily any structure in the function f that would help you to find two such inputs. At least, you would need to guess $\mathcal{O}(2^{n/2})$ different inputs before being able to find a pair on which f takes the same output.

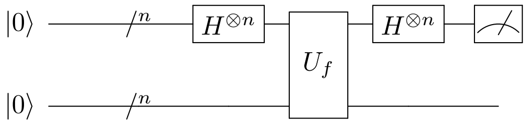

Figure 6: Quantum circuit for Simon's problem.

The quantum algorithm that solves this problem starts with the state

$$|\psi_0\rangle = |0\rangle^{\otimes n} |0\rangle^{\otimes n} \tag{6.12}$$

to which we apply a Hadamard on the n first qubits,

$$|\psi_1\rangle = H^{\otimes n} \otimes \mathbb{I}^{\otimes n} |\psi_0\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x\rangle |0\rangle^{\otimes n} \tag{6.13}$$

and evaluate the function $f(x)$ as always

$$|\psi_2\rangle = U_f |\psi_1\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x\rangle |f(x)\rangle \tag{6.14}$$

Next, we undo the Hadamard operation did before

$$\begin{aligned} |\psi_3\rangle = H^{\otimes n} \otimes \mathbb{I}^{\otimes n} |\psi_2\rangle &= \frac{1}{2^n} \sum_{x=0}^{2^n-1} \sum_{y=0}^{2^n-1} (-1)^{x\cdot y} |y\rangle |f(x)\rangle \\ &= \sum_{y=0}^{2^n-1} |y\rangle \left[\frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^{x\cdot y} |f(x)\rangle \right] \end{aligned} \tag{6.15}$$

The next step in the quantum algorithm is to measure in the first n registers, after this everything is classical processing of the information obtained. To cases must be distinguished:

$s = 0^n$ The function f is a one-to-one function, so the probability of obtaining y is given by

$$p_y = \left\| \frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^{x\cdot y} |f(x)\rangle \right\|^2 = \frac{1}{2^{2n}} \left\| \sum_{x=0}^{2^n-1} (-1)^{x\cdot y} |x\rangle \right\|^2 = \frac{1}{2^n} \tag{6.16}$$

because $f(x)$ will just reorder the inputs and the sum, because the states are orthogonal, can interchanged with the modulus which gives 2^n in total.

Thus, when $s = 0^n$, we can obtain any $y \in \{0,1\}^n$ with uniform probability 2^{-n}.

$s \neq 0^n$ The function is no longer one-to-one, some otputs are repeated. Name $S = \text{range}(f)$, then there must exist 2 distinct values x_z, x'_z such that $z = f(x_z) = f(x'_z)$ and it is necessary that $x_z = x'_z + s \mod 2$. So, the probability for y is

$$\begin{aligned} p_y &= \left\| \frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^{x\cdot y} |f(x)\rangle \right\|^2 \\ &= \left\| \frac{1}{2^n} \sum_{z\in A} \left[(-1)^{x_z\cdot y} + (-1)^{x'_z\cdot y} \right] |z\rangle \right\|^2 \\ &= \left\| \frac{1}{2^n} \sum_{z\in A} \left[(-1)^{x_z\cdot y} + (-1)^{(x_z+s)\cdot y} \right] |z\rangle \right\|^2 \\ &= \left\| \frac{1}{2^n} \sum_{z\in A} (-1)^{x_z\cdot y} \left[1 + (-1)^{s\cdot y}\right] |z\rangle \right\|^2 \end{aligned}$$

Thus, the possible results are

$$p_y = \begin{cases} 2^{-(n-1)} & s \cdot y = 0 \\ 0 & s \cdot y = 1 \end{cases} \tag{6.17}$$

So, the measurement always returns a string y that satisfies $s \cdot y = 0$ and the distribution is uniform over all the strings that satisfy the constraint. However, it is not possible to recover s yet, but we need more evaluations of U_f. In fact, in the best case we will need $n-1$ such that the system

$$\begin{cases} s \cdot y_1 = 0 \\ s \cdot y_2 = 0 \\ \vdots \\ s \cdot y_{n-1} = 0 \end{cases}$$

can be solved for s (remember that everything is modulo 2) if the strings $y_1, \dots, y_{n-1}$ are linearly independent. The probability that they are LI is about 0.289, so if that is not the cases with just $n-1$ evaluation, we can continue the process until we have $n-1$ independent equations which will quickly happen.

The important conclusion is that we can find s using $\mathcal{O}(n)$ evaluations of $f(x)$.

6.7 Trace estimation

Suppose we have a unitary gate U whose eigenvectors are known to be $\{|u_k\rangle\}$ but its eigenvalues $\{\lambda_k\}$ are unknown. We want to find the eigenvalues of the operator, notice that given that it is unitary the eigenvalues can be expressed as $\{e^{i\phi_k}\}$. We construct the reversible gate U_c from U that does the following: if the control qubit is set to $|1\rangle$ then it evaluates U on the second ($U_c\,|1\rangle\,|\varphi\rangle = |1\rangle\,U\,|\varphi\rangle$) but if it is set to $|0\rangle$ it does nothing.

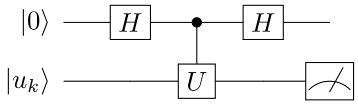

Figure 7: Quantum circuit for the problem.

Let's initialise the system at the state $|\psi_0\rangle = |0\rangle\,|u_k\rangle$ for some $|u_k\rangle$ and apply the Hadamard gate on the first qubit, $|\psi_1\rangle = |+\rangle\,|u_k\rangle$. Then, we apply the U_c gate on the second obtaining

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|u_k\rangle + |1\rangle\,(U\,|u_k\rangle)) = \frac{1}{\sqrt{2}}(|0\rangle\,|u_k\rangle + \lambda_k\,|1\rangle\,|u_k\rangle) \tag{6.18}$$

Finally, another Hadamard gate is applied on the first qubit giving

$$|\psi_3\rangle = \frac{1}{2}\left[(1+\lambda_k)\,|0\rangle + (1-\lambda_k)\,|1\rangle\right]|u_k\rangle \tag{6.19}$$

The simplest way to extract the value of λ_k is using the expectation value of Z with the last state

$$\langle Z\rangle = \langle\psi_3|Z|\psi_3\rangle = \frac{1}{2}(\lambda_k + \lambda_k) = \mathfrak{Re}\,\{\lambda_k\} \tag{6.20}$$

We were able to evaluate the real part of λ_k and using the fact that $|\lambda_k| = 1$ also the imaginary part can be known therefore the complete set of eigenvalues can be known.

In fact, we don't have to use any $|u_k\rangle$ at first, we could have used any other set of vectors. Suppose $d = \dim(U)$ and that all the vectors are equally distributed in probability, then the density matrix is $\rho_u = \sum p_k\,|u_k\rangle\!\langle u_k| = \mathbb{1}_d/d$ because $\{|u_k\rangle\}$ is a complete basis. So we could have used any other complete basis $\{|v_k\rangle\}$ (with equally distributed states) to initialise our system because $\rho_v = \mathbb{1}_d/d = \rho_u$ both systems are equivalent and lead to the same results.

After a sufficient number N of executions we can estimate the value of $\mathfrak{Re}\left\{[\right\}\operatorname{Tr}(U)]$ with an accuracy $1/\sqrt{N}$. That's in, in each iteration we will get a value $\mathfrak{Re}\left\{(\right\}\lambda_k)$ so after $d = \dim(U)$ iterations the mean between all the values will be $s = \frac{1}{d}\sum_{k=0}^{d-1}\mathfrak{Re}\left\{(\right\}\lambda_k) = \mathfrak{Re}\left\{(\right\}\operatorname{Tr}(U)/d)$ with some standard deviation σ. Repeating the experiment N times, the mean value would still be s but the standard deviation would be $\sigma/\sqrt{N} \leq \epsilon$ for an $\epsilon \propto 1/\sqrt{N}$. □

6.8 Grover search algorithm

Suppose we have a list of values with N elements $S = \{x_1, x_2, \ldots, x_N\}$, we are asked to find a certain $w \in S$ inside the list. Classically, with a simple linear search, it would take an average of $(N-1)/2$ comparisons to find w, this means an $\mathcal{O}(N)$. For small N this may be nice but if $N = 2^n$ the complexity grows exponentially fast.

In the quantum version we are again asked to find an element $|w\rangle$ from a list of $N = 2^n$ elements. We are provided with an oracle

$$f(x) = \begin{cases} 1 & x = w \\ 0 & x \neq w \end{cases} \tag{6.21}$$

implemented by a unitary operator U_f. Using quantum parallelism we will try to reduce the number of evaluations of this function as much as possible.

Initially, we don't know which is the state we are looking for so our situation is of complete ignorance, this state is represented by an equal superposition of all possible states

$$|\Phi\rangle = \frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}|x\rangle \tag{6.22}$$

where one of the $|x\rangle$ is our solution $|w\rangle$. Then, we can rewrite this state as

$$|\Phi\rangle = \sqrt{\frac{N-1}{N}}\sum_{x\neq w}|x\rangle + \frac{1}{\sqrt{N}}|w\rangle = \cos\frac{\theta}{2}|\alpha\rangle + \sin\frac{\theta}{2}|w\rangle \tag{6.23}$$

Did we solve the problem? No, if we measured now in the basis of N elements, the probability of collapsing to our state is still $1/N$.

Next, we apply the Grover's function G to the previous state defined as

$$G = KU_f \tag{6.24}$$

where K is called the inversion over the mean operator

$$K = 2|\Phi\rangle\langle\Phi| - \mathbb{I} \tag{6.25}$$

and U_f is the implementation of the oracle

$$U_f = |\alpha\rangle\langle\alpha| - |w\rangle\langle w| \tag{6.26}$$

The complete action of G is resumed in the basis $\{|\alpha\rangle, |w\rangle\}$ as

$$G = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{6.27}$$

which is just a rotation on the $\alpha - w$ plane of θ degrees.

Then, the application of G over the state $|\Phi\rangle$ returns the state

$$G|\Phi\rangle = \cos\left(\frac{\theta}{2} + \theta\right)|\alpha\rangle + \sin\left(\frac{\theta}{2} + \theta\right)|w\rangle$$

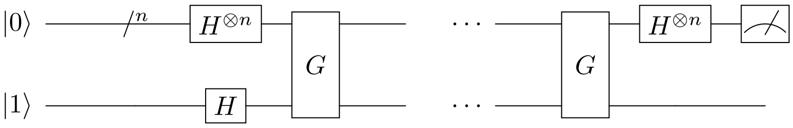

Figure 8: Quantum circuit implementing Grover's algorithm.

At this step the probability of finding $|w\rangle$ has increased while the probability of finding any other value has decreased, we are on the right path. We can continue applying G repeatedly until $p(w)$ maximises. In general, after k iterations the state would be

$$G^k |\Phi\rangle = \cos\left(\frac{\theta}{2} + k\theta\right) |\alpha\rangle + \sin\left(\frac{\theta}{2} + k\theta\right) |w\rangle \tag{6.28}$$

where the probability of finding w now is

$$p(w) = \sin^2\left(\frac{\theta}{2} + k\theta\right) \tag{6.29}$$

Maximising this function over k, this is taking $p(w) \approx 1$ one deduces that

$$k \approx \frac{1}{2}\left(\frac{\pi}{\theta} - 1\right) \approx \frac{1}{2}\left(\frac{\pi}{2}\sqrt{N} - 1\right) \tag{6.30}$$

where we used that $\sin(\theta/2) \approx \theta/2 = 1/\sqrt{N}$.

We can see that the number of iterations, the number of evaluations of $f(x)$ grows as $\mathcal{O}(\sqrt{N})$ which still is exponential for $N = 2^n$ but smaller than classically. In fig. 8, the Grover function U has to be applied this exact number of times to achieve maximum probability in the measure.

The hoe algorithm can be remake for 2 or more possible solutions. Suppose there are M possibilities, then we could write the state

$$|\Phi_M\rangle = \sqrt{\frac{N-M}{N}} |\alpha\rangle + \sqrt{\frac{M}{N}} |w\rangle = \cos\frac{\theta}{2} |\alpha\rangle + \sin\frac{\theta}{2} |w\rangle \tag{6.31}$$

where $|w\rangle$ is now an equiprobable superposition of all the possible solutions. The algorithm will be run in the same way, except that now the number of iterations has been reduced to

$$k_M \approx \frac{1}{2}\left(\frac{\pi}{2}\sqrt{\frac{N}{M}} - 1\right) \tag{6.32}$$

It is important to note that finding $M > N/2$ solutions may lead to incorrect results as the algorithm will reverse its work and increase the probability of the no-solutions instead of the solutions. To resolve this error, what we can do is increase the number of items in the list so that $M \ll N$ and continue normally.

6.9 Quantum Fourier Transform (QFT)

A discrete version of the Fourier transform can be defined as

$$\tilde{x}_k \equiv \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j e^{i2\pi jk/N} \tag{6.33}$$

where $x = (x_1, \dots, x_N)$ is the vector that is being transformed.

The quantum Fourier transform is based on essentially the same idea with the only difference that the vectors x and $\tilde{x}$ are the state vectors

$$|x\rangle = \sum_{j=0}^{N-1} x_j |j\rangle \qquad , \qquad |\tilde{x}\rangle = \sum_{k=0}^{N-1} \tilde{x}_k |k\rangle \tag{6.34}$$

Then, the action on the components of the state $|x\rangle$ is described by

$$\tilde{x}_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{i2\pi jk/N} x_j \tag{6.35}$$

In other words the incoming amplitude x_j of a given basis vector $|k\rangle$ (in original or "positio" space) is distributed among all basis vector (in Fourier or "momentum" space). What is, however, different is that both the original vector x and the transformed vector $\tilde{x}$ is recorded using the very same Hilbert space.

Quantum circuit In the case of quantum computation, the basis vectors $|j\rangle$ are the computational basis vectors for let's say n-qubits. Then it will be useful to adopt the binary representation

$$j : (0,1) \to \{0,1\}^n \quad \text{where} \quad j = 0.(j_1, \dots, j_m) \equiv \sum_{i=1}^{m} j_i 2^{-i} \tag{6.36}$$

Then, the Fourier decomposition can be expressed as

$$\begin{aligned}
|j\rangle = |j_1 j_2 \dots j_n\rangle &= \frac{1}{2^{n/2}} \sum_{k}^{2^n-1} \exp(i2\pi jk/2^n)\,|k_1 k_2 \dots k_n\rangle \\
&= \frac{1}{2^{n/2}} \sum_{k_1,\dots,k_n=0}^{1} \exp\left[i2\pi j \left(\sum_{l=1}^{n} k_l 2^{n-l}\right) 2^{-n}\right] |k_1 \dots k_n\rangle \\
&= \frac{1}{2^{n/2}} \sum_{k_1,\dots,k_n=0}^{1} \bigotimes_{l=1}^{n} \exp\big(i2\pi j k_l 2^{-l}\big)\,|k_l\rangle \\
&= \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \sum_{k_l=0}^{1} \exp\big(i2\pi k_l 2^{-l}\big)\,|k_l\rangle \\
&= \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \big[|0\rangle + \exp\big(i2\pi j 2^{-l}\big)\,|1\rangle\big] \\
&= \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \left[|0\rangle + \exp\left(i2\pi \left(\sum_{k=1}^{n} j_k 2^{n-k}\right) 2^{-l}\right) |1\rangle\right] \\
&= \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \left[|0\rangle + \exp\left(i2\pi \left(\sum_{k=n-l+1}^{n} j_k 2^{n-l-k}\right) 2^{-l}\right) |1\rangle\right] \\
&= \frac{1}{2^{n/2}} \big(|0\rangle + e^{i2\pi 0.j_n}\,|1\rangle\big)\big(|0\rangle + e^{i2\pi 0.j_{n-1}j_n}\,|1\rangle\big) \cdots \big(|0\rangle + e^{i2\pi 0.j_1 j_2 \dots j_n}\,|1\rangle\big)
\end{aligned}$$

The complete quantum circuit can be found in Nielsen and Chuang (chapter 5) which makes use of Hadamard gates and concatenated rotations of phase

$$R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{i2\pi/2^k} \end{pmatrix} \tag{6.37}$$

It is possible to see that the number of gates used grows as $\mathcal{O}(n^2)$ while the best classical algorithm (FFT) uses exponentially many gates $\mathcal{O}(n2^n)$.

Of course, keep in mind that not all of the information about the Fourier transformed state vector can be retrieved.

6.10 Quantum phase estimation

The job of the quantum phase estimation algorithm consist on finding the phase φ of an eigenvalue of U given a known eigenstate $|u\rangle$. In fact, we won't find exactly φ, but we will find an approximation φ' of n bits.

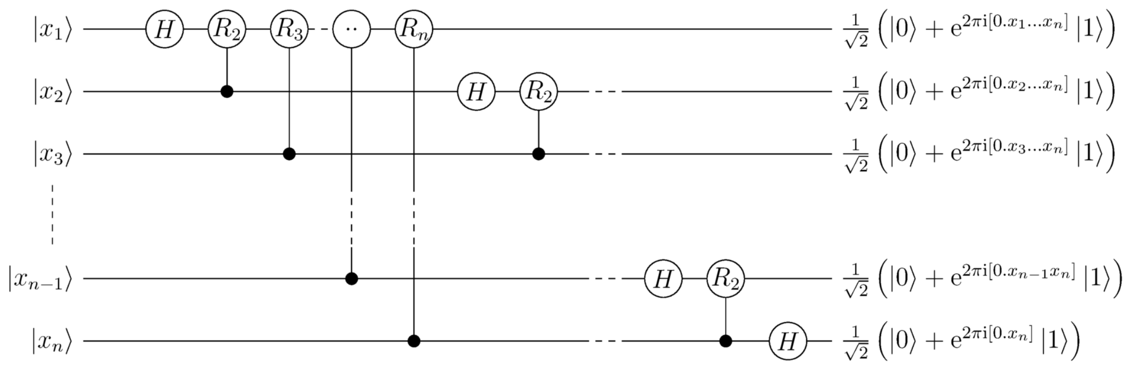
$|x_1\rangle$
$|x_2\rangle$
$|x_3\rangle$
$|x_{n-1}\rangle$
$|x_n\rangle$
H
R_2
R_3
$\cdot\cdot$
R_n
$\frac{1}{\sqrt{2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}[0.x_1 \ldots x_n]}|1\rangle\right)$
$\frac{1}{\sqrt{2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}[0.x_2 \ldots x_n]}|1\rangle\right)$
$\frac{1}{\sqrt{2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}[0.x_3 \ldots x_n]}|1\rangle\right)$
$\frac{1}{\sqrt{2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}[0.x_{n-1} x_n]}|1\rangle\right)$
$\frac{1}{\sqrt{2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}[0.x_n]}|1\rangle\right)$

Figure 9: Quantum Fourier Transform circuit.

In order to do so, we prepare that state

$$|\psi_0\rangle = |0\rangle^{\otimes t} |u\rangle \tag{6.38}$$

where t is given by [Nielsen and Chuang, 2010]

$$t = n + \left\lceil \log\left(2 + \frac{1}{2\epsilon}\right) \right\rceil \tag{6.39}$$

for some small ϵ.

Then, we create the superposed state by applying the Hadamard gates

$$|\psi_1\rangle = 2^{-t/2} \sum_{x=0}^{2^t-1} |x\rangle |u\rangle \tag{6.40}$$

and apply the unitary operator

$$|\psi_2\rangle = 2^{-t/2} \sum_{x=0}^{2^t-1} |x\rangle U_x |u\rangle = 2^{-t/2} \sum_{x=0}^{2^t-1} e^{i2\pi x\varphi} |x\rangle |u\rangle \tag{6.41}$$

Suppose that φ has the very special form $\varphi = y/2^t$, then the state $|\psi_2\rangle$ is just proportional to the QFT of the state $|y\rangle$. Thus, applying the inverse QFT on the first register we obtain

$$|\psi_3\rangle = |y\rangle |u\rangle \tag{6.42}$$

and so, a measurement of the first register will give us with certainty the value of y and with we will find the phase $\varphi = y/2^t$.

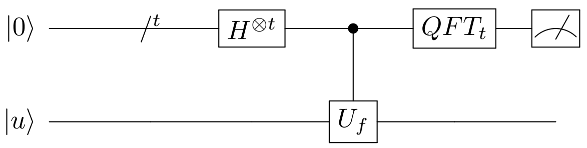

Figure 10: Quantum circuit for the phase estimation problem.

Of course, the previous is a very special case of φ but it helps us to explain the more general case. In general, φ will be a real number and the application of the inverse QFT will output the state (forgetting about the second register as it is not interesting to us anymore)

$$|\psi_3\rangle = 2^{-t} \sum_{x=0}^{2^t-1} \sum_{y=0}^{2^t-1} e^{2\pi i x\varphi} e^{-2\pi i x y/2^t)|y\rangle} = 2^{-t} \sum_{y=0}^{2^t-1} \left(\sum_{x=0}^{2^t-1} e^{2\pi i x(\varphi 2^t - y)/2^t} \right) |y\rangle \tag{6.43}$$

Then, measuring the state we will obtain the value y with probability

$$p(y) = \left| \frac{1}{2^t} \sum_{x=0}^{2^t-1} e^{2\pi i x(\varphi 2^t - y)/2^t} \right|^2 = \frac{1}{2^{2t}} \left| \frac{1-\exp[2\pi i(\varphi 2^t - y)]}{1-\exp[2\pi i(\varphi 2^t - y)/2^t]} \right|^2 \tag{6.44}$$

If we were on the previous case then the only possible value of y would be that satisfying $\varphi = y/2^t$ but here we can only demand to find that y for which $\varphi = y2^{-t} + \epsilon$ for $\epsilon < 1$ where ϵ represents the error made in the approximation. In terms of ϵ, the probability of finding the closest y is

$$p(y) = \frac{1}{2^{2t}} \left| \frac{1-\exp[2\pi i 2^t \epsilon]}{1-\exp[2\pi i \epsilon]} \right|^2 = \frac{1}{2^{2t}} \left| \frac{\sin(\pi 2^t \epsilon)}{\sin(\pi\epsilon)} \right|^2 \geq \frac{1}{2^{2t}} \left| \frac{22^t \epsilon}{\pi\epsilon} \right|^2 = \frac{4}{\pi^2} \approx 0.41 \tag{6.45}$$

where we have used that, if $\epsilon \leq 2^{-(t+1)}$ then $\sin(\pi\epsilon) \leq \pi\epsilon$ and $\sin \pi 2^t \epsilon \geq 22^t \epsilon$.

This result shows that we will measure the best t-bit estimate of φ with high probability. By increasing the number of qubits by $\mathcal{O}(\log(1/\epsilon))$ and ignoring those last qubits we can increase the probability to $1-\epsilon$.

As we've seen, this is not an exact algorithm, it is a method to find an approximation to some phase. However, it is certainly enough in much occasions as classical computation is also limited to a certain amount of bit of precision.

6.11 Quantum period finding

Now, suppose we are given a function $f(x) : \{0,1\}^n \to \{0,1\}^n$ that we know is periodic, this is $f(x) = f(x+kr)$ for $k = 1, \ldots, A$, but we don't know the period r. Our task is to find the value of r with highest probability.

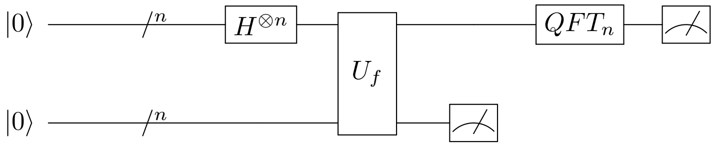

Figure 11: Quantum circuit for the period finding problem.

We shall start with the state

$$|\psi_0\rangle = |0\rangle^{\otimes n} |0\rangle^{\otimes n} \tag{6.46}$$

and apply a Hadamard gate on the first n qubits

$$|\psi_1\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle |0\rangle^{\otimes n} \tag{6.47}$$

Then, we evaluate our function using the unitary gate U_f as

$$|\psi_2\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle |f(x)\rangle \tag{6.48}$$

The tricky part starts, we will measure the state of the second register and suppose we obtain the value $y = f(x)$. The state will collapse to this exact value but, because there are multiple x that have the same $f(x)$, we obtain in the first register a superposition of all of them

$$|\psi_3\rangle = \frac{1}{\sqrt{A}} \sum_{x|y=f(x)} |x\rangle |y\rangle = \frac{1}{\sqrt{A}} \sum_{k=0}^{A-1} |x_0 + kr\rangle |y\rangle \tag{6.49}$$

where x_0 if the smallest x that satisfies $f(x) = y$.

The next step is to perform the QFT on the first registers which will take us to

$$|\psi_4\rangle = \frac{1}{\sqrt{A}}\sum_{k=0}^{A-1}\left[\frac{1}{\sqrt{N}}\sum_{l=0}^{N-1} e^{i2\pi\frac{kl}{A}}|l\rangle\right]|y\rangle = \frac{1}{\sqrt{AN}}\sum_{l=0}^{N-1}\left[\sum_{k=0}^{A-1} e^{i2\pi\frac{kl}{A}}\right]|l\rangle|y\rangle \tag{6.50}$$

Finally, we will measure in the first register where we will obtain some of the possible values for l with probability

$$p(l) = \frac{1}{AN}\left|\sum_{k=0}^{A-1}\exp\left\{i2\pi\frac{kl}{A}\right\}\right|^2 \tag{6.51}$$

At this step we are done with all the quantum stuff, our job now is to analyse the result obtained as to extract the information of the period r from it. We have two possibilities:

$N/r \in \mathbb{N}$ In this case, the number of elements in the superposition is exactly $A = N/r$ and so eq. (6.51) is exact. We can show then that, the only possible outcomes of the measurement in l are the ones that satisfy $l = Am$. Indeed,

- If $l = Am$ ($m \in \mathbb{Z}$), the probability becomes

$$p(l) = \frac{1}{AN}\left|\sum_{k=0}^{A-1}\exp\left\{i2\pi\frac{k}{A}Am\right\}\right|^2 = \frac{1}{AN}\left|\sum_{k=0}^{A-1}\exp\{i2\pi km\}\right|^2 = \frac{1}{AN}|A|^2 = \frac{A}{N} = \frac{1}{r}$$

- Instead, if $l \neq Am$, the probability vanishes as

$$p(l) = \frac{1}{AN}\left|\frac{e^{i2\pi l}-1}{e^{i2\pi l/A}-1}\right|^2 = 0$$

since the numerator vanishes.

The information that we can extract from this is that, with probability $1/r$ we obtain the result $l = Am$, put in other form, we obtain

$$\frac{l}{N} = \frac{m}{r} \tag{6.52}$$

Therefore, by simplifying the fraction l/N we can obtain r in the denominator if r is prime or a factor of $1/r$, but executing the algorithm a second time we can obtain finally r with high probability.

$N/r \notin \mathbb{N}$ This is a much more complicated case and so we will just state the basic results. It turns out that when the fraction N/r is not an integer, the probability to obtain a certain value l is given by

$$p\left(\left|\frac{l}{N}-\frac{m}{r}\right| \leq \frac{1}{2N}\right) \geq \frac{4}{\pi^2} \approx 0.41 \tag{6.53}$$

This is saying that, the probability to obtain a value l that, after divide it by N, is close to a fraction proportional to $1/r$ is approximately 0.41. In other words, four of every six times we will obtain a value l that is close to the solution. We can then construct r using a method of continuous functions that is not really important for us now.

6.12 Shor's algorithm

In the world we are living, the most used cryptography algorithm used is RSA (explained in section 8) which is based in the statement that "it is very difficult to factorise large numbers". However, if we found some way to efficiently factorise it, then we could decrypt most of the transactions that are done.

Shor's algorithm is capable of factorising very large numbers by using the quantum period finding algorithm and basic number theory. Suppose N is the L-bit number we want to factorise which we know it is a product of two prime numbers p, q ($N = pq$). The procedure goes as follow:

1. Randomly choose x in the range $1 < x < N-1$ such that x is coprime with N, this is, $\gcd(x, N) = 1$.
2. Use the period finding algorithm to determine the period r of the function $f(a) = x^a \mod N$.
3. If r is even and $x^{r/2} + 1 \neq 0 \mod N$ then compute $\gcd(x^{r/2} - 1, N)$ and $\gcd(x^{r/2} + 1, N)$ and test if one of these is a non-trivial factor. If so, we would have found p, then compute $q = N/p$. If not, start over again with a different x.

Example Let's fix $N = 91$ and choose $x = 4$, we can check that $\gcd(4, 91) = 1$. First, compute the period r, we will do so manually by computing $f(a)$ for various a: We see that the period $r = 6$

a	1	2	3	4	5	6	7	8	9	10	...
$f(x)$	4	16	64	74	23	4	16	64	74	23	...

as the values $f(a)$ start repeating. Once r is known, check that $\gcd(4^{6/2}, 91) + 1 = \gcd(64, 91) + 1 = 64 + 1 = 65 \neq 0 \mod N$. The algorithm has succed and we are ready to find one of the factors of 91 which is $p = \gcd(4^3 - 1, 91) = \gcd(63, 91) = 7$ and the other is $q = 91/7 = 13$.

Indeed, it seems an algorithm easy to implement but in order to factor out a number we need a quantum computer with a number of qubits of the order of the length of N which is usually either $2^{10}, 2^{11}$ or 2^{12} for extra security. The largest quantum computer out in the market has 72 qubits, in the past decade this value was 5, so if the number of qubits multiply by 10 after every decade we still have to wait around 30 years to see a physical realisation of this algorithm. Moreover, there already exist algorithms (post-quantum algorithms) for when this time comes.

7 Quantum information

In this chapter we review the basic definitions and properties of entropy in both classical and quantum information theory.

7.1 Shannon entropy

The key concept of classical information theory is the Shannon entropy. Suppose we learn the value of a random variable X. The Shannon entropy of X quantifies how much information we gain, on average, when we learn the value of X. An alternative view is that the entropy of X measures the amount of uncertainty about X before we learn its value. These two views are complementary; we can view the entropy either as a measure of our uncertainty before we learn the value of X, or as a measure of how much information we have gained after we learn the value of X.

The Shannon entropy of a variable X is defined by

$$H(X) = -\sum_x p_x \log_2 p_x \tag{7.1}$$

where p_x is the probability for the event x to occur.*

The best reason for this definition of entropy is that it can be used to quantify the resources needed to store information. The Shannon entropy quantifies the expected amount of memory (bits) that is required to record the values of X. Or, a different interpretation is that it expresses our ignorance on a test made on that system [Peres, 2006].

Note that, from the definition, it can be seen that Shannon's entropy is maximum when all events are equally likely. The uncertainty of the system is maximum because the events are uncorrelated so, at most, one 1 bit of information is obtained per question (measurement).

Binary entropy Consider the case of two possible states with probability p and $1-p$ with $p \in [0,1]$, the entropy of the system is

$$H_2(p) = -p \log p - (1-p)\log(1-p) \tag{7.2}$$

As commented before, H_2 has its maximum value at $p = 1/2$ when they are equally distributed and goes to zero as $p \to 0$ or 1. This is because the probability starts increasing for one of the two possibilities therefore, prior to the experiment, we have some sort of intuition on what the result will be, reducing the amount of information gained after the experiment.

Joint entropy Now consider two systems X and Y, the joint entropy of the two systems is

$$H(X,Y) = -\sum_{x,y} p(x,y)\log p(x,y) \tag{7.3}$$

where $p(x,y)$ is the probability that both events x and y happen together. The joint entropy measures our uncertainty about the pair (X,Y). But is $H(X,Y) = H(X) + H(Y)$?

The answer is yes only when the two systems are uncorrelated, in other words, what happens on one of the systems doesn't affect the other. This implies that the uncertainty is maximum because although we know completely the result of X, Y is still completely unknown (similarly to what happens in the binary entropy).

What if the two systems are correlated? Then, after knowing some information about X we would be able to say something about Y. To express the amount of information still unknown, we define the **conditional entropy** of X with respect to Y

$$H(X|Y) = H(X,Y) - H(Y) \tag{7.4}$$

*Note that we used log base 2 in our definition which is more useful than ln when talking about bits of information, so from now on log will explicitly mean base 2 although not indicated.

From the definition it is the total amount of information available minus the information stored only in Y. The conditional entropy is a measure of how uncertain we are, on average, about the value of X, given that we know the value of Y.

A second quantity, the **mutual information** content of X and Y, measures how much information X and Y have in common

$$H(X:Y) = I(X,Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{7.5}$$

Let's see some useful properties:

1. $H(X,Y) = H(Y,X)$ and $H(X:Y) = H(Y:X)$
2. $H(Y|X) \geq 0 \Rightarrow H(X:Y) \leq H(Y)$ with equality if and only if $Y \subset X$.
3. $H(X) \leq H(X,Y)$ with equality if and only if $Y \subset X$.
4. Subadditivity: $H(X,Y) \leq H(X) + H(Y)$ with equality iif X and Y are independent random variables.
5. $H(Y|X) \leq H(Y) \Rightarrow H(X:Y) \geq 0$ with equality iif X and Y are independent random variables.
6. Strong subadditivity: $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$ with equality iif $Z \to Y \to X$ forms a Markov chain.
7. Conditioning reduces entropy: $H(X|Y,Z) \leq H(X|Y)$

7.2 Von Neumann entropy

The Shannon entropy measures the uncertainty associated with a classical probability distribution. Quantum states are described in a similar fashion, with density operators replacing probability distributions. Von Neumann defined the entropy of a quantum state ρ by the formula

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho) = -\sum_k \lambda_k \log \lambda_k \tag{7.6}$$

given the eigenvalues $\{\lambda_k\}$ of the density matrix ρ.

Suppose ρ and σ are density operators, the relative entropy of ρ with respect to σ is

$$S(\rho||\sigma) = \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\rho \log \sigma) \tag{7.7}$$

which is a measure of distinguishability between two quantum systems. The quantum relative entropy can sometimes be infinite. In particular, the relative entropy is defined to be $+\infty$ if the kernel of σ (the vector space spanned by the eigenvectors of σ with eigenvalue 0) has non-trivial intersection with the support of ρ (the vector space spanned by the eigenvectors of ρ with non-zero eigenvalue), and is finite otherwise. The quantum relative entropy is non-negative, $S(\rho||\sigma) \geq 0$.

Properties

1. The entropy is non-negative and it is zero iff the state is pure.
2. In a $d$dimensional Hilbert space the entropy is at most $\log d$ with equality iff the system is in a completely mixed state with $\rho = \mathbb{I}/d$.
3. Suppose a composite system AB is in a pure state, then $S(A) = S(B)$.
4. Joint entropy theorem: suppose p_i are probabilities, $|i\rangle$ are orthogonal states for a system A and ρ_i is any set of density operators for another system B, then

$$S\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i) \tag{7.8}$$

5. Given a separable system $\rho \otimes \sigma$, the entropy of the complete system is given by

$$S(\rho \otimes \sigma) = S(\rho) + S(\sigma) \tag{7.9}$$

The joint, conditional and mutual entropy are defined following from their definition of Shannon entropy

$$S(X, Y) = -\operatorname{Tr}\left(\rho^{AB} \log \rho^{AB}\right) \tag{7.10}$$

$$S(X|Y) = S(X, Y) - S(Y) \tag{7.11}$$

$$S(X : Y) = S(X) + S(Y) - S(X, Y) = S(\rho^{AB} || \rho^{A} \otimes \rho^{B}) \tag{7.12}$$

Qubit What is the entropy of a single quantum bit of information? Think of it, if we only have one state in that system, there is no randomness at all, no questions have to be asked to the system because we already know the answer. Therefore, the entropy of a single qubit is 0. Mathematically, the density matrix of this system would be $\rho = |\psi\rangle\langle\psi|$ with eigenvalue 1 so $S(\rho) = 1 \log 1 = 0$.

Let's complicate a bit (jeje) the system, what if we have a lineal combination of $|0\rangle$ and $|1\rangle$? The answer is still the same, $S = 0$, because with a rotation we can return to a pure state and as a rotation is a unitary operation the entropy doesn't change.

Theorem 9. *The entropy of a system doesn't change under unitary operations U, $S(\rho) = S(U\rho U^{\dagger})$.*

So, any point in the surface of Bloch sphere has zero entropy, but states get progressively more mixed as they are moved towards the centre.

Any pair of states that verify $\langle\psi|\phi\rangle > 0$ cannot be perfectly distinguished (they have some correlation). Therefore, a mixture of this states has an entropy less than one. As $\langle\psi|\phi\rangle \to 1$, the less distinguishable they are, the less information is stored in them.

Entangled states When entangled states are considered, things change. For example, with the Bell state $|\Psi^{+}\rangle$ each qubit can store one bit $S(A) = S(B) = 1$ but by knowing one, the answer to the other is known so $S(A, B) = 0$. The mutual information is $H(A : B) = 2$ so $H(A|B) = S(A) - H(A : B) = -1$.

Principle 11. *An entangled system has negative conditional entropy.*

A quantum coin can know more about another system than that system can possibly know about itself.

7.3 Thermodynamic entropy

It is a well known principle of physics that the entropy of a system always increments

$$dS \geq 0 \tag{7.13}$$

this is called the second law of thermodynamics and equality is only reached in ideal and reversible process.

Without going into details the entropy of a system with energy Q at temperature T is

$$dS = \frac{\delta Q}{T} \tag{7.14}$$

Now, consider that we want to build a machine that takes heat from a cold environment to a warmer environment. Let Q be the amount of heat transferred from T_C to T_H ($T_H > T_C$) then the change in entropy of this process is

$$\Delta S = \frac{Q}{T_H} - \frac{Q}{T_C} = Q\frac{T_C - T_H}{T_C T_H} < 0 \tag{7.15}$$

Thus, violating the second law of thermodynamics. This is telling us that, naturally, the heat will transfer from warmer reservoirs to cooler but not vice verse. There exist and arrow of time indicating the directions of the events.

Maxwell demon (Szilard's engine) [Maruyama et al., 2009] Consider the quantum information variation of this problem. A chamber of volume V contains a gas, which consists of a single molecule (Fig. 1(a)). As a first step of the process, a thin, massless, adiabatic partition is inserted into the chamber quickly to divide it into two parts of equal volumes. The demon measures the position of the molecule, either in the right or in the left side of the partition. The demon records this result of the measurement for the next step. Then, he connects a load of a certain mass to the partition on the side where the molecule is supposed to be in, according to his recorded result of the previous measurement. Keeping the chamber at a constant temperature T by a heat bath, the demon can let the gas do some work W by quasistatic isothermal expansion (the partition now works as a piston). The gas returns to its initial state, where it now occupies the whole volume V, when the partition reaches the end of the chamber. During the expansion, heat Q is extracted from the heat bath and thus $W = Q$ as it is an isothermal process. Hence, Szilard's engine completes a cycle after extracting heat Q and converting it to an equal amount of mechanical work.

As the gas is expanded isothermally, the amount of extracted work is

$$W = k_B T \int_{V/2}^{V} \frac{dV}{V} = k_B T \ln 2 \tag{7.16}$$

The factor $k_B T \ln 2$ appears often in the following discussions on thermodynamic work, so we will take it as a unit and call it '1 bit' when there is no risk of confusion. This will be especially useful when we coordinate discussions of the information theoretic 'bit' with the thermodynamic work.

The demon apparently violates the second law. As a result of the perfect conversion of heat Q into work W, the entropy of the heat bath has been reduced by $Q/T = W/T = k_B \ln 2$. According to the second law, there must be an entropy increase of at least the same amount somewhere to compensate this apparent decrease. Szilard attributed the source of the entropy increase to measurement. He wrote, "The amount of entropy generated by the measurement may, of course, always be greater than this fundamental amount, but not smaller".

The answer to the demon paradox remain unknown until Charles Bennett in 1982 considered the role of information processing by the demon. Since information processing must be carried out by a certain physical system, there should be a one-to-one correspondence between logical and physical states. Logical states may be described as an abstract set of variables on which some information processing can be performed. Then, a reversible logical process, which means an injective (one-to-one) mapping for logical states, corresponds to a reversible physical process.

However, a logically irreversible process is noninjective, i.e. many-to-one, mapping. Such a process does not have a unique inverse as there may be many possible original states for a single resulting state. The key here is that memory erasure is a logically irreversible process because many possible states of memory should be set to a single fixed state after an erasing procedure.

Now, let's focus on the erasure of information. The physical system for the demon's memory can be modelled as a one-molecule gas in a chamber of volume V , which is divided into two parts, the left L and the right R, by a partition. The demon memorizes the measurement result by setting the position of the molecule in this box. If the molecule in Szilard's engine may be found in the left and the right sides with equal probability, i.e. 1/2, then the minimum amount of work that needs to be invested and dissipated into the environment is $k_B T \ln 2$.

Szilard's engine turns out to be a isotropic process as $\Delta S = 0$ but in general, for two systems at temperature T_C and $T_H > T_C$ the maximum efficiency of a machine operating between both states is

$$\eta = \frac{Q_H - Q_C}{Q_H} = 1 - \frac{T_C}{T_H} \tag{7.17}$$

known as Carnot efficiency, which is the maximum that one can get without violating the second law of thermodynamics.

Erasure of information The last *gedankenexperiment* has real consequences in the quantum information regime.

Principle 12 (Landauer)**.** *The erasure of n bits of memory at temperature T requires $nk_B T \ln 2$ units of energy.*

This principle is telling us that the storage of information does not come for free but a finite amount of energy has to be spend in this purpose.

This principle can be generalised to any function f which maps $X \xrightarrow{f} Y$:

Principle 13. *Physical realisation of any function that maps an input x distributed according to a random variable X to $y = f(x)$ requires energy expenditure of*

$$W = k_B T H(X|Y) \ln 2 \tag{7.18}$$

with $H(X|Y) = H(X) - H(Y)$ because $H(Y|X) = 0$ since $Y = f(X)$.

For example, in a two level system where the entropy is given by $h(p)$ the amount of work that can be extracted is $k_B T h(p)$. Or, think of a XOR gate that maps $\{00, 11\} \to 0$ and $\{01, 10\} \to 1$, the entropy of X is $\ln 2$ but the entropy of Y is $\ln 1$ thus, the amount of bits erased is $k_B T \ln 2$.

Maxwell's demon is now exorcized. The entropy decrease, or the equivalent work the demon could give us, should be completely consumed to make his memory state come back to its initial state. The state of the whole system, consisting of the heat engine and the demon, is restored after completing a thermodynamic cycle, without violating the second law.

Temperature The temperature that appears in the above expression is not our classic temperature but we should redefine it as *the minimal energetic cost needed to erase 1 bit in the environment* so $T \propto E/\text{bit}$.

8 Quantum cryptography

In fact, in the following sections, we will explain the most typical protocols for quantum key distribution which becomes the important step in order to handle a private conversation later on. The process of encrypting a message is made as in the classical theory, using the RSA algorithm.

The key plays the most important role in cryptography as it enables to securely encrypt and decrypt information between two parties, usually called Alice and Bob. However, if a malicious agent, the eavesdropper or Eve in short, has the key then their communication channel is now longer secure. This type of cryptosystems are called private-key cryptosystems and we will investigate its security in the following sections.

8.1 Conditions of security

Before heading into actual protocols to share keys and messages we should investigate the conditions of security, those conditions that will tell us if a communication system is secure.

We will put ourselves in the worst of the cases, Eve has complete control on the channel and she is able to capture every bit of information that we send. However, she doesn't know the key k. Alice wants to send a message m to Bob and uses the encoding function $\mathrm{Enc}(m,k)=\tilde{m}$ to generate a string $\tilde{m}$ which will be send to Bob. He will use the key k and the string $\tilde{m}$ to recover the original message using a decoding function $\mathrm{Dec}(\tilde{m},k)=m$.

Firstly, Eve has been listening to the whole conversation and knows $\tilde{m}$ perfectly, can she extract any information on m? The answer to this question has to be negative in all cases, thus we require that

$$p(m)=p(m\mid\tilde{m}) \tag{8.1}$$

which means that the original message m and the encoded string $\tilde{m}$ are completely independent. Absolutely no information is gained by having access to the channel! If eq. (8.1) is satisfied we say that the protocol is secure.

Secondly, we must also impose a condition which might seem trivial but without which communication wouldn't be possible. We must have that

$$m=\mathrm{Dec}(\mathrm{Enc}(m,k)) \qquad \forall m,k \tag{8.2}$$

We are always capable of recovering the original message without losing information. If eq. (8.2) is satisfied we say that the protocol is correct.

The two conditions impose a restriction in the size of the keys, let $\mathscr{K}$ be the space of all keys with total size[*] $|\mathscr{K}|$ and let $\mathscr{M}$ be the space of all message with total size $|\mathscr{M}|$. Then,

Theorem 10. *An encryption scheme (*Enc, Dec*) is secure and correct if and only if* $|\mathscr{K}|\geq|\mathscr{M}|$.

In following section we will always try to find the smallest $\mathscr{K}$ that still satisfies the previous condition to make a scheme secure and correct.

One time pad The simplest example of a secure and correct scheme is the so called one time pad. Moreover, this protocol saturates the condition of theorem 10.

Suppose Alice wants to send a message $m\in\{0,1\}^n$ to Bob and they both share a key $k\in\{0,1\}^n$ which is as long as the message. The encryption/decryption scheme work as follows:

> **Protocol 1.** One time pad
> *A message* $m\in\{0,1\}^n$ *is encrypted and decrypted using a key* $k\in\{0,1\}^n$ *with the operations:*
>
> $$\mathrm{Enc}(m,k)=m\oplus k=(m_1\oplus k_1,m_2\oplus k_2,\dots,m_n\oplus k_n) \tag{8.3}$$
> $$\mathrm{Dec}(\tilde{m},k)=\tilde{m}\oplus k \tag{8.4}$$

[*] By size we understand the total number of symbols in the space.

where $a \oplus b = a + m \mod 2$.

Proof. The proof of correctness is almost trivial, one just has to note that for any bit $m, k \in \{0, 1\}$ it holds that $(m \oplus k) \oplus k = m \oplus (k \oplus k) = m \oplus 0 = m$.

For the proof of security, note that for a uniformly random key k where each bit can have the value zero or one with equal probability it holds that $p(\tilde{m} \mid m) = p(m \otimes k \mid m) = 2^{-n}$. Moreover, since $\tilde{m}$ is also uniformly distributed, i.e. $p(m) = \sum_m p(m)p(\tilde{m} \mid m) = 2^{-n}$, the probability of recovering m knowing $\tilde{m}$ is

$$p(m \mid \tilde{m}) = \frac{p(m, \tilde{m})}{p(\tilde{m})} = \frac{p(\tilde{m} \mid m)p(m)}{p(\tilde{m})} = p(m)$$

satisfying the condition on eq. (8.1). □

8.2 The quantum approach, how to make Eve ignorant?

8.3 Classical cryptography: RSA algorithm

All current cryptography algorithms are based on the difficulty to factorise a product of two prime numbers. RSA is one of the most used protocols and work as follows:

1. Choose two prime numbers p and q and evaluate their product $n = pq$. The prime numbers p and q are kept private but n is published.
2. Evaluate the value of Euler's ϕ function $\phi(n) = (p - 1)(q - 1)$ and choose a number $1 < e < \phi(n)$ such that e is coprime with $\phi(n)$ (their maximum common divisor is 1).
3. Find d such that $de = 1 \mod \phi(n)$, that is the inverse of e modulo $\phi(n)$.
4. Alice publishes the numbers (n, e) and Bob then encodes his message m by performing the operation $c = m^e \mod n$.
5. Alice receives the encoded message c and recovers it using $m = c^d \mod n$.

Of course, this whole process will be no-sense if anyone could factorise n, find p and q since d is trivially found, so the message would be exposed.

8.4 BB84

It is important to remember that this protocol is not intended to transmit a message but to transmit a private key which can be used later to encode the real message as explained above. The steps to obtain the key go as follow:

1. Alice generates a sequence of $4n$ random bits $a = (a_1 a_2 \dots a_n)$ and she encodes each bit in a quantum state by choosing randomly one of the two basis $Z = \{0 : |\uparrow\rangle, 1 : |\downarrow\rangle\}$ or $V = (X \pm Z)/\sqrt{2} = \{0 : |\nearrow\rangle, 1 : |\nwarrow\rangle\}$.
2. Bob receives this sequence of states and he measures them by measuring each of them in the basis Z or V, chosen randomly. The result is a sequence of bits $b = (b_1 b_2 \dots b_n)$.
3. Alice and Bob publish the basis they've used to encode and to measure, respectively. They will compare them and discard the bits where Bob measured in a different basis in which Alice prepared it.

   The resulting sequence should have about $2n$ bits, if not they start over again.
4. Alice picks a subset of n bits and compares them with the same Bob's subset. She then evaluates the *Quantum Bit Error Rate* (percentage of error), if it is large than $8'11\%$ we say that there has been an eavesdropper and we start over again.
5. At this point, the key has n bits and can be considered safe but there are some extra steps that increase the privacy.

- Information reconciliation: Alice and Bob choose some even number of bits and evaluate their sum, $c_k + c_j$, if they get the same result then they discard c_k and save b_j again.
- Privacy amplification: Alice and Bob replace some bits c_k by the sum with its adjacent c_{k+1}.

At the end of the whole process, we will end up with a short key as most bits are removed for safety.

8.5 Ekert91

1. Alice and Bob share a set of n entangled states $|\Phi^+\rangle$ generated in a safe source.
2. Alice measures her part of the state in one of the basis $Z^A_{\theta_1}, Z^A_{\theta_2}, Z^A_{\theta_3}$ and similarly for Bob $Z^B_{\varphi_1}, Z^B_{\varphi_2}, Z^B_{\varphi_3}$.
3. They share the bases that they used and put the bits where they agree first.
4. Then they publish the set of bits where they didn't concide and evaluate the correlation function
$$S = E(Z^A_{\theta_1}, Z^B_{\varphi_3}) + E(Z^A_{\theta_1}, Z^B_{\varphi_2}) + E(Z^A_{\theta_2}, Z^B_{\varphi_3}) - E(Z^A_{\theta_2}, Z^B_{\theta_2}) \tag{8.5}$$
Quantum mechanics tells us that $S = 2\sqrt{2}$ (see section 3.5) while classically $S = \sqrt{2}$. In case there was an eavesdropper, the correlation function will have a value $S < 2\sqrt{2}$.
5. If S is close to $2\sqrt{2}$ thye can take the key, if not they must start over again.

9 Quantum Error Correction

In practice, it is very difficult to keep a qubit stable in its state since environment interacts with it and usually it is entangled to it. The environment introduces noise to the system, for instance it can take the state $|0\rangle$ to a superposition $\sqrt{1-p}\,|0\rangle + \sqrt{p}\,|1\rangle$. If the noise is small, $p \ll 1$, we will recover the original state with probability $1-p \approx 1$, but for not so small values of p the uncertainty increases.

This is the reason why we introduce quantum error correction protocols. As a simple example, consider the channel explain previously named *bit flip channel*, pictured in fig. 12. This channel flips the original bit, $|i\rangle \to |1 \oplus i\rangle$, with probability p. A way to reduce the probability of error, we will send 3 identical qubits instead of one, $|i\rangle \to |i\rangle_L = |iii\rangle$. The channel may or may not flip any of the individual states, in any case, whenever we want to recover the state we will measure the 3 states and decide which was our state using the *majority voting* protocol, i.e. choose the value that has more bits. Thus, the error probability is the probability that two or three qubits were flipped so $p_e = 3p^2(1-p) + p^3 = 3p^2 - 2p^3$.

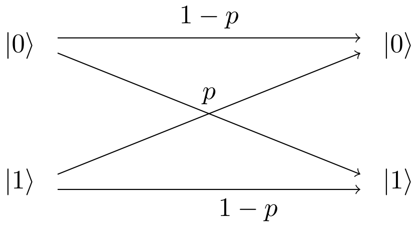

Figure 12: Bit flip channel with flipping probability p.

The previous example is just the same as in classical mechanics, we shall study in the following sections a specific example which only happens in quantum mechanics but the key idea is the same: send multiple copies of the same qubits. So, instead of sending the state $\alpha\,|0\rangle + \beta\,|1\rangle$ we will send $\alpha\,|0\rangle_L + \beta\,|1\rangle_L$ where $\{|0\rangle_L, |1\rangle_L\}$ are called logical qubits. The specific form of the logical qubits depend on the algorithm and the error we want to correct, sometimes they are also called error-correcting qubits. In the following section we will see examples of error-correcting codes and the general approach to quantum error correction.

9.1 Indirect measurement

In the following sections we will need the concept of indirect measurement. We already know that a standard measurement A on the state $|\psi\rangle$ will collapse the state to any of the orthogonal subspaces of A, thus destroying the original state. The indirect measurement is built in order to extract information from a state without destroying it. This process has to be done with the introduction of an ancillary qubit.

For instance, in fig. 13 we see the process that makes an indirect measure on the qubit $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ of the projector P with eigenvalues ± 1.

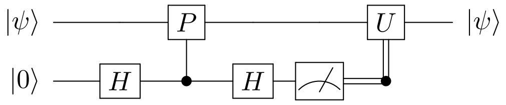

Figure 13: Circuit that implements an indirect measurement of the projector P.

The state of the circuit initially is $|\psi_0\rangle = |\psi\rangle\,|0\rangle$ and after the first Hadamard it turns to $|\psi_1\rangle = |\psi\rangle\,|+\rangle$. Then, the control operation produces the entangled state

$$|\psi_3\rangle = \frac{1}{\sqrt{2}}\left(|\psi\rangle\,|0\rangle + (P\,|\psi\rangle)\,|1\rangle\right) \tag{9.1}$$

It follows that the state prior to the measurement is

$$|\psi_4\rangle = \left(\frac{\mathbb{I}+P}{2}\right)|\psi\rangle|0\rangle + \left(\frac{\mathbb{I}-P}{2}\right)|\psi\rangle|1\rangle \tag{9.2}$$

The form of this state induces us to think that the circuit projects the state into the positive and negative eigenspace of P, depending on the result of the measure. This is very useful since, if we choose a $|\psi\rangle$ such that $P|\psi\rangle = +|\psi\rangle$, the result of the measure will always be 0 unless there has been an error in the circuit. In this situation, we use this information to perform a controlled unitary operation U on the state to recover our original qubit.

The process is clearly seen with an example. Consider $|\psi\rangle = |+\rangle$ and $P = X$. Then, $|\psi_4\rangle = |+\rangle|0\rangle$ and the measure will always trigger $+1$ so no change needs to be made on the state. However, if an error occurred, suppose a Z gate was applied before the application of P then the state will turn to $|-\rangle$ and $|\psi_4\rangle = |-\rangle|1\rangle$. The state has collapsed to the negative eigenspace so the measure will output -1. We have been able to detect the error and now we are able to correct it by applying an $U = Z$ gate to the first register recovering the initial state.

We've seen an example of indirect measurement that induces a simple quantum error correction. This will be useful when we analyse more complicated codes in what follows.

Finally, let us introduce some nomenclature. This process of indirect measurement + quantum error correction using an ancilla qubit is usually called *parity check* as we use the parity of $|\psi\rangle$ under P to detect an error. The gate P is called an *stabiliser* of $|\psi\rangle$ if $P|\psi\rangle = +|\psi\rangle$.

9.2 Bit flip code

Let's make a more systematic analysis of the previous algorithm. We start with our agent Alice sending a state $|\psi\rangle$ using three qubits. Suppose that $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ we will have $|\psi\rangle_L = \alpha|0\rangle_L + \beta|1\rangle_L$. She sends this state through the channel in fig. 12 which potentially flips 1,2,3 or non or the qubits. The error, in practical terms, is introduced when an X gate is applied to any of the three qubits composing our logical qubit. Bob, after receiving this state, performs a measurement to detect if there has been any bit flip and then corrects its state in order to recover the original $|\psi\rangle$.

It is important to note that this system is only capable of recovering the original state if there has been only a flip. Notice that three individual bit flips are required to take $|0\rangle_L \to |1\rangle_L$, hence if we assume $|\psi\rangle_L = |0\rangle_L$, a single bit flip on any qubit leaves the final state closer to $|0\rangle_L$ than $|1\rangle_L$. The distance between two codeword states, d, defines the number of errors that can be corrected, t, as $t = \lfloor (d-1)/2 \rfloor$ Devitt et al. [2013]. In this case, $d = 3$, hence $t = 1$.

There are two ways to proceed now:

A) Apply a POVM $\{Pi_j\}_{i=0}^{3}$ on the state received to detect the position of the flip, assigning the hypothesis that there was a change in the j-th position if the result j is obtain, except for $j = 0$ which we identify with the no flip. The form of these operators is

$$\Pi_0 = |000\rangle\langle 000| + |111\rangle\langle 111| \tag{9.3a}$$
$$\Pi_1 = |100\rangle\langle 100| + |011\rangle\langle 011| \tag{9.3b}$$
$$\Pi_2 = |010\rangle\langle 010| + |101\rangle\langle 101| \tag{9.3c}$$
$$\Pi_3 = |001\rangle\langle 001| + |110\rangle\langle 110| \tag{9.3d}$$

We can check that $\{\Pi_j\}$ form a compete set of orthogonal projectors, moreover, the state post measurement isn't changed at all for any ψ. This allows us to detect the position of the change with 100% certainty and to correct it applying a X gate on the position of the change (or the identity if no flip occurred).

B) The other method is to include an auxiliary system composed of two qubits attached to the state $|\psi\rangle$ (after the possible change) and apply a sequence of C_{NOT} gates to change the state of the ancilla (see fig. 14) to one of the following

$$|\psi_0\rangle = (\alpha\,|000\rangle + \beta\,|111\rangle)\,|00\rangle \tag{9.4a}$$
$$|\psi_1\rangle = (\alpha\,|100\rangle + \beta\,|011\rangle)\,|01\rangle \tag{9.4b}$$
$$|\psi_2\rangle = (\alpha\,|010\rangle + \beta\,|101\rangle)\,|10\rangle \tag{9.4c}$$
$$|\psi_3\rangle = (\alpha\,|001\rangle + \beta\,|110\rangle)\,|11\rangle \tag{9.4d}$$

The four states are orthogonal to each other so we can perform a measure on the Z basis of the last two qubits. This is tel us with certainty the position of the flip and with this information we will be able to recover the state as before, by applying a X gate on that position (or do nothing if no flip occurred).

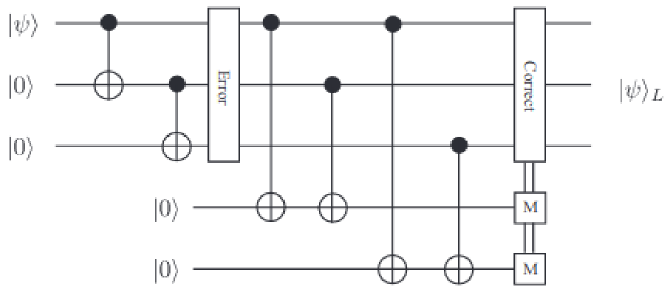

Figure 14: Circuit to encode and correct a single bit flip using an auxiliary system.

C) There is even a third strategy, which is to perform a joint measure of the first and second qubit and of the second and third, that is, $Z_1Z_2 \equiv Z \otimes Z \otimes \mathbb{I}$ and $Z_2Z_3 \equiv \mathbb{I} \otimes Z \otimes Z$. There result of each separately can be either ±1 so in total we have 4 possibilities as desired to distinguish perfectly the 4 cases above.

In the nomenclature introduced before $\{Z_1Z_2, Z_2, Z_3, Z_1Z_3\}$ are stabilisers of the 3-bit flip code, all of them leave the state $|\psi\rangle_L$ invariant. Moreover, since they are projectors with eigenvalues ±1 we can use them as an indirect measure as a way to detect a possible bit flip in any position. Note that, we don't need to measure the three observables, but with only two we cover all the possible outcomes.

The analysis is made simpler by writing the operators explicitly:

$$Z_1Z_2 = (|00\rangle\langle 00| + |11\rangle\langle 11|) \otimes \mathbb{I} - (|01\rangle\langle 01| + |10\rangle\langle 10|) \otimes \mathbb{I} \tag{9.5a}$$
$$Z_2Z_3 = \mathbb{I} \otimes (|00\rangle\langle 00| + |11\rangle\langle 11|) - \mathbb{I} \otimes (|10\rangle\langle 10| + |01\rangle\langle 01|) \tag{9.5b}$$

Essentially, Z_iZ_j will output +1 if the states i and j are the same, and −1 if the states are different. Therefore, by knowing these output we can identify the syndrome and correct it (see table below).

Table 1: Syndrome detection and correction procedure for a bit flip error-correcting code.

Z_1Z_2	Z_2Z_3	Syndrome	Correction
$+1$	$+1$	No error	$\mathbb{I}$
$+1$	-1	Error 3rd qb	$\mathbb{I} \otimes \mathbb{I} \otimes X$
-1	$+1$	Error 1st qb	$X \otimes \mathbb{I} \otimes \mathbb{I}$
-1	-1	Error 2nd qb	$\mathbb{I} \otimes X \otimes \mathbb{I}$

Essentially, this three methods are equivalent, only that sometimes it is easier to implement the later as it is made of known and standard gates. The former doesn't need to add extra states to the analysis but we would have to find a way to implement those projectors which is a non-trivial task.

What is crucial to the success of these measurements is that neither measurement gives any information about the amplitudes of α and β of the encoded quantum states, and thus neither measurement destroys the superposition of quantum states that we with to preserve using the code.

This error-correction procedure works perfectly, provided bit flips occur on one or fewer qubits which happens with probability $(1-p)^3 + 3p(1-p)^2 = 1 - (3p^2 - 2p^3)$. The probability of an error remaining uncorrected is therefore $3p^2 - 2p^3$, just as the classical *majority voting* protocol before.

9.3 Phase flip code

This channel introduces a phase to the state $|0\rangle$ and $|1\rangle$ with probability p according to the eigenvalue of Z, i.e. $|i\rangle \to Z\,|i\rangle$. Obviously, if the state send is one of the eigenvectors of Z nothing will happen but as soon as we send a superposed state we have that $\alpha\,|0\rangle + \beta\,|1\rangle \to \alpha\,|0\rangle - \beta\,|1\rangle$ up to the extreme case that $|+\rangle \to |-\rangle$, and vice-versa.

This gives us an intuition of this problem that can help us solve it without any trouble as the phase flip error in the Z basis can be interpreted as a bit flip error in the X basis. This suggest using the states $|0\rangle_L = |+++\rangle$ and $|1\rangle_L = |---\rangle$ as logical states. Then, we replace the measurement projectors by $\Pi'_j = H^{\otimes 3}\Pi_j H^{\otimes 3}$ and run the protocol as before.

Similarly, we can apply this transformation to the 3rd case, so instead of having measurements in the Z basis we will measure in the X basis and proceed as in table 1 (whenver you see a Z replace it with a X, and vice-versa).

We can say that the bit flip and the phase shift are unitary equivalent, since there is a unitary operator U such that the action of one channel is the same as the other, provided the first channel is preceded by a U and followed by a $U^\dagger$.

9.4 9 qubit Shor code

Shor [1995] came up with the idea to use a coding scheme of 9 qubits in order to correct any arbitrary error on a single qubit. It makes use of the three qubits bit flip and phase flip codes, which are able to correct an X or Z transformation on a single qubit[*].

The qubits are first encoded using the phase flip code: $|0\rangle \to |+++\rangle$ and $|1\rangle \to |---\rangle$; and then, each of the separate qubits is replicated into three copies as in the bit flip code. The logical qubits are

$$|0\rangle_L \equiv \frac{1}{2\sqrt{2}}\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right) \tag{9.6a}$$

$$|1\rangle_L \equiv \frac{1}{2\sqrt{2}}\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right) \tag{9.6b}$$

each composed of 9 qubits.

The detection of a bit flip in any qubit is made using any of the three methods explained above. Here, we will stick to the third case at it doesn't require the introduction of auxiliary qubits or any POVM. We measure them in pairs of consecutive qubits via Z_jZ_{j+1} from $j = 1, \ldots, 8$ obtaining 2^8 possible results. For example, if a bit flip occurred in the 5th qubit we will obtain a -1 for the measurement operator Z_4Z_5 and -1 for Z_5Z_6 leading to the conclusion that there was a bit flip error in the 5th qubit (2nd qubit of the 2nd group).

Instead, if there was a phase flip error. a change of sing would have happened in any of the three groups. The detection is made by measuring in the observables $\{X_1X_2X_3X_4X_5X_6, X_4X_5X_6X_7X_8X_9\}$. Again, each observable separately will output $+1$ if the sign is the same in the two groups and -1 if the sign is different. The exact position of the phase flip can be extracted (just as in table 1) and corrected by applying $Z_{3i-2}Z_{3i-1}Z_{3i}$ on the group $i \in \{1, 2, 3\}$ that suffered the change.

Because both measurements are non-destructive, we can concatenate both and correct any bit flip error on a qubit and any phase flip error. thus, the Shor code enables the correction of combined bit and phase flip errors on a singe qubit.

[*] Remember that, since the Pauli matrices do not commute, we can produce a Y rotation with a combination of X and Z rotations. Therefore, a code that can correct both X and Z rotations also corrects Y rotations.

In fact, this code is more powerful than that as it can correct any arbitrary small rotation in the Bloch sphere with just two discrete correction codes for specific errors [see Nielsen and Chuang, 2010, Section 10.2].

9.5 General quantum error correction protocol

A quantum error correction protocol generally consist on 3 stages:

1. A quantum state is encoded into a quantum error-correcting code, formally it is encoded into some subspace C of a larger Hilbert space. We introduce the projector P which maps any state to this subspace.

   For instance, in the bit flip case, the subspace C was composed of two states $\{|000\rangle, |111\rangle\}$ that are used to encode all the qubits. The projector onto this subspace is just $P = |000\rangle\langle 000| + |111\rangle\langle 111|$.

2. The noise acted on the system and we to perform a measurement to diagnose the type of error, that is, the error syndrome. It is important that the measurement distinguishes without error all the types of syndromes, the projectors forming the POVM are orthogonal, otherwise the next recovery step won't work.

3. Finally, based on the outcome of the measure, a recovery operation is performed on the state to return it to the original one.

With this said, we suppose that the noise is described by a quantum operation $\mathcal{E}$ and the complete error-correction procedure is effected by a trace-preserving quantum operation $\mathcal{R}$. Then, if the quantum error correction is successful we should have that, for all ρ, to following condition is satisfied

$$(\mathcal{R} \circ \mathcal{E})(\rho) \propto \rho \tag{9.7}$$

where the reason for the $\propto$ is that we don't require $\mathcal{E}$ to be trace-preserving, thus allowing to a more general set of operations like measurement. We are ready to announce the following theorem that gives us a way to construct $\mathcal{R}$ from $\mathcal{E}$.

Theorem 11. *Let C be a quantum code and let P be a projector onto C. Suppose $\mathcal{E}$ is a quantum operation with elements $\{E_i\}$. a necessary and sufficient condition for the existence of an error-correction operation $\mathcal{R}$ correction $\mathcal{E}$ on C is that*

$$PE_i^\dagger E_j P = \alpha_{ij} P \tag{9.8}$$

for some hermitian matrix α.

If such an $\mathcal{R}$ exist, we say that $\{E_i\}$ constitutes a correctable set of errors. For instance, take Shor's code for 9 qubits that is said to correct any type of single qubit error. The most general form of such quantum operation $\mathcal{E}$ can be written with operation elements $\{E_i\}$ that are a linear combination of the Pauli matrics and the identity,

$$E_i = e_{i0}\mathbb{I} + e_{i1}\sigma_1 + e_{i2}\sigma_1\sigma_3 + e_{i3}\sigma_3$$

where we have used that $\sigma_1\sigma_3 \propto \sigma_2$. Therefore, to check that the Shor code corrects against arbitrary single qubit errors on the k-th qubit it is sufficient to verify that the equations

$$P\sigma_i^k \sigma_j^k P = \alpha_{ij} P \qquad i = 0, 1, 2, 3 \tag{9.9}$$

are satisfied. The projector P for the Shor code is given by $P = |0_L\rangle\langle 0_L| + |1_L\rangle\langle 1_L|$ with $|0\rangle_L$ and $|1\rangle_L$ as in eq. (9.6).

9.6 Fault-tolerant quantum computation

In our discussion so far, we have assumed that we can encode quantum information and perform recovery from errors without making any mistakes. But,of course, error recovery will not be flawless. Recovery is itself a quantum computation that will be prone to error. If the probability of error for each bit in our code block is ϵ, then it is reasonable to suppose that each quantum gate that we employ in the recovery procedure has a probability of order ϵ of introducing an error. If our recovery procedure is carelessly designed, then the probability that the procedure fails (e.g., because two errors occur in the same block) may be of order ϵ. Then we have derived no benefit from using a quantum error-correcting code; in fact, the probability of error per data qubit is even higher than without any coding. So we are obligated to consider systematically all the possible ways that recovery might fail with a probability of order ϵ, and to ensure that they are all eliminated. Only then is our procedure fault tolerant, and only then is coding guaranteed to pay off once ϵ is small enough [Preskill, 1998].

The basic idea of fault-tolerant quantum computation is to compute directly on encoded quantum states in such a manner that decoding is never required. We replace each qubit with an encoded block of qubits and each gate with a procedure for performing an encoded gate acting on the encoded state. The problem comes when one of this gates fails and the error propagates through the code. We define the fault-tolerance of a procedure to be the property that if only one component in the procedure fails then the failure causes at most one error in each encoded block of qubits output from the procedure. By component we understand gates, measurement, wires, state preparation...

We say that a gate U is *transversal* on a code C if its action before encoding the state leaves the same result as applying individual qubit gates after encoding. In fig. 15 we see a general scheme of a gate acting on a codified state. If the logical gate $U_L = U^{\otimes n}$, where n is the number of qubits forming the code, then the gate is transversal, otherwise U_L can have a totally different form. In any case, this logical gate has to verify the equality $U_L \left|\psi\right\rangle_L = C(U \left|\psi\right\rangle)$ giventhe original state $\left|\psi\right\rangle$ and the original gate U.

$$\begin{tikzcd}
|\psi\rangle \arrow[r, "C"] \arrow[d, "U"] & |\psi\rangle_L \arrow[d, "U_L"] \\
U|\psi\rangle \arrow[r, "C"] & U_L|\psi\rangle_L = C(U|\psi\rangle)
\end{tikzcd}$$

Figure 15: Application of a fault tolerant gate.

For example, the Z and X gates are transversal under the 3-qubit code but H is not. The importance of tranversal gates is that they do not propagate the error to other qubits, however not all universal gates are transversal. This is in contrast to non-transversal operators, where, for example,an encoded gate coupling every subsystem in a code block might convert an error on a single subsystem into an error on every subsystem of the code block.

It would be nice if there exist a universal set of gate that are transversal under a certain error-correcting code C allowing fault-tolerant quantum computation but the following theorem by Eastin and Knill [2009] forbids this.

Theorem 12. *For any nontrivial local-error-detecting quantum code, the set of transversal, logical unitary operators is not universal.*

9.7 Threshold theorem and concatenated codes

The idea behind a concatenated code is to reduce the effective error by recursively applying error correcting codes. In the first stage, each qubit is encoded in a quantum code C_0 whose qubits are themselves encoded in a quantum code C_1, those into C_2 and so forth. Shor's code is an example of concatenated code in two stages.

Suppose that our code C takes one qubit to c qubits, if the failure probability of each component is p then the probability of failure after the first encoding is $cp \cdot p = cp^2$. After, two encoding procedures, the error becomes $c(cp^2)^2$ and so after k concatenations the error is $(cp)^{2^k}/c$.

If our systems grows polynomially with the input size n, that is $p(n)$, and we wish to achieve an accuracy ϵ in the final result, each gate in our algorithm must have an accuracy $\epsilon/p(n)$ so

$$\frac{(cp)^{2^k}}{c} \leq \frac{\epsilon}{p(n)} \tag{9.10}$$

Only when, $p < p_{th} = 1/c$ this k can be found. The condition $p < p_{th}$ is known as the *threshold condition* since provided it is satisfied we can achieve arbitrary accuracy in our quantum computation.

The number of gates needed to simulate a circuit with $p(n)$ gates up to an accuracy ϵ is

$$\mathcal{O}\left(\text{poly}\left(\log p(n)/\epsilon\right) p(n)\right) \tag{9.11}$$

provided that $p < p_{th}$.

9.8 Entanglement distillation

Teleportation and dense coding, two of the most important protocols in quantum communication, rely on having a maximally entangled state shared between two parties Alice and Bob. However, thos protocols as explained in this text, assume that the Alice sends one part of her bipartite state to Bob without error. Sadly, this is an ideal process, but in real life we are faced with errors on the communication line.

Entanglement distillation is a process by which two separate observes, by applying local operations to a supply of not-too-impure entangled states can prepare a smaller number of entangled pairs of arbitrarily high purity [Bennett et al., 1996]. Alice will prepare n copies of a bipartite state ρ and sends the second parts to Bob. In any case, a single copy of those states can be expressed as a Werner state

$$\rho = W_F = F\,|\Phi_+\rangle\langle\Phi_+| + \frac{1-F}{3}\left(|\Psi_+\rangle\langle\Psi_+| + |\Psi_-\rangle\langle\Psi_-| + |\Phi_-\rangle\langle\Phi_-|\right) \tag{9.12}$$

since the Bell states form a basis of $\mathcal{H}_2^{\otimes 2}$, where F is the fidelity with respect to the state $|\Phi_+\rangle$, the state we need to have in order to apply successful communication.

References

E. Andersson, S. M. Barnett, C. R. Gilson, and K. Hunter. Minimum-error discrimination between three mirror-symmetric states. *Phys. Rev. A*, 65:052308, Apr 2002. doi: 10.1103/PhysRevA.65.052308.

J. Bae and L.-C. Kwek. Quantum state discrimination and its applications. *Journal of Physics A: Mathematical and Theoretical*, 48(8):083001, 2015.

E. Bagan, R. Muñoz-Tapia, G. Olivares-Rentería, and J. Bergou. Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes. *Physical Review A*, 86(4):040303, 2012.

S. M. Barnett. Minimum-error discrimination between multiply symmetric states. *Phys. Rev. A*, 64:030303, Aug 2001. doi: 10.1103/PhysRevA.64.030303.

S. M. Barnett and S. Croke. Quantum state discrimination. *Adv. Opt. Photon.*, 1(2):238–278, Apr 2009. doi: 10.1364/AOP.1.000238.

J. S. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science*, 560:7–11, 2014.

C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76(5):722–725, Jan 1996. ISSN 1079-7114. doi: 10.1103/physrevlett.76.722. URL http://dx.doi.org/10.1103/PhysRevLett.76.722.

A. Chefles and S. M. Barnett. Optimum unambiguous discrimination between linearly independent symmetric states. *Physics Letters A*, 250(4):223 – 229, 1998. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(98)00827-5.

C.-L. Chou. Minimum-error discrimination among mirror-symmetric mixed quantum states. *Phys. Rev. A*, 70:062316, Dec 2004. doi: 10.1103/PhysRevA.70.062316.

E. Davies. Information and quantum measurement. *IEEE Transactions on Information Theory*, 24(5):596–599, 1978.

M. E. Deconinck and B. M. Terhal. Qubit state discrimination. *Physical Review A*, 81(6):062304, 2010.

S. J. Devitt, W. J. Munro, and K. Nemoto. Quantum error correction for beginners. *Reports on Progress in Physics*, 76(7):076001, 2013.

B. Eastin and E. Knill. Restrictions on transversal encoded quantum gate sets. *Physical review letters*, 102(11):110502, 2009.

A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 5 1935. doi: 10.1103/PhysRev.47.777.

C. A. Fuchs. *Distinguishability and accessible information in quantum theory*. PhD thesis, University of New Mexico, 1996.

C. A. Fuchs. Quantum mechanics as quantum information (and only a little more). *arXiv preprint quant-ph/0205039*, 2002.

C. A. Fuchs and C. M. Caves. Ensemble-dependent bounds for accessible information in quantum mechanics. *Physical Review Letters*, 73(23):3047, 1994.

C. A. Fuchs and A. Peres. Quantum theory needs no ‘interpretation’. *Physics Today*, 53(3):70–71, 2000.

N. Gisin. Quantum cloning without signaling. *Physics Letters A*, 242(1):1 – 3, 1998. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(98)00170-4.

M. Gu. *Physics of Information Lecture Notes.* NTU, 2018.

D. Ha and Y. Kwon. Complete analysis for three-qubit mixed-state discrimination. *Physical Review A*, 87(6):062302, 2013.

C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2): 231–252, 1969.

A. Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3 (4):337 – 394, 1973a. ISSN 0047-259X. doi: https://doi.org/10.1016/0047-259X(73)90028-6.

A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973b.

K. Maruyama, F. Nori, and V. Vedral. Colloquium: The physics of Maxwell's demon and information. *Reviews of Modern Physics*, 81:1–23, Jan. 2009. doi: 10.1103/RevModPhys.81.1. https://arxiv.org/abs/0707.3400.

M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2010. doi: 10.1017/CBO9780511976667.

A. Peres. *Quantum theory: concepts and methods*, volume 57. Springer Science & Business Media, 2006.

J. Preskill. Fault-tolerant quantum computation. In *Introduction to quantum computation and information*, pages 213–269. World Scientific, 1998.

P. Raynal. Unambiguous state discrimination of two density matrices in quantum information theory. *arXiv preprint quant-ph/0611133*, 2006.

T. Rudolph, R. W. Spekkens, and P. S. Turner. Unambiguous discrimination of mixed states. *Physical Review A*, 68(1):010301, 2003.

J. J. Sakurai and J. Napolitano. *Modern Quantum Mechanics.* Cambridge University Press, 2 edition, 2017. doi: 10.1017/9781108499996.

P. W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical review A*, 52(4):R2493, 1995.

G. Weir, S. M. Barnett, and S. Croke. Optimal discrimination of single-qubit mixed states. *Phys. Rev. A*, 96:022312, Aug 2017. doi: 10.1103/PhysRevA.96.022312.