Quantum information 116031 – Lecture Notes

J.E. Avron[1]

June 22, 2021

[1]Comments and typos welcome. Send to avronj@gmail.com.

Contents

Chapter 1

Qubit

1.1 Bit and Qubit

The binary digit, bit in short[1], is the currency classical information: $b \in \{0, 1\}$. It describes the state of a classical two state system, such as a coin. Strictly speaking, there are no bits in classical physics since all observables, position, momenta, angles, are continuous, not discrete. A classical bit is an idealized notion. More important is the fact that classical physics is itself only an approximate theory.

The underlying physical theory is quantum. In contrast with classical physics quantum mechanics offers observables with discrete values. But, it is not quite a bit.

A qubit is a *two state quantum system*. It too is an idealized notion. Often a qubit is a physical system with two nearly or better strictly degenerate energy levels which are almost isolated from the rest of the world. Examples are:

- The two polarization states of photon
- The two spin states of the electron or nucleon
- Two optical cavities
- Any two isolated modes in atoms, ions, or Josephson junctions (whatever this is).

The qubi is the currency of quantum information.

The important difference between a bit and a qubit is superposition. An ideal classical coin can be **either** up **or** down. Schrödinger cat can be **both** dead and alive.

[1] Popularized by Shanon who attributed it to Tukey. We'll meet both later.

Figure 1.1: At low energies, the two lowest energy levels act like an approximate qubit

1.2 The Hilbert space of a qubit

The Hilbert space of a qubit is $\mathbb{C}^2$. We can pick a basis $\mathbb{C}^2$ anyway we want. But anu such basis will have two basis vectors which we denote

$$|a\rangle, \quad a \in \mathbb{Z}_2$$

I shall consistently denote by $|a\rangle$ with a Roman letters, basis vectors and by $|\psi\rangle$, with Greek letters, a general state, which is a superposition in this basis:

$$|\psi\rangle = \sum_{a\in\mathbb{Z}_2} \psi_a |a\rangle, \quad \psi_a \in \mathbb{C}, \quad \langle\psi|\psi\rangle = 1 \tag{1.1}$$

Normalized vectors $|\psi\rangle$ in the Hilbert space are called pure states. The pure $|\psi\rangle$ lie on the unit sphere:

$$1 = \langle\psi|\psi\rangle = |\psi_0|^2 + |\psi_1|^2 \tag{1.2}$$

Geometrically, this is S^3, the 3-sphere in 4-D.

Since $|\psi\rangle$ and $e^{i\gamma}|\psi\rangle$ are physically indistinguishable there is a circle in S^3 that represent the same physical state. The space of physically distinct states is therefore

$$S^3/S^1 \sim S^2$$

Geometrically, this is the 2-sphere in 3-D, which we can easily visualize. This space of distinct states of a qubit is known as $CP(1)$.

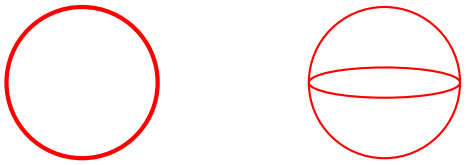


Figure 1.2: Left: S^1 in $\mathbb{R}^2$ is given by $x^2 + y^2 = 1$. Right: S^2 in $\mathbb{R}^3$ is given by $x^2 + y^2 + z^2 = 1$, etc.

1.3 Projections

A nice way to get rid of the redundant overall phase in $|\psi\rangle$ is to look at the associated projection

$$P_\psi = |\psi\rangle\langle\psi| \tag{1.3}$$

Clearly

$$P_\psi^2 = P_\psi$$

which simply expresses the fact that P_ψ is a projection. Since $P_\psi = P_\psi^\dagger$ we may also view P_ψ as an observable. It has eigenvalues 1 and eigenvalue 0

$$P_\psi|\psi\rangle = |\psi\rangle, \quad P_\psi|\psi_\perp\rangle = 0$$

where $|\psi_\perp\rangle$ is a state orthogonal to $|\psi\rangle$.

As a general rule, we shall only measure projections, so the result of the measurement will always be either $a \in \mathbb{Z}_2$. This fact does not depend on the state of the qubit: If the qubit is in the state $|\psi\rangle$ and we measure the projection P_ϕ then we get 0 and 1 with probabilities:

$$Prob(1) = \langle\psi|P_\phi|\psi\rangle = |\langle\psi|\phi\rangle|^2, \quad Prob(0) = \langle\psi|(\mathbb{1} - P_\phi)|\psi\rangle = 1 - |\langle\psi|\phi\rangle|^2$$

This makes a qubit like a bit: You always find either head or tail.

You can think of a projection concretely as representing a detector, and the 1 means the detector clicks and 0 that it does not.

There are S^2 worth of projections. If you think of the qubit as spin, you may think of these as measuring the direction in which the spin is pointing. The same is true for any 2-level system only that the direction is a direction in Hilbert space rather than in physical coordinate space.

Exercise 1.1. *Show that if H is a Hermitian operator so that $H^2 = \mathbb{1}$ then*

$$P_\pm = \frac{\mathbb{1} \pm H}{2}$$

are orthogonal projections.

1.4 The computational basis

Qubit comes with a distinguished basis: The basis in which we are supposed to measure, or read the qubit. For example, if we had a quantum computer and wanted to read its output, we need to be told a-priori in what basis to read the output. This basis is called the *computational basis*.

In some applications we shall read certain qubits in one basis and other qubits in another basis. In some cases, e.g. in cryptography, the choice of basis will be a secret that we share with people we trust.

The computational basis is sometimes associated with the Hamiltonian of the (isolated) qubit. Since energy eigenstates evolve in time by

$$|a\rangle \mapsto e^{-iE_a t}|a\rangle$$

and the rules of QM say that states that differ by an overall phase are physically indistinguishable, the energy eigenstates behaves like a classical bit in the sense that once you set the bit to be $|a\rangle$ it remains $|a\rangle$ until you decide to operate on it (e.g. by changing the Hamiltonian).

1.5 Mixed states

$|\psi\rangle$ gives a complete description of a quantum state. Suppose that the state is not completely specified. For example, you have a source that with probability p_j gives the state $|\psi_j\rangle$. This situation of incomplete knowledge is described by a *density matrix*

$$\rho = \sum p_j |\psi_j\rangle\langle\psi_j|$$

Clearly

$$\begin{aligned} Tr\,\rho &= Tr\,\left(\sum_j p_j |\psi_j\rangle\langle\psi_j|\right) \\ &= \sum_j p_j Tr\,|\psi_j\rangle\langle\psi_j| \\ &= \sum_j p_j \langle\psi_j|\psi_j\rangle \\ &= \sum_j p_j = 1 \end{aligned}$$

ρ, being the convex combination of positive operators (projections) is a positive operator:

$$\rho \geq 0 \tag{1.4}$$

The special case

$$\rho_{1/2} = \frac{\mathbb{1}}{2}$$

is called fully mixed. We know nothing about the state of the qubit. We only know we have a qubit.

1.5.1 Classical probability theory

Density matrices allow to describe classical probability theory with the bra and ket notation. A classical coin with probability p of being up and $1-p$ down, is described by

$$\rho_p = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$$

The case of complete ignorance corresponds to $p = 1/2$. In this case

$$\rho_{1/2} = \frac{\mathbb{1}}{2}$$

Exercise 1.2. *Write the density matrix for a quantum cat in a dead-alive superposition*

$$\rho_\pm = |\pm\rangle\langle\pm|, \quad |\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}} \tag{1.5}$$

and compare it with the dead or alive density matrix.

	Quantum probabilities	Classical probabilities
Space	$\underbrace{\mathcal{H}}_{\text{Hilbert space}}$	$\underbrace{\Omega}_{\text{Sample space}}$
Information on system	$\underbrace{\rho \geq 0}_{\text{positive matrix}}$	$\underbrace{p_j \geq 0}_{\text{positive function}}$
Observable	Matrix (operators) on $\mathcal{H}$	Functions on Ω
Events	Projections	points j in Ω
Expectation	$\langle F\rangle = Tr(F\rho)$	$\langle F\rangle = \sum_{j\in\Omega} F_j p_j$
Independence	$\rho_A \otimes \rho_B$	$p_A p_B$
Mixtures of systems	$\underbrace{\sum p_{jk}\rho_j \otimes \rho_k}_{separable}$	$\underbrace{p(j,k) = \sum p_{jk}\delta_{jj'}\delta_{kk'}}_{general}$

Table 1.1: The table compares classical probability with quantum probabilities

1.6 Pauli gates

Define the operators Z and X, in the computational basis, by

$$Z|a\rangle = (-1)^a|a\rangle \quad X|a\rangle = |a \oplus 1\rangle$$

where $1 \oplus 1 = 0$. These are denoted graphically

$$|\psi\rangle \text{ ——}\boxed{Z}\text{—— } Z|\psi\rangle \text{ , } |\psi\rangle \text{ ——}\boxed{X}\text{—— } X|\psi\rangle$$

Z and X are both unitary and hermitian. Since Z is diagonal in the computational basis, it identifies the computational basis. The projection

$$\underbrace{\frac{\mathbb{1} - Z}{2}}_{\text{projection}} |a\rangle = a|a\rangle$$

measures the qubit in the computational basis.

X is the NOT operation that flips the qubit.

It is easy to see that

$$Z^2 = X^2 = \mathbb{1}, \quad XZ + ZX = 0$$

Since X and Z anti-commute their product is anti-hermitian and so

$$Y = iXZ$$

is Hermitian. I is also unitary. It is easy to see that

$$Y^2 = \mathbb{1}, \quad YX + XY = YZ + ZY = 0$$

The eigenstate of X are denoted by

$$X|+\rangle = |+\rangle, \quad X|-\rangle = -|-\rangle$$

where

$$\sqrt{2}|+\rangle = |0\rangle + |1\rangle, \quad \sqrt{2}|-\rangle = |0\rangle - |1\rangle$$

In physics X, Y, Z are known as the Pauli matrices

$$X = \sigma_x, \quad Y = \sigma_y, \quad Z = \sigma_z$$

A representation is

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = i \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.6}$$

The Pauli group has 16 elements and is given by

$$\pm X, \quad \pm Y, \quad \pm Z, \quad \pm iX, \quad \pm iY, \quad \pm iZ, \quad \pm \mathbb{1}, \quad \pm i$$

1.7 Bloch ball

The eigenvalues of the 2×2 matrix ρ are given by the solution of the quadratic equation

$$\det(\lambda \mathbb{1} - \rho) = \lambda^2 - \lambda Tr\, \rho + \det \rho = 0,$$

ρ is positive if both the trace and the determinant are positive.

Write

$$\rho = \frac{\rho_0 \mathbb{1} + \mathbf{r} \cdot \boldsymbol{\sigma}}{2}, \quad \mathbf{r} = (x, y, z), \quad \mathbf{r} \cdot \boldsymbol{\sigma} = xX + yY + zZ$$

We can interpret ρ as a state if

$$Tr\rho = \rho_0 = 1 \quad \text{and} \quad 4 \det \rho = \rho_0^2 - \mathbf{r} \cdot \mathbf{r} \geq 0$$

It follows that quantum states of a qubit are described geometrically by the unit ball in 3-D

$$\rho(\mathbf{r}) = \frac{\mathbb{1} + \mathbf{r} \cdot \boldsymbol{\sigma}}{2}, \quad |\mathbf{r}| \leq 1$$

This is the Bloch ball.

The interior of the ball $|\mathbf{r}| < 1$ describes mixed states and its bundary , the sphere, $|\mathbf{r}| = 1$ describes the pure states. If $|\mathbf{r}| > 1$ the matrix ρ is not positive and does not describe states.

Exercise 1.3. *Show that*

$$Tr\big(\rho(\mathbf{r})\rho(\mathbf{r}')\big) = \frac{1 + \mathbf{r} \cdot \mathbf{r}'}{2} \tag{1.7}$$

In particular, antipodal points on the Bloch sphere represent orthogonal vectors in Hilbert space; Orthogonal states in the Hilbert space are represented by anti-parallel Euclidean vectors.

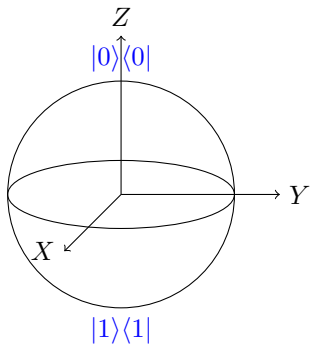


Figure 1.3: Bloch ball: Pure states lie on the surface, the Bloch sphere. Mixed states lie in the interior. The center is the fully mixed state. Orthogonal pure states are antipodal points on the surface.

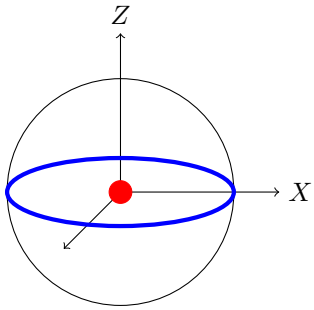


Figure 1.4: The red dot at the center represents the "Dead or alive" state. Every point on the equator represents a superposition representing the states that are both 'Dead and alive".

1.7.1 Poincare sphere– Polarization

The electric field of right circularly polarized plane wave propagating along the z-axis is

$$|R\rangle = (\hat{\mathbf{x}} + i\hat{\mathbf{y}})e^{i\phi}, \quad \phi = kz - \omega t$$

A dictionary between polarization states and qubit states

$$|R\rangle \mapsto |0\rangle, \quad |L\rangle \mapsto |1\rangle,$$

where R and L correspond to right and left circular polarization.

$$\sqrt{2}|R\rangle = |H\rangle + i|V\rangle, \quad \sqrt{2}|L\rangle = |H\rangle - i|V\rangle$$

where H and V correspond to horizontal and vertical polarization. Alternatively,

$$|H\rangle \mapsto |+\rangle, \quad |V\rangle \mapsto |-\rangle$$

With this identification the circular polarization are associated with the poles of the Bloch sphere and the Horizontal and Vertical polarization lie at antipodal points on the equator.

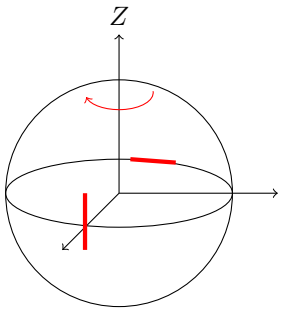


Figure 1.5: Poincare ball. The poles are circular polarizations. The equator represents linear polarization. The center of the ball represents unpolarized light. Antipodal points on the surface are orthogonal states.

Exercise 1.4. *Find the location of the two diagonal polarizations D and $\bar{D}$ on the Bloch sphere. Show that the equator is the locus of linear polarizations.*

1.8 State preparation

Consider a polarizer, a bit like what you have in sun-glasses, that perfectly transmits a photon if it is $|H\rangle$ and perfectly absorbs it if it is $|V\rangle$. The action of the polarizer is described by the projection $P_H = P_{|+\rangle}$. If a photon gets through we call this a success and we have prepared the state $|+\rangle$. If it failed, we lost the photon.

More generally consider the projection P_ψ which prepares the state ψ. If we start with a qubit in state ρ then success has probability

$$Tr(\rho P_\psi) = \langle\psi|\rho|\psi\rangle$$

and failure has the probability

$$Tr\big(\rho(\mathbb{1} - P_\psi)\big) = 1 - \langle\psi|\rho|\psi\rangle$$

Note that no matter what ρ is only $|\psi\rangle$ can be prepared.

There are to extreme special cases worth noting

- If ρ is fully mixed the success probability is $1/2$ no matter what ψ is

- If $\rho = |\psi\rangle\langle\psi|$ the probability of success is 1. The measurement is "non-demolition".

Exercise 1.5. *Show that*

$$Tr(P_\psi P_\phi) = |\langle\psi|\phi\rangle|^2 = \frac{1}{2} + \frac{\mathbf{r}_\psi \cdot \mathbf{r}_\phi}{2} \tag{1.8}$$

1.9 Mixtures have multiple decomposition

You can always write a mixed state in the basis of its eigenvectors uniquely

$$\rho = \sum p_j |\psi_j\rangle\langle\psi_j|, \quad \langle\psi_j|\psi_k\rangle = \delta_{jk}$$

This correspond pictorially to representing a point in the Bloch sphere by two antipodal pure states.

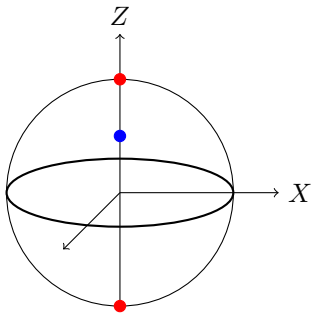


Figure 1.6: The blue dot on the z axis is a weighted sum of 2 red dots, representing pure states at the poles.

However you can get the same state ρ by mixing more than 2 pure states as illustrated in the figure.

The moral of this is that in general, given a state ρ you can not tell from how many pure states it has been constructed, and which states they were.

1.10 Tomography of a qubit

One of the first things you learn in QM is that if you are given and unknown state $|\psi\rangle$, you can not find out $|\psi\rangle$ by measuring the state since measurements in QM prepare from $|\psi\rangle$ a new state: An eigenvector of the measured observable. The situation changes if you happen to have an large supply of identical quantum systems that are all in the same state.

Suppose you have an unlimited supply of a single identical qubits all in the same unknown state ρ. You can determine the state by measuring the

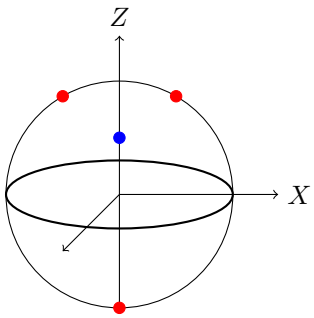


Figure 1.7: Now the red dot on the z axis is expressed as a weighted sum of 3 pure states.

expectation values of

$$Tr\ (X\rho), \quad Tr\ (Y\rho), \quad Tr\ (Z\rho),$$

From which you can reconstruct ρ. This is quantum tomography.

Example 1.6. *It is instructive to compare the density matrix describing a superposition (of a dead and alive cat) with the density matrix of a cat in an unknown dead or alive state:*

$$\rho_{sup} = \underbrace{\frac{1}{2}\begin{pmatrix} 1 & e^{i\phi} \\ e^{-i\phi} & 1 \end{pmatrix}}_{dead\ and\ alive}, \quad \rho_{mix} = \underbrace{\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{dead\ or\ alive}$$

If you are only allowed to measure Z you can not make full tomography and can nor distinguish between a mixture and a superposition. You can not distinguish a qubit from a bit.

Remark 1.7. *This is what happens in measurements: A measurement apparatus is macroscopic object with many degrees of freedom that are in practice inaccessible. This means that for the system+apparatus you can not distinguish superpositions from mixtures.*

Chapter 2

Single qubit gates

Isolated quantum states evolve unitarily. We shall write this graphically as

$$|\psi\rangle \longrightarrow \boxed{U} \longrightarrow U|\psi\rangle$$

The action of U on the (computational) basis is

$$U|a\rangle = \sum_b \underbrace{U_{ba}|b\rangle}_{\text{note order}} \iff U_{ba} = \langle b|U|a\rangle \tag{2.1}$$

U acts on density matrices by conjugation

$$\rho \mapsto U\,\rho\,U^\dagger$$

Unitaries map states to states

$$\rho \geq 0 \Leftrightarrow U\rho U^\dagger \geq 0, \quad tr\,\rho = 1 \Leftrightarrow tr\,(U\rho U^\dagger) = 1$$

Hence, U maps the Bloch ball to itself. It also maps pure states to pure states, and leaves the fully mixed state invariant because

$$\rho^2 = \rho \Leftrightarrow (U\rho U^\dagger)^2 = U\rho U^\dagger, \quad U\mathbb{1}U^\dagger = \mathbb{1},$$

2.1 Unitaries: Rigid rotations of Bloch ball

We define the scalar product between states by

$$Tr\rho\rho'$$

Unitary transformations preserve the scalar product

$$Tr\,(\rho\rho') = Tr\,(U\rho U^\dagger)\,(U\rho' U^\dagger)$$

As a consequence, they also preserve the distance

$$Tr(\rho - \rho')^2$$

As we shall now show unitary transformations corresponds to a rigid rotation of the Bloch ball.

To see this write:

$$\rho - \rho' = \frac{(\mathbf{r} - \mathbf{r}') \cdot \boldsymbol{\sigma}}{2}$$

Using the fact that

$$Tr\big(\mathbf{x} \cdot \boldsymbol{\sigma}\, \mathbf{y} \cdot \boldsymbol{\sigma}\big) = x_j x_k\, Tr(\sigma_j \sigma_k) = 2\, \mathbf{x} \cdot \mathbf{y} \tag{2.2}$$

We find

$$Tr(\rho - \rho')^2 = \frac{(\mathbf{r} - \mathbf{r}')^2}{2} \tag{2.3}$$

This means that the Euclidean distance in the Bloch ball is proportional to the Hilbert space distance between states.

Rigid transformations of Euclidean space are translations and rotation (and inversions). Since unitaries take the Bloch ball to itself, translations are not allowed. Inversions are not allowed because they dot preserve the commutation relations of X, Y, Z. We are left with rigid rotation of the ball.

To find the axis of rotation U we need to distinguish two cases:

- $U = e^{i\alpha}\mathbb{1}$ is degenerate: The rotation is the identity.
- U is non-degenerate and has two distinct eigenvalues.
  - The two eigenvectors define two unique directions in Hilbert space

    $$U|\psi\rangle = e^{i\alpha}|\psi\rangle, \quad U|\psi_\perp\rangle = e^{i\beta}|\psi_\perp\rangle$$

    which are invariant under the action of U.
  - $|\psi\rangle$ and $|\psi_\perp\rangle$ are associated with antipodal points on the Bloch sphere.
  - The line connecting the antipodes is the axis of rotation.
  - The angle of rotation is $\alpha - \beta$.

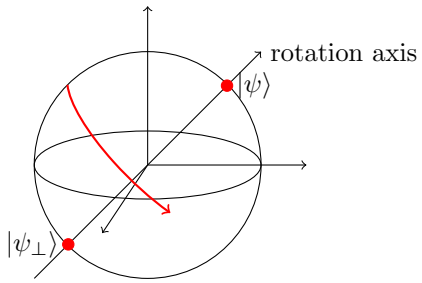


Figure 2.1: H gate is a rotation by π of the Bloch sphere around the -z-x axis

It follows that if $U \neq \pm\mathbb{1}$ while $U^2 = \mathbb{1}$ then U rotates the Bloch ball by π.

2.2 NOT gate

X rotates the Bloch ball by π around the X axis. and is represented by the unitary X.

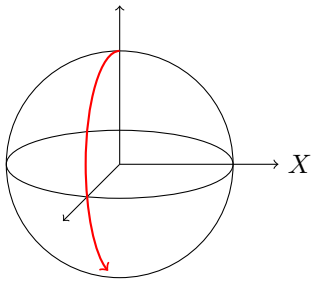
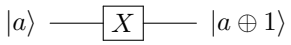


Figure 2.2: NOT gate is a rotation by π of the Bloch sphere around the x axis



2.3 Z gate

The Z gate is a π rotation of the Bloch sphere about the Z axis:

$$|\pm\rangle \text{ —— } \boxed{Z} \text{ —— } |\mp\rangle$$

This gate has no classical analog since, in the computational basis, it is just a phase gate

$$|a\rangle \text{ —— } \boxed{Z} \text{ —— } (-)^a|a\rangle$$

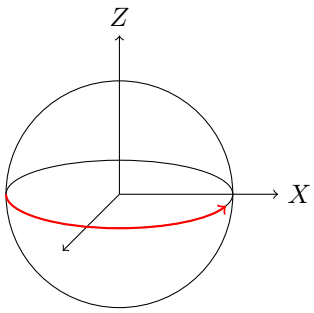


Figure 2.3: Z gate is a rotation by π of the Bloch sphere around the z axis

Exercise 2.1. *Determine τ and $\mathbf{B}$ so that the pulsed Hamiltonian $H(t) = \tau\delta(t)\mathbf{B}\cdot\boldsymbol{\sigma}$ implement the X and Z gates.*

2.4 Hadamard gate

The Hadamard gate is defined by

$$H = \frac{X+Z}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \tag{2.4}$$

Since X and Z anti-commute

$$H^2 = \mathbb{1} \tag{2.5}$$

so H too is a rotation by π of the Bloch ball. It is a rotation about the $X+Z$ axis and it interchanges the X and Z axis. If you do not trust geometry this follows from

$$HZ = XH$$

As a consequence, U may be viewed as the unitary map between the computational basis and the $|\pm\rangle$ basis:

$$|a\rangle \text{ —— } \boxed{H} \text{ —— } |(-)^a\rangle$$

A good notation can lead to deep results. This is the case for

$$H|a\rangle = \frac{1}{\sqrt{2}} \sum_{b\in 0,1} (-)^{ab}|b\rangle \quad a, b \in 0, 1 \tag{2.6}$$

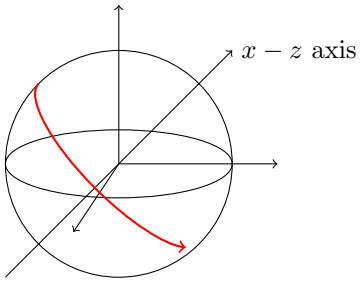


Figure 2.4: H gate is a rotation by π of the Bloch sphere around the -z-x axis

2.5 Pauli generate rotations

The Pauli matrices have the commutation relation

$$\left[\frac{X}{2}, \frac{Y}{2}\right] = i\frac{Z}{2}, \quad \text{and cyclic}$$

These are the commutation relation of angular momentum

$$[L_x, L_y] = iL_z, \quad \text{and cyclic}$$

Since the angular momenta are the generators of rotations

$$U(\mathbf{n}) = e^{-i\mathbf{n}\cdot\boldsymbol{\sigma}/2}, \tag{2.7}$$

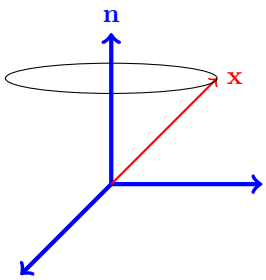
is a rotation. The axis of rotation is $\mathbf{n}$ since

$$U\mathbf{n}\cdot\boldsymbol{\sigma}U^\dagger = \mathbf{n}\cdot\boldsymbol{\sigma}$$

Rotation of a vector $\mathbf{x}$ by angle φ about the $\hat{\mathbf{n}}$ axis does not affect the part parallel to the axis. The part perpendicular to it, rotates:

$$\mathbf{x}' = R_n(\varphi)\mathbf{x} = \underbrace{(\mathbf{x}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}}_{parallel} - \underbrace{(\mathbf{x}\times\hat{\mathbf{n}})\sin\varphi + (\mathbf{x}\times\hat{\mathbf{n}})\times\hat{\mathbf{n}}\cos\varphi}_{perpendicular} \tag{2.8}$$

One can show from this, with some more work, that the angle of rotation is $|\mathbf{n}|$.



Using the fact

$$(\mathbf{n}\cdot\boldsymbol{\sigma})^2 = \mathbf{n}^2\mathbb{1}$$

and power expanding one finds a nice and useful formula for the rotation matrix:

$$e^{-i\mathbf{n}\cdot\boldsymbol{\sigma}/2} = \mathbb{1}\cos\left(\frac{|\mathbf{n}|}{2}\right) - i\,\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\,\sin\left(\frac{|\mathbf{n}|}{2}\right) \tag{2.9}$$

2.6 Geometry of $SU(2)$

$SU(2)$ are the unitary 2×2 matrices with $\det U = \mathbb{1}$.

Theorem 2.2. *A unitary 2×2 matrix U with $\det U = 1$ can be identified uniquely with a point in S^3:*

$$U = e^{-i\mathbf{u}\cdot\boldsymbol{\sigma}/2} = \mathbb{1}\cos\frac{|\mathbf{u}|}{2} - i\mathbf{u}\cdot\boldsymbol{\sigma}\sin\frac{|\mathbf{u}|}{2}, \quad |\mathbf{u}| \le 4\pi$$

parametrized by $|\mathbf{u}| \le 4\pi$. Note that $\mathbf{u} = 0$ and $\mathbf{u} = 4\pi$ represent a single point $U = \mathbb{1}$ and $|\mathbf{u}| = 2\pi$ also represent a single point, $U = -1$. The geometry becomes apparent in a 4-D Euclidean representation

$$\left(\cos\tfrac{1}{2}|\mathbf{u}|, \mathbf{u}\sin\tfrac{1}{2}|\mathbf{u}|\right)$$

where each point has unit length. The space is therefore S^3. U is a rigid rotation of the Bloch ball by angle $|\mathbf{u}|$ around the axis $\mathbf{u}$. In particular the angle of rotation is

$$2\cos\left(\frac{|\mathbf{u}|}{2}\right) = Tr\,U \tag{2.10}$$

2.7 Gauge freedom

The overall phase of a state is not a physical entity. We could have defined the computational basis by

$$|a'\rangle = \sum_a U_{a,a'}|a\rangle, \quad U = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix} \tag{2.11}$$

and simultaneously redefine

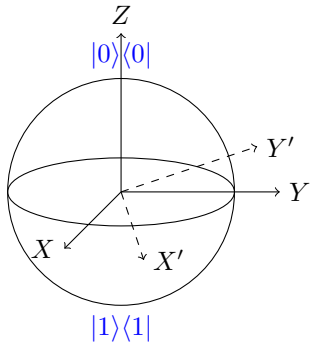
$$Z' = UZU^* = Z, \quad X' = UXU^*, \quad etc. \tag{2.12}$$

In the new basis, the state ρ takes the form

$$\rho' = U\rho U^* = \begin{pmatrix} \rho_{00} & e^{i(\phi_0-\phi_1)}\rho_{01} \\ e^{i(\phi_1-\phi_0)}\rho_{10} & \rho_{11} \end{pmatrix} \tag{2.13}$$

The diagonal is gauge invariant but the off diagonal terms are not. Only their magnitude is gauge invariant.

Since U is a rotation about the Z axis, you may relate gauge freedom to the freedom to choose the point along the equator of the Bloch sphere where the matrix X of Eq. (1.6) represents the X axis.



2.8 Universal single qubit gates

Logical operations in classical computers can be done with few standard gates. A discovery made in the MSc thesis of Shannon at MIT. If the computation is complex we may need many gates, and use some gates many times, but the number of gate *types* does not increase with the complexity of the computation. This issue is related to the fact that Turing machines are, by definition, *finite*.

What about quantum gates? Can we cover the Bloch sphere starting with the computational basis and operating with a finite number of gate types?

The X and Z gate generate the Pauli group which is a finite subgroup of the group of the unitary group $U(2)$. Starting with the state $|0\rangle\langle 0|$ all the Pauli group can do is generate the computational basis.

The the X, Z and H generate a larger, but still finite group. Starting with the state $|0\rangle\langle 0|$ all you can get are the 4 states:

$$|0\rangle\langle 0|, \quad |1\rangle\langle 1|, \quad |+\rangle\langle +|, \quad |-\rangle\langle -|$$

Exercise 2.3. *How many different words do X, Z and H with the relations*

$$Z^2 = X^2 = H^2 = \mathbb{1}, \quad ZH = HX, \quad ZX + XZ = 0$$

generate?

2.9 Irrational rotations

The gate

$$e^{i\pi \frac{p}{q} Z}, \quad p, q \in \mathbb{N}, \quad gcd(p, q) = 1$$

rotates the Bloch sphere by the angle $\pi \frac{p}{q}$ about the Z axis. Using such a rotation we can map $|+\rangle$ to q states that are evenly distributed along the equator.

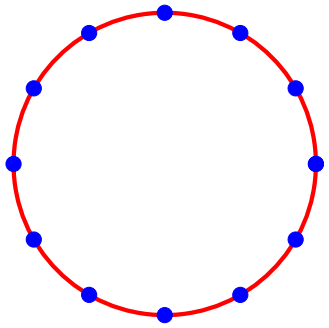


Figure 2.5: A rational rotation by 30 degrees.

This makes it intuitively clear that to generate an arbitrarily good approximation to $e^{i\alpha Z}$ for any α, we need just a single gate: One that generates an irrational rotation about the Z axis, e.g.

$$e^{i\pi Z/\sqrt{2}}$$

Exercise 2.4. *Prove this? Hint: Use the fact from number theory that any irrational α can be approximated by rationals so that*

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^2}$$

with arbitrarily large p and q.

2.10 T Phase gate

The phase gate T is defined by:

$$-\boxed{T}- \quad = \quad \begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix}, \quad \omega = e^{i\pi/4}, \quad T|a\rangle = \omega^a |a\rangle$$

Since

$$T^4 = Z$$

and Z rotates the Bloch ball by π, it follows that T rotates the Bloch ball by $\pi/4$ about the z axis. Another way to see this is to write

$$T = e^{i\pi/8} T_8, \quad T_n = \begin{pmatrix} \bar{z}_n & 0 \\ 0 & z_n \end{pmatrix}, \quad z_n = e^{i\pi/n}$$

Evidently, T and T_8 give the same rotation of the Bloch ball. Since $\det T_8 = 1$ we can use Eq. 2.10 to conclude that the rotation angle is $\pi/4$.

Although T and H are both rational rotations (about different axis) their product is an irrational rotation. As we shall now see T is, in a certain sense, the simplest rotation around Z that has this property.

Write

$$iHT_n = \frac{i}{\sqrt{2}} \begin{pmatrix} \bar{z}_n & z_n \\ \bar{z}_n & -z_n \end{pmatrix},$$

Since $\det(iHT_n) = 1$ we can use Eq. 2.10 to find the angle of rotation θ_n

$$\cos \frac{\theta_n}{2} = \frac{\sin(\pi/n)}{\sqrt{2}}$$

With $n = 1, 2, 4$ you get rational rotations by $\pi, 3\pi/2, 4\pi/3$ which are all rational rotations. However with $n = 8$ the rotation angle θ_8 is irrational. (I do not prove that.) Since HT is proportional to iHT_8, we learn that HT gives an irrational rotation.

Exercise 2.5. *Find the corresponding axis of rotation.*

Similarly, THwill give us the same irrational rotation but about a a different axis.

H and T generate an infinite group with infinitely many different words, e.g

$$HT^{n_1} HT^{n_2} H \ldots, \quad n_j \in \{1, \ldots, 7\}$$

Remark 2.6. *You may also want to worry about the question: Suppose I want to approximate arbitrary rotation with n digits of accuracy given universal gates. How does the number of actual gates (not gate types) scale with n?*

2.11 Complete control

We say that we have complete control of a qubit if we can rotate the Bloch vector by arbitrary rotation. Rotations of a qubit, like ordinary rotation, parametrized by three angles for example the three Euler angles:

You might think that to generate arbitrary rotation you'd need to be able to rotate by arbitrary angle about three different directions. After all, a pilot has three controls for pitch, yaw and roll.

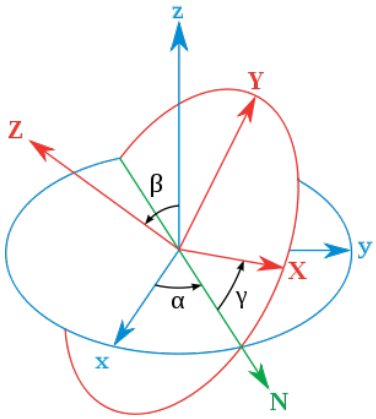


Figure 2.6: Euler angles rotate the frame (x, y, z) to (X, Y, Z). N is the line of nodes. (Figure taken from Wikipedia.)

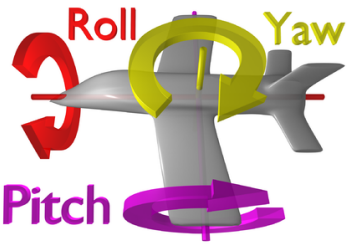
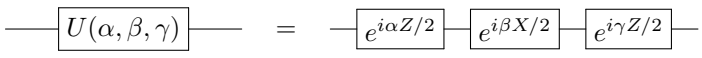


Figure 2.7: Pitch, yaw and roll. (Figure taken from Wikipedia.)

However, it turns out that actually two controls suffice. This is because rotations are non-commutative:

$$[\mathbf{L}\cdot\mathbf{a}, \mathbf{L}\cdot\mathbf{b}] = i\mathbf{L}\cdot\mathbf{a}\times\mathbf{b}$$

In fact, Euler only uses rotations about the x and z axis.



As we can generate arbitrary rotations about the HT and TH axes, we can generate any rotation.

Theorem 2.7 (Mor et. al)**.** *An arbitrary single qubit unitary can be approximated arbitrarily well with two qubit types: H and T*

Chapter 3

Appendix: Linear algebra

Any orthonormal basis $|j\rangle$, with $j = 1, \ldots N$ in $\mathbb{C}^N$ gives a resolution of the identity

$$\sum_j |j\rangle\langle j| = \mathbb{1}$$

Linear operators are represented by matrices. In particular, if $|j\rangle$, $j \in 1, \ldots, N$ is the computational basis in $\mathbb{C}^N$ then

$$A|j\rangle = \sum_{k=1}^{N} A_{kj}|k\rangle$$

The (funny) order on the rhs is dictated by $\langle k|A|j\rangle = A_{kj}$.

Definition 3.1. *An operator A is real (=Hermitian)/positive if every expectation value $\langle\psi|A|\psi\rangle$ is real/positive for any vector $|\psi\rangle$.*

Exercise 3.2. *Show that if A is real then $A = A^\dagger$*

Exercise 3.3. *Show that $AA^\dagger$ and $A^\dagger A$ are positive and have identical eigenvalues except possibly for 0.*

Exercise 3.4. *Show that if $\lambda \neq 0$ is an eigenvalue of $AA^\dagger$ it is also an eigenvalue of $A^\dagger A$ and vice versa.*

Definition 3.5. *Let $|j\rangle$ be any orthonormal basis. The trace of an operator A is*

$$Tr\, A = \sum_j \langle j|A|j\rangle$$

The trace is independent of the choice of the basis.

Exercise 3.6. *Show that*

$$Tr\ |\phi\rangle\langle\psi| = \langle\psi|\phi\rangle$$

3.1 Functions of operators

What do we mean by a function of an (Hermitian) operator in Hilbert space $\mathcal{H}$? Since H is self-adjoint it can be diagonalized. Suppose H has discrete spectrum with eigenvalues h_j and P_j the associated projections. Then

$$H = \sum h_j P_j$$

We then *define*

$$f(H) = \sum f(h_j) P_j$$

Exercise 3.7. *Show that if* $Tr\, H = 0$ *then* $\det e^{iH} = 1$

In the case of $\mathbb{C}^2$ we can say more. Any traceless Hermitian 2×2 matrix can be written as

$$H(\mathbf{b}) = \mathbf{b} \cdot \sigma, \quad \mathbf{b} \in \mathbb{R}^3$$

By Ex. **??**,

$$H^2 = |\mathbf{b}|^2 \mathbb{1}$$

and then by Ex 1.1 the spectral projections are

$$P_\pm = \frac{\mathbb{1} \pm \mathbf{b} \cdot \sigma}{2}$$

and so

$$f(H) = \mathbb{1} f_+(|\mathbf{b}|) + i\,(\mathbf{b} \cdot \sigma)\, f_-(|\mathbf{b}|), \quad 2f_\pm(x) = f(x) \pm f(-x)$$

In particular, for any traceless Hermitian 2×2 matrix

$$e^{-iH(\mathbf{b})} = \mathbb{1} \cos(|\mathbf{b}|) - i\, \mathbf{b} \cdot \sigma \sin(|\mathbf{b}|)$$

Exercise 3.8. *What are the corresponding formulas when* $Tr\, H \neq 0$.

3.2 Unitaries

A unitary matrix U may be interpreted as the transformation from one base in the Hilbert space to another base:

$$U = \sum_{j=1}^{N} |b_j\rangle\langle a_j|$$

where $|a_j\rangle$ and $|b_j\rangle$ are two bases in $\mathbb{C}^N$. A basis independent way of expressing this is:

Definition 3.9. *U is unitary matrix in $\mathbb{C}^N$ if $U^{-1} = U^\dagger$ or, equivalently*

$$U^\dagger U = U U^\dagger = \mathbb{1}$$

Exercise 3.10. *Show that the spectrum of a unitary lies on the unit circle and that eigenvectors associated with distinct eigenvalues are orthogonal.*

Every unitary matrix is diagonalizable.

Remark 3.11. *For finite dimensional matrices the inverse of a matrix is both a right inverse and a left inverse. For infinite dimensional matrices this need not be true. U is called an isometry if $U^\dagger U = \mathbb{1}$.*

The right shift on $\mathbb{N}$, i.e.

$$(R\psi)(n) = \begin{cases} \psi(n-1) & n \geq 1 \\ 0 & n = 0 \end{cases}$$

is an isometry.

Exercise 3.12. *Show that if U is an isometry then $UU^\dagger$ is an orthogonal projection.*

Chapter 4

Gates: Realizations

4.1 Mirror

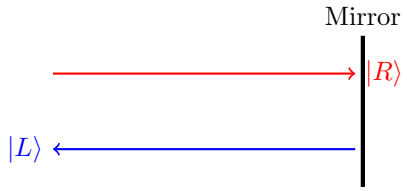


Figure 4.1: By conservation of angular momentum a mirror interchanges circular R and L polarization.

By conservation of angular momentum, reflection from a mirror interchanges left and right circular polarizations. This fixes the gate (up to an overall phase) to be

$$M = \begin{pmatrix} 0 & \alpha \\ \bar{\alpha} & 0 \end{pmatrix}, \quad |\alpha| = 1$$

Consider now an incoming $|H\rangle$. This comes out as $(\bar{\alpha}, \alpha)$. We are free to orient the incoming and outgoing frames so that what we call horizontal agrees in both. This says that

$$M = X$$

4.2 Hadamard – Beam splitter

The photon can be in one of two states: moving to the right or moving up. An even beam splitter sends

$$|0\rangle_{in} \mapsto \frac{e^{i\alpha}|0\rangle_{out} + e^{i\beta}|1\rangle_{out}}{\sqrt{2}}$$

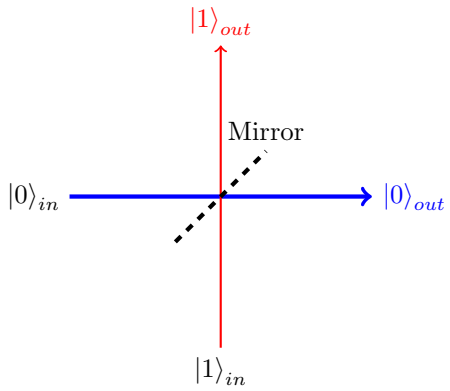
Since a unitary gate sends orthogonal states to orthogonal states, it follows that

$$|1\rangle_{in} \mapsto \frac{e^{i\alpha}|0\rangle_{out} - e^{i\beta}|1\rangle_{out}}{\sqrt{2}}$$

In fact we are free to redefine the basis so that

$$e^{i\alpha}|0\rangle_{out} \mapsto |0\rangle_{out}, \quad e^{i\beta}|1\rangle_{out} \mapsto |1\rangle_{out}$$

So the beam splitter is Hadamard



4.2.1 Classical perspective

Example 4.1. *A plane electromagnetic wave of unit amplitude impinges on a beam splitter from the left*

$$\mathbf{E}_{in} = ye^{i(x-t)}, \quad \mathbf{B}_{in} = ze^{i(x-t)}$$

. The beam is split to two outgoing waves, one right

$$\mathbf{E}_{rt} = (by - az)e^{i(x-t)}, \quad \mathbf{B}_{rt} = (ay + bz)e^{i(x-t)}$$

and one up

$$\mathbf{E}_{up} = (dx - cy)e^{i(z-t)}, \quad \mathbf{B}_{up} = (cx + dy)e^{i(z-t)}$$

The boundary conditions across the beam splitter are that $\mathbf{B}_{\perp}$ *and* $\mathbf{E}_{\parallel}$ *are continuous.*

- *Explain.*
- *Show that the boundary conditions imply*

$$c = 1 - b, \quad d = -a$$

- *Determine* b *for a* 50% *beam splitter*
- *Did you get Hadamard?*

4.3 Half and Quarter wave plates

On a length scale which is much larger than atomic spacing, a crystal looks homogeneous but un-isotropic. On such scales the dielectric constant can be represented a symmetric matrix. In the frame of its principal axis the matrix looks like

$$\varepsilon = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}$$

It follows that plane wave propagating in the z direction will have different speed for x and y polarization if $\varepsilon_{xx} \neq \varepsilon_{yy}$. Since the frequency ω is determined by the source (laser) the different propagation speeds imply different wave lengths for the two polarizations. A plate of of width Δz will cause a phase shift between the two polarizations

$$\alpha = (k_x - k_y)\Delta z = \omega \Delta z \left(\frac{1}{c_x} - \frac{1}{c_y} \right)$$

We choose the x and y polarizations to correspond to the $|\pm\rangle$ basis. In this basis the plate is described by the unitary

$$e^{i\alpha Z/2} = \mathbb{1} \cos \alpha/2 + iZ \sin \alpha/2$$

Since H transforms between the X and Z basis, the plate is represented in the Z basis by the gate

$$\begin{aligned} He^{i\alpha Z/2}H &= \mathbb{1} \cos \alpha/2 + iHZH \sin \alpha/2 \\ &= \mathbb{1} \cos \alpha/2 + iXHH \sin \alpha/2 \\ &= e^{i\alpha X/2} \end{aligned}$$

With $\alpha = \pi/2$ this rotates the Bloch sphere by $\pi/2$ about the X axis. It converts circular polarization to linear polarization. This is a quarter wavelength plate. With $\alpha = \pi$ it rotates the Bloch sphere by π and interchanges $|R\rangle$ to $|L\rangle$.

Exercise 4.2. *In what way $e^{i\pi X/4}$ is similar to and different from H?*

Exercise 4.3. *In 3D movies, the pictures projected into your two are are filtered according to two orthogonal polarizations. Since people, especially in India, tilt their heads when watching movies, the filtering is by circular polarizations. In practice, filters are always linear polarizers. How would you design a filter for circular polarization?*

4.4 Spin gates

If the qubit is a spin 1/2 with Magnetic moment μ then applying a magnetic field $\mathbf{B}$ is associated with the Hamiltonian

$$H = \frac{\mu}{2} \mathbf{B} \cdot \boldsymbol{\sigma}$$

Schrödinger equation says that this generates the evolution

$$i\dot{U} = H(t)U, \quad , U_0 = \mathbb{1}$$

If H is time independent

$$U(t) = e^{-iHt}$$

This rotates the Bloch sphere at constant rate about the direction $\mathbf{B}$.

Exercise 4.4. *Calculate the rate of rotation of the Bloch sphere, in the magnetic field of the earth, 1 Gauss, and for* 1 T*, for nuclear spin 1/2 and electronic spin 1/2, where* $H = \mu \mathbf{s} \cdot \mathbf{B}$ *and* μ *the magnetic moment.*

Consider now a pulsed Hamiltonian with

$$\mathbf{B}(t) = \mathbf{\Phi}\delta(t)$$

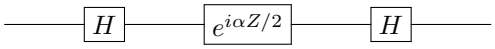
The corresponding unitary is

$$U = e^{-i\mathbf{\Phi}\cdot\boldsymbol{\sigma}/2}$$

4.5 Mach-Zehnder interferometer

Mach-Zehnder is made with two beam splitters, each represented by H and two mirrors that, we assume, act like the identity[1].

The Mach-Zehnder interferometer can be represented by circuit



There is some freedom in decorating the diagram with $|0\rangle$ and $|1\rangle$ along the path. and I have chosen $|0\rangle$ to represent the blue (lower) path and the $|1\rangle$ the upper path. α is the difference in optical lengths between the two paths.

Exercise 4.5. *Suppose* $\alpha = 0$*. If the incoming photon is in the* $|0\rangle$ *state, what is the state of the outgoing photon?*

[1] Alternatively, if the mirrors are represented a s an X gate then one can stick with horizontal propagation being identified with $|0\rangle$ and vertical with $|1\rangle$.

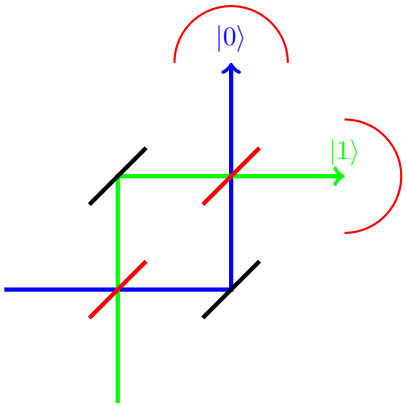


Figure 4.2: Mach-Zehnder interferometer. The two red arcs are detectors. The red slanted lines are beam splitters. The black slanted lines are mirrors. The multiple paths of the photon are shown in blue and green. The green lines represent the channel $|1\rangle$ and the blue $|0\rangle$. The two arms in the interferometer may have different optical paths expressed in the relative phase $e^{i\alpha}$.

Chapter 5

Quantum tricks I

5.1 Random number generator

Random numbers are a basic primitive in many applications.

- Lottery
- Algorithms in computer science.
- Integration of high dimensional functions (Monte Carlo)
- Cryptography
- Computer simulations.
- Economics

Unfortunately, as a matter of principle, there are no true random numbers in classical physics, simply because classical physics is deterministic. What appear to be random numbers is an expression of *incompete knowledge*: If your adversary knows more than you your random numbers may not be as random to him. Security in cryptography requires random sequence that are unpredictable. Since you do not know who your opponent is, or what he knows, it is best to assume that he knows all that can be known.

Example 5.1 (Coin toss)**.** *In some science museums you can find deterministic coin tossing machines: You can tune the machine so that in every coin too the coin will land on the face it showed initially. Lets go though this example:*

- *You see the initial face of the coin*
- *The coin, with radius a, is kicked at the rim*
- *You can estimating the hight of a coin toss h to better than $\pi a/8$*
- *When the coin hits the table it comes to rest without bouncing*

- *Are allowed to choose head or tail after you saw how high the coin went*
- *You can then tell if it is going to be head or tail.*

The center of mass orbit is a parabola and the angular rotation is constant

$$x = -\frac{1}{2}gt^2 + vt, \quad \theta = \omega t$$

You do not know the initial conditions v, ω. But, if the motion of the coin is due to a kick at the rim of a coin of radius a you know:

$$\delta p = mv, \quad \delta N = a\delta p = I\omega = \frac{ma^2\omega}{2} \Longrightarrow 2v = \omega a$$

The hight of the coin toss

$$2gh = v^2$$

determines the initial velocity v. The number of turns of the coin makes till its center of mass is back at $x = 0$ is

$$\theta = 2\omega\frac{v}{g} = \frac{4v^2}{ga} = \frac{8h}{a}$$

The number of turns is

$$\frac{\theta}{2\pi} = \frac{4h}{\pi a}$$

If this number is close to an integer or close to half integer you can tell head from tail.

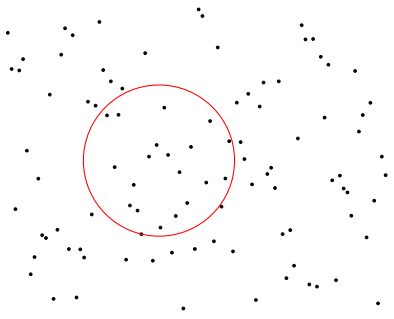


Figure 5.1: You can compute the (relative) area of a complicated shape by counting the number of points that fall insider the shape, when the points are uniformly distributed in the unit square.

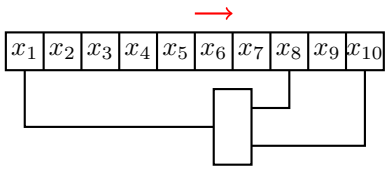
Randomness in QM shows up not because of incomplete knowledge, but rather as a fundamental feature of reality. Hence, in principle, only QM can provide truly random sequences.

Some of classical algorithms that have, or are being used, to generate random numbers are amusing.

- During the second world war the Russian had units of soldiers assigned to throwing dice. The human random number generators could not produce random numbers at the rate the army needed and the army had to recycle the list. This is something you do not want to do in cryptography.
- In 1955 RAND corporation published a list called: A million random digits. This is of course not a good list for cryptography, but it can be useful for algorithmic purposes such as Monte Carlo integration.
- Casinos use roulette table to generate, what gamblers believe, are random numbers. Roulettes are deterministic mechanical devices. Casino allow you to look at the roulette and place your bet after it starts spinning, provided, of course, the wheel is spinning. If you are clever and quick enough, you can guess the initial position and velocity by observing the wheel. C. Shannon used an early version wearable computer, to recompute probabilities after the wheel start spinning so that his chances of winning were better than the house. He was good enough for the house to throw him out.
- The industrial way of making random numbers is by deterministic algorithms known as pseudo-number generators for example, the Linear Shift Feedback Register. The basic scheme is
  - You need a register with n bits with n is large.
  - The register is updates deterministically

$$x \mapsto f(x),\ x = \underbrace{x_1 x_2 \dots x_n}_{\text{binary rep}} \in \{0, \dots N-1\}$$

    with f a standard and known function
  - $f(0) = 0$
  - $(f \circ f \dots)(x)$ goes through all $x \neq 0$ configurations for $x \neq 0$.
  - The last bit x_n is the output
  - You can not tell the current state of the register x by observing the last n outputs.



The trouble with all these methods, is that either they are obviously predictable, like pseudo-number generators, or in principle predictable, as any classical mechanical device must be.

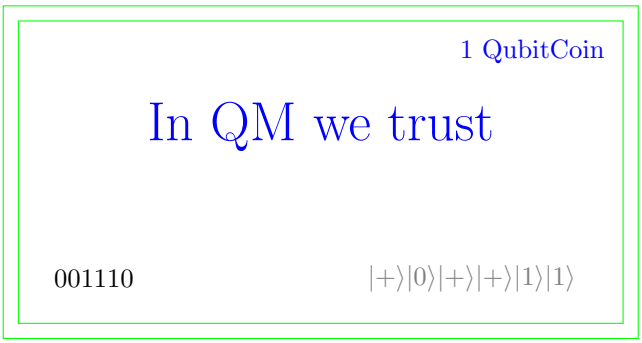
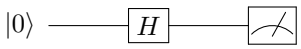


Figure 5.2: Quantum banknote

Here is a quantum circuit that produces bona fide random numbers



The Hadamard gate generates deterministically $|+\rangle$. But the measurement in the computational basis, gives 0 or 1 with probability 1/2:

$$Prob(0) = \frac{1 + \hat{\mathbf{x}} \cdot \hat{\mathbf{z}}}{2} = \frac{1}{2}, \quad Prob(1) = \frac{1 - \hat{\mathbf{x}} \cdot \hat{\mathbf{z}}}{2} = \frac{1}{2}$$

The procedure is deterministic but the outcome is random. Unlike classical physics, QM randomness is not the result of incomplete knowledge. No super being can predict the result. (We shall say more about why we believe hat this is the case when we discuss hidden variables.)

The Achilles heel of QM is the measurement devices, which, by definition, are classical. So, if an adversary gets access to the measurement apparatus, he may be able to tinker with it and corrupt the random numbers.

5.2 Quantum money

Quantum information was born in about 1970 when S. Wiesner, a grad student, proposed the idea of quantum money. His paper was repeatedly rejected and only appeared in 1983. Wiesner proposed money that can not be counterfeited in principle. In fact, any attempt to make a copy of the paper would result is erasing the original legitimate banknote.

Protocol:

- The bank issues bank notes that carry an n-digit classical binary serial number $x \in \mathbb{Z}_2^n$ that all can see, and a secret quantum serial number, $q(x) \in \mathbb{Z}_2^n$ made from n-qubits.

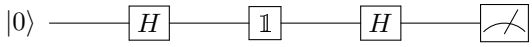
- The bank uses a true random number generator to assign to $x \in \mathbb{Z}_2^n$ the secret $q(x) \in \mathbb{Z}^n$.
- The bank uses a true random number generator to assign to $x \in \mathbb{Z}_2^n$ a secret random number whose entries are $b(x) \in \{Z, X\}^n$.
- If the j-th binary digit, $b_j(x) = Z$ the j-qubit is written in the Z basis. That is, $q_j \mapsto |q_j\rangle$, If $b_j(x) = X$ the j-qubit is written in the X basis. That is $q_j \mapsto H|q_j\rangle$.
- The bank can reliably verify the quantum serial number: $\{|0\rangle, |+\rangle, |1\rangle, |-\rangle\}$ and the verification does not alter the qubits.
- Someone who does not have access to the secret table will need to guess the basis. He will make, with high probability, a $n/2$ wrong guesses about the choice of basis. Reading in the basis he guessed he will make about $n/2$ errors in reading the secret $q(x)$ and the note he will forge will be easily identified as forgery and in addition he had ruined his own legitimate bank note.

This, of course, raises the question if it is possible to copy unknown quantum information without reading it. We shall see that this not possible.

5.3 Vaidman Elitzur Bomb

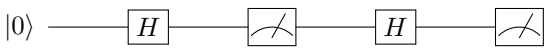
The story, like a typical news item from the middle east, starts with bombs: You have a collection of them. They have a quantum trigger which makes them explode when a photon is reflected from the trigger. Some of the bombs are duds: Their trigger is stuck and the photon is reflected and the mirror does not measure anything. Can you find bombs that are guaranteed to be alive? This brings up the issue of quantum non-demolition: Can you make a quantum measurement that will allow you to learn something you did not already know about a quantum system, without modifying it?

Place the bomb so that the trigger is a mirror in Mach Zehnder. Suppose the bomb is dead. A circuit representing a dead bomb is made with unitary gates:



and $\mathbb{1}$ represents the bomb. It does nothing. The incoming state is the same as the outgoing. Hence, only the detector C (on the right) clicks. Put this bomb in a stockpile marked "Dead and Alive".

Now suppose the bomb is alive. It acts like a measurement device: It is a which path detector. The circuit that represents this is



The measurement, represented by a meter, prepares one of two states: $|1\rangle$ represents the path that went through NW mirror and $|0\rangle$ the path via SE mirror which is the trigger of the bomb.

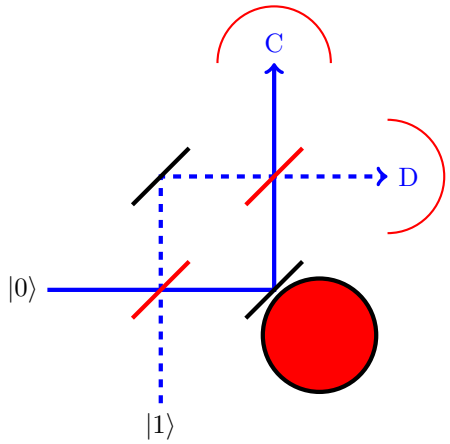
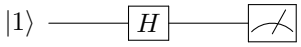


Figure 5.3: The Elitzur Vaidman Bomb testing apparatus is built around a Mach-Zehnder interferometer. The two paths represent the computational basis $|a\rangle$. If the bomb is dead, and the photon comes in at $|0\rangle$, it comes out as $|0\rangle$ and only the detector C clicks. If the bomb is alive, then then it determines the path the photon took. If the photon took the upper arm, the bomb did not explode and there is 50% chance that detector D will click. This allows to identify the bombs which are alive.

Suppose the measurement determines that the particle went through the SE mirror. The trigger was activated, the bomb explodes, and you need a new lab.

However, the curious thing about QM is that the which path measurement done by the bomb placed at SE mirror will in half the cases determine that the particle went the other way, through the NW mirror. In this case, the bomb does not explode. Instead, it prepared the system in state $|1\rangle$ and the corresponding right half of the circuit (following the meter) is replaced by



In this case, there 50% chance that the C-detector will click and 50% that the D-detector will. Since the D detector never clicks for dead bombs, you know that the bomb is alive, and you put it in the pile "Alive Only".

5.4 Quantum key distribution

5.4.1 Encryption and decryption

A key is a sequence of bits

$$k = \{k_1, \ldots k_n\}, \quad k_n \in \{0, 1\}$$

which Alice and Bob share. The message $m = \{m_1, \ldots, m_n\}$ is another sequence of bit of the same length. Alice encrypts the message $e = E(m, k)$ and broadcast it which Bob then has to decrypt to get m. A simple encryption-decryption scheme is:

- Encryption: bitwise addition (mod 2)

$$m \mapsto E(m, k) = \{m_1 \oplus k_1, \ldots, m_n \oplus k_n\}$$

- Decryption:

$$D \mapsto D(E(m, k), k) = \{m_1 \oplus k_1 \oplus k_1, \ldots, m_n \oplus k_n \oplus k_n\} = m$$

The space of all messages has 2^n messages. Similarly, the space of all keys has 2^n elements and so does the space of all encryptions.

5.4.2 Security

What do we mean by saying that an encryption scheme is secure? A definition, going back to Shannon, is:

Definition 5.2. *The encryption $e = E(m, k)$ of a message m is secure if intercepting the encryption e gives no information on m:*

$$P(m|e) = P(m)$$

where $P(m)$ is the probability of the message m and $P(m|e)$ is the conditional probability for the message m given the encryption e.

In a secure encryption, the message and its encryption are independent

$$\underbrace{P(m, e) = P(m|e)P(e)}_{\text{Bayes}} = P(m)P(e)$$

where $P(m, e)$ is the joint probability for the message m and the encryption e. This implies that the key must be chosen independently of the message.

5.4.3 One time key pad

In a one time key pad the key is chosen independently of the message with uniform distribution:

$$P(m, e, k) = \delta(e = m \oplus k)P(m, k) = \delta(e = m \oplus k)P(m)2^{-n}$$

It follows that

$$P(m, e) = \sum_k \delta(e = m \oplus k)P(m)2^{-n} = P(m)2^{-n}$$

and

$$P(e) = \sum_m P(m)2^{-n} = 2^{-n}$$

is uniformly distributed and

$$P(m, e) = P(m)P(e)$$

We can write this as

$$P(m, e) = P(m)P(e) = P(m|e)P(e) \Longrightarrow P(m|e) = P(m)$$

We have therefore proved:

Theorem 5.3 (Shannon)**.** *One time pad encryption is secure.*

Note that $P(m)$ may have any distribution. For example, Bob is allowed to ask Alice to encrypt for him the single single message

War starts tomorrow at 08:00

The point is that one Alice encrypts the message a second time, Bob will not be able to decipher it.

Exercise 5.4. *Suppose a message m has correlations between consecutive bits expressed by unequal probabilities of the various two-bit words*

$$P(0, 0) > P(0, 1) > P(1, 0) > P(11)$$

Show that encryption with one time pad of identically distributed bit erases the correlations.

The one time pad is secure if used once. If you use the same key for two (binary) messages m and m' then you know everything about $m \oplus m'$. You have half the information about the two messages.

The weakness of the one-time pad is the absence of a reliable method for key sharing: If you have a secure way to transmit the key you can use it to transmit m.

Figure 5.4: From Left: A 360×360 image; A random array of black (rgb=0,0,0) and white (rgb=1,1,1) pixels; The encrypted image; The decrypted image.

Example 5.5. *A Mathematica program that generates the figure*

```
qi=ImageData[Graphics[Text[Style[QI,80]]]];
ran=Table[RandomInteger[{0,1}]{1.,1.,1.},{j,1, 360},{k,1,360}];
encript=Mod[ran+qi,2]; decript=Mod[encript+ran,2];
GraphicsRow[{Image[qi],Image[ran],Image[encript],Image[decript]},
Frame→ All]
```

5.4.4 Quantum key distribution (BB84)

Quantum mechanics provides a secure way to share a (classical) key.

- Alice chooses a (random classical binary) sequence $b_A = \{Z\ldots, X\ldots\}$ of length $4n$.
- Alice chooses random key k of (classical) bits 0,1 of length $4n$
- Alice encoded k in qubits in the (random) base b taken from X, Z and send the qubits to Bob.
- Bob measures his qubits in a randomly chosen basis $b_B = \{Z\ldots, X\ldots\}$
- Bob and Alice broadcast the bases b_B and b_A
- Alice and Bob trash the bits where their basis elements disagree (about 1/2)
- Alice or Bob broadcast (a random) half of the un-trashed bits to check for eavesdropping.
- If there is perfect agreement, they use the remaining (approximately) n bits as private key, and they are confident no one was eavesdropping
- If Eve is eavesdropping, she would most likely use the wrong basis for half of the qubits she intercepted. As a consequence a quarter of the (intercepted) bits will disagree.
- If Alice and Bob suspect eavesdropping, they try again.

Remark 5.6. *This is not a proof of security. I did not specify how much power the eavesdropper (Eve) has. I have assumed that Alice prepares her qubits in pure states and Bob receives pure states. What is Eve is allowed to mess up Alice preparation? What if Alice qubits are actually entangled with Eve's qubits? I will not address this here since we still have not learned about entanglement. The actual proof of security of BB84 is actually difficult.*

Chapter 6

Alice and Bob

6.1 Two bits

Two classical bits allow us to count to 4:

$$\underbrace{ab}_{binary} = 2 \times a + b, \quad a, b \in 0, 1$$

Geometrically, 2 bits can be represented by the unit square in 2-dimensions. Two classical bits are just that, two single classical bits. We shall see that two

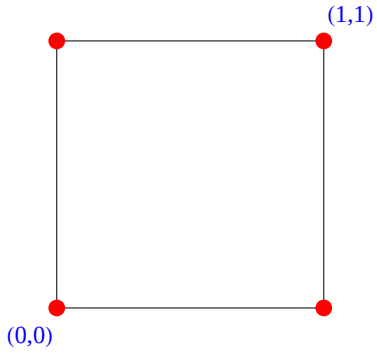


Figure 6.1: 2 bits correspond to the 4 corners of the unit square

qubits are more than what you would naively expect from two single qubits.

6.2 Hilbert space of two qubits: Tensor products

The Hilbert space of two qubits is $\mathbb{C}^4$ with a special structure that pays attention to who owns the qubits:

$$\mathbb{C}^4 = \underbrace{\mathbb{C}^2}_{Alice} \otimes \underbrace{\mathbb{C}^2}_{Bob}$$

A vector in the Hilbert space, in the computational basis, takes the form

$$|\psi\rangle = \sum_{a,b\in 0,1} \psi_{ab} |a\rangle_A \otimes |b\rangle_B$$

Think of the subscript ab as a binary representation of a number.

The the unit sphere in $\mathbb{C}^4$ is S^7:

$$1 = \langle\psi|\psi\rangle = \sum_{a,b\in 0,1} |\psi_{ab}|^2$$

Since $|\psi\rangle$ is physically equivalent to $e^{i\gamma}|\psi\rangle$ the space of physically distinct (pure) states is six dimensional

$$S^7/S^1$$

This the space is known as $CP(3)$. It is the space of rank one projections, $|\psi\rangle\langle\psi|$. It is clearly compact and six dimensional. (It is, however, not S^6.) It has interesting geometry.

The first thing to notice is that $CP(3)$ is larger than what you'd expect classically. The space of pure states of a qubit is S^2. Classically you'd expect the pure states of 2 qubits to be $S^2 \otimes S^2$. But it is not. It has 6 dimensions not 4.

6.3 Computational basis

The computational basis is the common eigenvectors of

$$Z \otimes Z, \quad Z \otimes \mathbb{1}, \quad \mathbb{1} \otimes Z$$

These observables are mutually commuting and so can be measured simultaneously. Measuring them Alice and Bob to prepares one of the 4 computational basis states:

$$|a\rangle_A \otimes |b\rangle_B \quad a, b, \in 0, 1$$

We shall use several shorthands

$$|a\rangle_A \otimes |b\rangle_B = |ab\rangle_{AB} = |2a + b\rangle \quad a, b, \in 0, 1$$

and

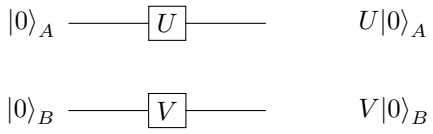
$$|00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

6.4 Pure product states

Definition 6.1. *The state $|\psi\rangle \otimes |\phi\rangle$, (equivalently, the associated projection, $|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|$), is called a pure product state.*

Pure product states can be identified with a point on two Bloch spheres: Alice's and Bob's. They are similar to classical states.

Pure product states are mapped to pure product states under local operations where Alice and Bob each each make their qubit go through a unitary gate



This is expressed by

$$(U_A \otimes V_B)\,|\psi\rangle_A \otimes |\phi\rangle_B = U_A|\psi\rangle_A \otimes V_B|\phi\rangle_B$$

This means that Alice and Bob can rotate their qubits anywhere on their respective Bloch spheres.

If Alice and Bob have a pure product state and do any single qubit measurement on their qubits, they will still get a pure product state. This follows from

$$|\psi\rangle \otimes |\phi\rangle \mapsto P_A \otimes P_B|\psi\rangle \otimes |\phi\rangle = P_A|\psi\rangle \otimes P_B|\phi\rangle$$

Product states describe the situation where Alice qubit is independent of Bob's qubit. For example, the probability that Alice finds her qubit pointing in the $\hat{\mathbf{m}}$ direction and Bob finding his pointing in the $\hat{\mathbf{n}}$ direction is

$$\begin{aligned} Prob(P(\hat{\mathbf{m}}) \otimes P(\hat{\mathbf{n}})|\phi \otimes \psi) &= Prob(P(\hat{\mathbf{m}})|\phi)\, Prob(P(\hat{\mathbf{n}})|\psi) \\ &= \langle\psi|P(\hat{\mathbf{m}})|\psi\rangle\langle\phi|P(\hat{\mathbf{n}})|\phi\rangle \end{aligned}$$

where

$$P(\hat{\mathbf{m}}) = \frac{\mathbb{1} + \hat{\mathbf{m}} \cdot \boldsymbol{\sigma}}{2}$$

To kick $|\psi\rangle \otimes |\phi\rangle$ out of the space of pure product Alice and Bob need to do a bona fide 2-qubit operation. For example, they need to measure $X \otimes X$ where the assocciated projection is

$$P = \frac{\mathbb{1} + X \otimes X}{2}$$

This is not something Alice and Bob can do by measuring the two qubits individually. They need to do something on the pair.

6.5 The algebra of tensor products

In the computational basis

$$\begin{aligned}(\mathbf{A}\otimes\mathbf{B})\,|jk\rangle &= \mathbf{A}|j\rangle\otimes\mathbf{B}|k\rangle\\ &= \sum_m \mathbf{A}_{mj}|m\rangle\otimes\sum_n \mathbf{B}_{nk}|n\rangle\\ &= \sum_{m,n}\mathbf{A}_{mj}\mathbf{B}_{nk}|mn\rangle\\ &= \sum_{m,n}(\mathbf{A}\otimes\mathbf{B})_{mn;jk}|mn\rangle\end{aligned}$$

$|jk\rangle$ is vector, with indices that runs on the binaries 00 to 11. The matrix $\mathbf{A}\otimes\mathbf{B}$ is the 4×4:

$$(\mathbf{A}\otimes\mathbf{B})_{mn;jk} = \mathbf{A}_{mj}\mathbf{B}_{nk}$$

Example 6.2.

$$\mathbb{1}\otimes X = \begin{pmatrix} X & 0\\ 0 & X\end{pmatrix},\quad X\otimes\mathbb{1} = \begin{pmatrix} 0 & \mathbb{1}\\ \mathbb{1} & 0\end{pmatrix},$$

Note that

$$(\mathbf{A}\otimes\mathbf{B})^\dagger = \mathbf{A}^\dagger\otimes\mathbf{B}^\dagger$$

Exercise 6.3. *Show that in the computational basis,*

$$\mathbf{A}\otimes\mathbf{B} = \begin{pmatrix} A_{00}\mathbf{B} & A_{01}\mathbf{B}\\ A_{10}\mathbf{B} & A_{11}\mathbf{B}\end{pmatrix},$$

A simple, yet often useful, result is: $X\otimes X$, $Y\otimes Y$, and $Z\otimes Z$ are all mutually commuting.

6.5.1 Partial trace

Definition 6.4 (Partial trace)**.** *The partial traces of $\mathbf{A}\otimes Y$ are defined by*

$$Tr_A\big(\mathbf{A}\otimes\mathbf{B}\big) = \mathbf{B}\ Tr_A\mathbf{A},\quad Tr_B\big(\mathbf{A}\otimes\mathbf{B}\big) = \mathbf{A}\ Tr_B\mathbf{B}$$

and then extended by linearity to $\sum X_j\otimes Y_j$ for any linear operator on $\mathcal{H}_A\otimes\mathcal{H}_B$.

Alternatively, if the operator $\mathbf{D}$ on the joint space has matrix elements $\mathbf{D}_{\alpha\beta;\alpha'\beta'}$ with α labeling the basis in Alice Hilbert space and β is Bob's then its partial trace has matrix elements:

$$(\mathbf{D}_A)_{\alpha\alpha'} = \sum_\beta \mathbf{D}_{\alpha\beta;\alpha'\beta}$$

It follows that

$$Tr_{\mathcal{H}_1\otimes\mathcal{H}_2}\,(\mathbf{A}\otimes\mathbf{B}) = (Tr_{\mathcal{H}_1}\,\mathbf{A})\ (Tr_{\mathcal{H}_2}\,\mathbf{B})$$

It is also easy to verify that

Proposition 6.5. *Partial trace maps positive operators (on the total space) to positive operators (on Alice space) and is trace preserving.*

6.6 The state of subsystems

Suppose Alice and Bob share ρ_{AB}. We would like to introduce a notion of a state ρ_A that captures what Alice know about her qubit if Bob's qubit is inaccessible to her. Alice observables are of the form

$$\mathbf{A} \otimes \mathbb{1}$$

since Bob is doing nothing on his qubit. The expectation values for any of here observables is

$$\begin{aligned} Tr\, \mathbf{A} \otimes \mathbb{1}\, \rho_{\mathbf{A}B} &= \sum \mathbf{A}_{\alpha\alpha'} \delta_{\beta\beta'} \rho_{\alpha'\beta';\alpha\beta} \\ &= \sum \mathbf{A}_{\alpha\alpha'} \rho_{\alpha'\beta;\alpha\beta} \\ &= Tr_A\big(\mathbf{A}\rho_A\big) \end{aligned} \tag{6.1}$$

where

$$\rho_A = Tr_B\left(\rho_{AB}\right)$$

The reduced density matrix ρ_A encodes the results of measurements Alice can do on her qubit.

Remark 6.6. *Partial trace is the analog of marginals in probability: The joint probability distribution $p(x, y)$ is the analog of ρ_{AB}. It describes the knowledge of Alice and Bob about the joint system. The marginals*

$$p(x) = \sum_y p(x, y) \Longleftrightarrow \rho_A = Tr_B \rho_{AB}$$

express what Alice knows about her subsystem.

Exercise 6.7. *Compute the partial traces ρ_A and ρ_B for*

$$\rho = \frac{1}{4} \sum \rho_{\mu\nu} \sigma_\mu \otimes \sigma_\nu, \quad \rho_{\mu\nu} \in \mathbb{R}, \quad \rho_{00} = 1 \tag{6.2}$$

Suppose you are told that ρ_A and ρ_B are fully mixed. What does this imply on $\rho_{\mu\nu}$?(Answer: $\rho_{\mu 0} = 0$).

6.7 Purification

Purification is the converse of partial tracing: We can always purify a density matrix at the price of adding an auxiliary system, known as ancilla (Latin for "maid").

Given a density matrix of the system S represented as a convex sum of pure states

$$\rho_S = \underbrace{\sum_j p_j |\psi_j\rangle\langle\psi_j|}_{\text{not necessarily orthogonal}}$$

The corresponding pure state of the joint system+ancilla is

$$|\Psi\rangle = \sum_j \sqrt{p_j}|\psi_j\rangle_S \otimes |j\rangle_A, \quad \langle j|k\rangle_A = \delta_{jk}$$

One verifies that

$$\rho_S = Tr_A|\Psi\rangle\langle\Psi|$$

6.8 Tomography of two qubits

A basis in the space of 4×4 matrices is

$$\sigma_\mu \otimes \sigma_\nu, \quad \mu, \nu \in \{0, 1, 2, 3\}$$

Any trace normalized state can be written as

$$\rho = \frac{1}{4}\left(\mathbb{1} + \sum_{\mu\nu\neq 00} \rho_{\mu\nu}\sigma_\mu \otimes \sigma_\nu\right), \quad \rho_{\mu\nu} \in \mathbb{R}$$

If we have a black box that spits out copies of an unknown ρ we can determine ρ by measuring 15 observables, e.g.

$$\rho_{\mu\nu} = Tr(\rho\, \sigma_\mu \otimes \sigma_\nu)$$

In practice it is often natural to replace the observables

$$\mathbb{1}, \quad X, \quad Y, \quad Z$$

by observables that are projections. One reason to do that is that projections represent detector counts. A set of linearly independent projections is, for example

$$\frac{\mathbb{1}+X}{2}, \quad \frac{\mathbb{1}+Y}{2}, \quad \frac{\mathbb{1}+Z}{2}, \quad \frac{\mathbb{1}-Z}{2}$$

From these you can construct $4^2 = 16$ pairs of correlations between detectors. The price you have to pay is that to determine $\rho_{\mu\nu}$ in the $\sigma_\mu \otimes \sigma_\nu$ basis you need to solve a linear algebra problem.

Exercise 6.8. *Suppose you are told that the state is pure. How many and which correlations would suffice to determine the state.*

6.8.1 Gauge invariance

Suppose Alice and Bob agree on the Z direction. The freedom to choose an arbirtary phase in the computational basis is expressed as gauge freedom

$$|ab\rangle' = U|ab\rangle, \quad U = \begin{pmatrix} e^{i\phi_1} & 0 & 0 & 0 \\ 0 & e^{i\phi_2} & 0 & 0 \\ 0 & 0 & e^{i\phi_3} & 0 \\ 0 & 0 & 0 & e^{i\phi_4} \end{pmatrix}$$

Under gauge transformation the tomography changes by

$$\rho_{\mu\nu} \mapsto \rho_{\mu\nu} e^{i(\phi_\mu - \phi_\nu)}$$

The diagonals are gague invariant, and the absolute values of the off diagonal elements are gauge invariant. Moreover, the phase for a product of any cosed cycle is guage invariant

$$\rho_{12}\rho_{23}\rho_{31} \mapsto \rho_{12}\rho_{23}\rho_{31}$$

Chapter 7

Entanglement

Most vectors $|\psi\rangle \in \mathbb{C}^4$ are not pure products: $|\psi_A\rangle \otimes |\psi_B\rangle$. This follows from a simple counting argument:

$$\dim(2 \text{ qubits pure states}) = 6, \quad \dim(2 \text{ qubits pure product states }) = 4$$

Pure states that are not pure products are called *entangled*.

The mother of entangled states are Bell states. There are four of them. Here is one

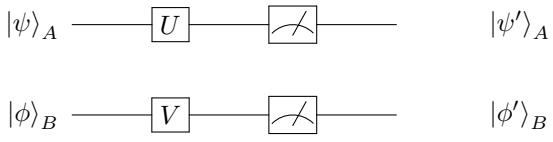
$$\sqrt{2}|\beta_0\rangle = |0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B \tag{7.1}$$

It may not always be obvious if a state is or is not pure product. For example

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = |+\rangle \otimes |+\rangle \tag{7.2}$$

the lhs is a superposition in the computational basis, but the rhs shows that, in suitable basis, it is is a pure product. In the next section we shall describe a mathematical tool that will decide if a pure state is entangled on not.

The fact that most pure states are entangled does not mean that it is necessarily easy to make them. If Alice and Bob start from a pure product state $|\psi_A\rangle \otimes |\psi_B\rangle$ and all they are allowed to do are local operation, the state will transform to another pure product state:



7.1 Schmidt decomposition

How can one tell if a (pure) state shared by two parties is entangled or not? When is general 2-parties state pure product:

$$\sum_{a,b,\in 0,1} \psi_{ab}|a\rangle \otimes |b\rangle ? = ?|\psi\rangle \otimes |\phi\rangle$$

The representation on the left depends on the choice of single qubit basis, as we have seedn in Eq. 7.2. How can we choose an optimal basis so that the number of terms in the superposition is minimal? The Schmidt decomposition selects the optimal basis adjusted to the state.

The basic idea is similar to diagonalization of a normal matrix: An $N \times N$ normal matrix is encoded in the N eigenvalues, if we chose the optimal basis–the basis of eigenvectors.

The coefficients ψ_{ab} can be organized as a 2×2 matrix. The matrix of coefficients need not be Hermitian. For example,

$$|0\rangle \otimes |1\rangle \leftrightarrow \psi_{ab} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The singular value decomposition, that we shall now describe, allows to to represent any $N \times N$, matrix, not necessarily normal, by N entries, known as singular values, that are natural generalizations of the eigenvalues of Normal matrices.

To see how this comes about recall the polar decomposition of a complex number $z = re^{i\theta}$. This simple fact has an analog for (square[1]) matrices:

Theorem 7.1. *Every square matrix A can be written as a product of a unitary and a positive matrix:*

$$A = U|A|, \quad U^\dagger = U^{-1}, \quad |A| = \sqrt{A^\dagger A}$$

(uniquely).

Example 7.2.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{unitary} \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{positive}$$

This leads to the singular values decomposition (SVD):

Theorem 7.3 (SVD)**.** *Every square matrix A can be written in terms of the product of two unitary matrices U, W and a diagonal positive matrix D*

$$A = (unitary)(diagonal)(unitary) = W\,D\,U$$

where

$$D_{nm} = \delta_{n,m} Eigenvalue_n \left(\sqrt{A^\dagger A}\right)$$

D_n are known as "singular values". The decomposition is unique up to reordering of the singular values in D and the freedom to multiply D on the left by a diagonal unitary and on the right by its inverse).

[1] Append columns or rows of zeros if not square.

Proof: Since $|A|$ is a positive matrix, we can write

$$|A| = VDV^\dagger$$

with V unitary and D diagonal and positive. Combined with the polar decomposition gives the result. SVD has numerous applications.

SVD leads to

- Any state partitioned between Alice and Bob can be brought into canonical form

$$\sum_{mn} \psi_{mn}|m\rangle \otimes |n\rangle = \underbrace{\sum_m \lambda_m |\psi_m\rangle \otimes |\phi_m\rangle}_{canonical} \tag{7.3}$$

  with

$$\langle \psi_m|\psi_n\rangle = \delta_{mn} \quad \langle \phi_m|\phi_n\rangle = \delta_{mn}$$

- λ_m are called the Schmidt coefficients.
- The number of non-zero Schmidt coefficients is called the Schmidt rank.
- A state is a product state if its Schmidt rank is 1 and is entangled it its Schmidt rank is larger than 1.
- A state is maximally entangled if all $|\lambda_m|$ are equal.

Note that in the lhs of Eq. (7.3) different terms are orthogonal because one of the factors $|m\rangle \otimes |n\rangle$ is orthogonal whereas in the rhs both factors $|\psi_m\rangle \otimes |\phi_m\rangle$ are mutually orthogonal. This allows to identify when a state is in Schmidt form.

Proof:

$$\begin{aligned}
\sum_{j,k=1}^{N} \psi_{jk}|j\rangle \otimes |k\rangle &= \sum_{jkm} W_{jm} D_m V^\dagger_{mk} |j\rangle \otimes |k\rangle \\
&= \sum_m D_m \left(\sum_j W_{jm}|j\rangle \right) \otimes \left(\sum_k V^\dagger_{mk}|k\rangle \right) \\
&= \sum_{m=1}^{N} D_m \underbrace{W|m\rangle \otimes V^*|m\rangle}_{unitary\ change\ of\ base}
\end{aligned} \tag{7.4}$$

The λ_m can be chosen non-negative. (The Schmidt coefficients give the equivalence class of states that are related by local unitary operations.) Clearly

$$\langle Wm'|Wm\rangle = \langle m'|W^\dagger W|m\rangle = \delta_{mm'}$$

7.2 Bell pairs

The Bell states, named after John Bell whom we shall meet again later, are the 2-qubits states given by:

$$|\beta_\mu\rangle = \frac{1}{\sqrt{2}} \sum_{a,b=0}^{1} (\sigma_\mu)_{ab} |ab\rangle, \quad \mu \in 0, \ldots, 3, , \quad a, b \in 0, 1$$

Explicitly

$$\sqrt{2}|\beta_0\rangle = |00\rangle + |11\rangle, \quad \sqrt{2}|\beta_1\rangle = |10\rangle + |01\rangle$$

$$i\sqrt{2}|\beta_2\rangle = |10\rangle - |01\rangle, \quad \sqrt{2}|\beta_3\rangle = |00\rangle - |11\rangle$$

- Being in Schmidt form Bell states are maximally entangled.
- $$\langle \beta_\mu | \beta_\nu \rangle = \frac{1}{2} Tr(\sigma_\mu \sigma_\nu) = \delta_{\mu\nu}$$
- Bell states are an orthogonal basis for $\mathbb{C}^2$.
- My notation differs from Nielsen and Chuang by the overall phase in $|\beta_2\rangle$.

Exercise 7.4. *Show that*

$$\underbrace{\sqrt{2}|ab\rangle = \sum_\mu (\sigma_\mu)_{ba} |\beta_\mu\rangle}_{Note\ ordering\ of\ ab}$$

Exercise 7.5. *Purify the two qubits state*

$$\rho = p|\beta_0\rangle\langle\beta_0| + (1-p)\frac{\mathbb{1}}{4}$$

(Hint: Write the resolution of the identity in terms of Bell states).

7.2.1 Syndrome

Often, interesting wave functions turn out to be quite complicated and you do not learn much by looking at the wave function. Instead, you can characterizer them by enough commuting observables. Since

$$X \otimes X, \quad Y \otimes Y, \quad Z \otimes Z$$

are mutually commuting, have eigenvalues ± 1, and satisfy

$$\big(X \otimes X\big)\big(Y \otimes Y\big)\big(Z \otimes Z\big) = (XYZ) \otimes (XYZ) = -\mathbb{1} \otimes \mathbb{1}$$

The four Bell states correspond to the eigenvalues:

$$\begin{aligned} |\beta_0\rangle: &\quad \{1,-1,1\} \\ |\beta_1\rangle: &\quad \{1,1,-1\} \\ |\beta_2\rangle: &\quad \{-1,-1,-1\} \\ |\beta_3\rangle: &\quad \{-1,1,1\} \end{aligned}$$

Exercise 7.6. *Show that*

$$|\beta_0\rangle\langle\beta_0| = \frac{1}{4}(\mathbb{1}\otimes\mathbb{1} + X\otimes X - Y\otimes Y + Z\otimes Z) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

$$|\beta_3\rangle\langle\beta_3| = \frac{1}{4}(\mathbb{1}\otimes\mathbb{1} - X\otimes X + Y\otimes Y + Z\otimes Z) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix},$$

Similarly, for the projections on the remaining Bell states

$$|\beta_{1,2}\rangle\langle\beta_{1,2}| = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \pm 1 & 0 \\ 0 & \pm 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

7.2.2 Rotations: Bell singlet

The generator of the joint rotation of the two qubits is

$$\mathbf{J} = \frac{1}{2}(\boldsymbol{\sigma}\otimes\mathbb{1} + \mathbb{1}\otimes\boldsymbol{\sigma})$$

and explicitly

$$2J_z = Z\otimes\mathbb{1} + \mathbb{1}\otimes Z, \quad 2J_x = X\otimes\mathbb{1} + \mathbb{1}\otimes X, \quad 2J_y = Y\otimes\mathbb{1} + \mathbb{1}\otimes Y.$$

The Bell state $|\beta_2\rangle$ is an eigenstate of the total angular momentum with eigenvalue 0. It follows that it is invariant under rotations and so is isotropic. It is also known as the singlet.

7.3 Generating Bell pairs

Consider a quantum dot, where the excited state $|e^2h^2; J=0\rangle$ is a bound pair of two electrons and two holes with total angular momentum $J=0$. The dot has a degenerate intermediate level $|eh; J=\pm 1\rangle$ which is optically connected to

top state. The ground state is an empty dot $|0; J = 0\rangle$. The first e-h pair can recombine to emit a red photon. It can do that in one of two paths. QM doess not chose a path–it uses both

$$\left|e^2h^2; J = 0\right\rangle \mapsto |eh, J = -1\rangle \otimes |L\rangle + |eh, J = 1\rangle \otimes |R\rangle$$

This state has the dot is entangled with the red photon. The remaining electron-hole pair can recombine to leave the dot in its ground state while emitting a blue photon.

$$|eh; J = -1\rangle \otimes |L\rangle + |eh, J = 1\rangle \otimes |R\rangle \mapsto \underbrace{|0; J = 0\rangle}_{\text{empty dot}} \Big(|R\rangle \otimes |L\rangle + |R\rangle \otimes |R\rangle\Big)$$

The dot and the photons are in a product state. But, now the two photons are entangled.

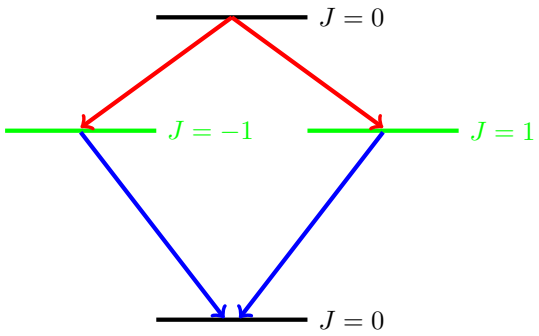


Figure 7.1: A three level system. The top state has $J = 0$ and decays to a ground state, also with $J = 0$, through intermediate levels with $J = \pm 1$, emitting two photons in th process. The first photon is red and the second photon is blue. There are two decay paths. The right arm emits $|L\rangle \otimes |R\rangle$ where R, L denote the circular polarization of the photons. The left arm emits $|R\rangle \otimes |L\rangle$

7.4 Dense coding

An interesting feature of Bell states is their non-locality: Alice can affect Bobs' qubit even if she has no direct access to it: Either Alice or Bob can turn any Bell state to any other Bell state by local operations

$$\sigma_\mu \otimes \mathbb{1}|\beta_0\rangle = |\beta_\mu\rangle, \quad \mathbb{1} \otimes \sigma_\mu^t|\beta_0\rangle = |\beta_\mu\rangle$$

(The annoying transpose only affects σ_2 and is just an overall sign.) In other words, starting with one Bell state, Alice can generate all Bell states, which

span a basis in the Hilbert. This is not possible in the computational basis for Alice can not chage $|00\rangle$ to the vector $|11\rangle$.

Transmission of a single qubit of a Bell pair from Alice to Bob is equivalent to the transmission of two classical bits between them.

Protocol:

- Alice and Bob share $|\beta_0\rangle$.
- Alice wants to transmit $\mu \in \{0, 1, 2, 3\}$
- Alice acts on her qubit by σ_μ. This turns the state to $|\beta_\mu\rangle$.
- Alice sends Bob her qubit.
- Bob measures $X \otimes X$ and $Z \otimes Z$ (syndrom) on the Bell pair and determines μ.

The one qubit Bob got from Alice is worth 2 classical bits to Bob.

7.5 Entanglement = incomplete knowledge about subsystems

If Alice and Bob share a pure state $|\Psi\rangle_{AB}$ with Schmidt decomposition

$$|\Psi\rangle_{AB} = \sum_m \lambda_m |\psi_m\rangle \otimes |\phi_m\rangle$$

then Alice system has the density matrix

$$\rho_A = \sum_m |\lambda_m|^2 |\psi_m\rangle\langle\psi_m|$$

and Alice is entangled with Bob iff ρ_A is mixed and is maximally entangled if her state is fully mixed. Only if $|\Psi\rangle_{AB} = |\psi\rangle \otimes |\phi\rangle$ is pure product then Alice system is in a pure statee: $Tr_B |\Psi\rangle_{AB}\langle\Psi| = |\psi\rangle\langle\psi|$ is a projection.

The result is interesting as it illustrates a basic difference between the quantum and the classical worlds. In classical physics, if you know all there is to know about Alice and Bob, you also know all there is to know about Alice. This is not true for entangled states. In fact, if the state is fully entangled then Alice knows nothing about her qubit.

If Alice and Bob share a Bell pair, then Alice knows nothing about her qubit: The state is fully mixed:

$$\rho_A = Tr_B |\beta_\mu\rangle\langle\beta_\mu| = \tfrac{1}{2}\mathbb{1}$$

Only pure product states behave like classical systems.

Exercise 7.7. *Show that if Alice and Bob are maximally entangled, Alice's qubit is a perfect balanced coin in the sense that measurement of the qubit in any direction* $\mathbf{m}$ *of the Bloch sphere has*

$$\langle\beta_\mu|(\mathbf{m}\cdot\sigma)\otimes\mathbb{1}|\beta_\mu\rangle = 0$$

7.6 Perfect correlations and total ignorance

As we have seen when Alice and Bob share a Bell state Alice knows nothing about her own qubit. And so does Bob. Interestingly, the two coins are nevertheless perfectly correlated. Consider for simplicity the case that Alice and Bob share $|\beta_2\rangle$:

$$\rho_0 = |\beta_2\rangle\langle\beta_2| = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{7.5}$$

The probability that Alice finds her qubit pointing in the $\mathbf{m}$ direction and Bob finds his pointing in the $\mathbf{n}$ direction is:

$$Prob(P(\mathbf{m})\otimes P(\mathbf{n})|\beta_2) = \langle\beta_2|\underbrace{P(\mathbf{m})\otimes P(\mathbf{n})}_{projections}|\beta_2\rangle = \frac{1-\mathbf{m}\cdot\mathbf{n}}{4}, \tag{7.6}$$

where

$$P(\mathbf{m}) = \frac{\mathbb{1}+\mathbf{m}\cdot\boldsymbol{\sigma}}{2}, \quad |\mathbf{m}| = 1$$

The conditional probability, conditioned on Alice finding her qubit in $\mathbf{m}$, is

$$\frac{Prob(P(\mathbf{m})\otimes P(\mathbf{n})|\beta_2)}{Prob(P(\mathbf{m})\otimes\mathbb{1}|\beta_2)} = \frac{1-\mathbf{m}\cdot\mathbf{n}}{2}$$

The correlations are perfect when $\mathbf{m} = -\mathbf{n}$ where the rhs is 1 so that Bob's qubit becomes a faithful slave of Alice's qubit.

Exercise 7.8. *Show that*

$$\begin{aligned}\langle\beta_\mu|(\mathbf{m}\cdot\sigma)\otimes(\mathbf{n}\cdot\sigma)|\beta_\mu\rangle &= Tr\left(P_\mu\,(\mathbf{m}\cdot\sigma)\otimes(\mathbf{n}\cdot\sigma)\right) \\ &= \begin{cases} m_1n_1 - m_2n_2 + m_3n_3 & \mu = 0 \\ m_1n_1 + m_2n_2 - m_3n_3 & \mu = 1 \\ -m_1n_1 - m_2n_2 - m_3n_3 & \mu = 2 \\ -m_1n_1 + m_2n_2 + m_3n_3 & \mu = 3 \end{cases}\end{aligned}$$

7.7 Correlations and signaling

Contrary to a popular myth, correlations do not imply signaling. Bell likes to tell the story of his friend Bertlesman, who always wears one red sock and one blue. But chooses right and left randomly. If Alice lifts Bertlesman hose on the left foot, she knows the color on both feet. But her knowledge has not been transmitted to Bob at the other leg.

7.8 Remote state preparation, Heralding

Protocol

- Alice and Bob share the singlet Bell state,

$$|\beta_0\rangle = |0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B$$

- Alice measures her qubit in the computational basis.
- Alice tells Bob what she found.
- Bob now knows that his qubit is in the same state as Alice's without actually measuring it.
- Alice prepared Bob's qubit at a distance.

You should worry how does a state change instantaneously upon measuring a remote part of the system and if there is conflict with relativity. There no problem with this because

- Alice measurement gives a probabilistic result so her measurement does not transmit information to Bob who has no idea what the result has been.
- Heralding is deterministic provided Alice has a classical channel to transmit her result to Bob. Since classical transmissions have finite propagation speed relativity is safe. Heralding is not a method to signal faster than light.

7.9 Separable and entangled states

So far we have discussed entanglement for pure states. Let us now extend the notion of entanglement to mixed states.

Let us start with pure product states

$$\rho = \rho_A \otimes \rho_B$$

The operator $P_A \otimes P_B$ with P_A and P_B projections represent the occurrence of the event that Alice prepares the state P_A and Bob prepared P_B. If Alice and Bob share the state ρ this occurs with probability

$$\begin{aligned} Prob(P_A \otimes P_B|\rho) &= Tr\ (\rho P_A \otimes P_B) \\ &= Tr_A\ (\rho_A P_A)\ Tr_B\,(P_B \rho_B) \\ &= Prob(P_A|\rho_A)\, Prob(P_B|\rho_B) \end{aligned} \tag{7.7}$$

This means that Alice and Bob probabilities are independent.

Consider now the case that the quantum state of Alice and Bob is a mixture of product states:

$$\rho = \sum_j p_j (\rho_A)_j \otimes (\rho_B)_j, \quad 1 \geq p_j \geq 0 \tag{7.8}$$

One way to think about this state is to imagine that Alice has a pool of states $(\rho_A)_j$ and Bob has a pool $(\rho_B)_j$. Alice and Bob prepare ρ by throwing a dice. If the dice gives j they pick $(\rho_A)_j$ and $(\rho_B)_j$. The do that many time and use the state to make various measurements. ρ is the state they prepared on the average.

Since Alice and Bob use a the same dice, they are correlated. These correlations are classical. This motivates the definition of special class of of states, called *separable* states.

- A state is separable if it is of the formEq. 7.8.
- The separable states are a convex subset of the set of all states.
- The fully mixed state is separable:

$$\frac{\mathbb{1}_A \otimes \mathbb{1}_B}{2^{n+m}}$$

- There are density matrices $\rho \geq 0$ that are non-separable.
- A density matrix that is not separable is called *entangled*.

As we shall see this definition generalizes the definition of entanglement that we gave for pure states.

The definition of separability suggests that entangled states can not be described by classical probability theory–they have non-classical correlations. We shall see that this is indeed the case.

7.10 Bell states for q-dits

Let $\mathbb{C}^d$ be the Hilbert space of Alice with the computational basis $|n\rangle,\ n \in \mathbb{Z}_d$. Define

$$S|n\rangle = \omega|n\rangle, \quad T|n\rangle = |n-1\rangle, \quad \omega = e^{2\pi i/d}$$

Then, the $d^{@}$ states

$$|\beta_{jk}\rangle = \frac{1}{\sqrt{d}} \sum_{n=1}^{d} T^j|n\rangle \otimes S^k|n\rangle$$

are maximally entangled and mutually orthogonal. They are the generalization of Bell states to q-dits.

Exercise 7.9. *Show this.*

7.11 Partial transpose

Given a bipartite partition, the partial transpose P is the linear operation defined on the base vectors of the partition $\boldsymbol{\sigma_\mu} \otimes \boldsymbol{\sigma_\alpha}$ by

$$\left(\boldsymbol{\sigma_\mu} \otimes \boldsymbol{\sigma_\alpha}\right)^P = \boldsymbol{\sigma_\mu} \otimes \boldsymbol{\sigma}_{\boldsymbol{\alpha}}^t$$

Since σ_y is anti-symmetric while σ_x and σ_z are symmetric:

$$\sigma_y^t = -\sigma_y, \quad \sigma_x^t = \sigma_x, \quad \sigma_z^t = \sigma_z$$

In particular, if $\rho \geq 0$ is separable:

$$\rho = \sum_j p_j (\rho_A)_j \otimes (\rho_B)_j \tag{7.9}$$

with $(\rho_A)_j \geq 0$, $(\rho_B)_j \geq 0$ then also ρ^P is also a (positive) state $\rho^P \geq 0$ and is also separable:

$$\rho^P = \sum_j p_j (\rho_A)_j \otimes (\rho_B)_j^t, \tag{7.10}$$

since $(\rho_A)_j \geq 0$ and $(\rho_B)_j^t \geq 0$.

The set of states that satisfies $\rho^P \geq 0$ is a convex set.

Exercise 7.10. *Show that the partial transpose of the projection on the Bell state $|\beta_0\rangle\langle\beta_0|$ is the swap.*

Remark 7.11. *Transpose is basis dependent in the following sense*

$$A \mapsto A^t \iff U^\dagger A U \mapsto (U^\dagger A U)^t \neq U^\dagger A^t U$$

7.11.1 Time reversal

For self-adjoint operators transpose is the same as complex conjugation

$$A \mapsto A^t = (A^\dagger)^t = (\bar{A}^t)^t = \bar{A}$$

Complex conjugation has an interesting interpretation. Namely, time reversal in quantum mechanics is related to complex conjugation. This can be seen in several ways. For example, under time reversal

$$\mathbf{x} \mapsto \mathbf{x}, \quad \mathbf{p} \mapsto -\mathbf{p}$$

To preserve the uncertainty principle[2]

$$[p_j, x_k] = -i\hbar\delta_{jk}$$

[2] More precisely, this says that time reversal is (an anti-linear) operator made from a unitary times complex conjugation: $U*$.

we need to replace i by $-i$. The same argument works for the angular momentum commutation

$$[J_j, J_k] = -i\hbar\epsilon_{ijk}J_i$$

Yet another way to see this is to look at Schrödinger equation

$$-i\hbar\partial_t|\psi\rangle = H|\psi\rangle \Longrightarrow \overline{-i\hbar\partial_t|\psi\rangle} = \overline{H|\psi\rangle} \Longrightarrow i\hbar\partial_t|\bar{\psi}\rangle = \bar{H}|\bar{\psi}\rangle$$

is like flipping time.

7.12 Peres test

The truly remarkable fact about the partial transpose is that it is, in general, not a complete positivity preserving map. By this I mean that there are positive $\rho \geq 0$ such that ρ^P is not positive. This leads to Peres test:

If a density matrix $\rho \geq 0$ has a partial transpose that is non-positive, then ρ is entangled.

The case of two-qubits is special in that:

- The test is both necessary and sufficient condition for entanglement. (I shall not prove this.)
- ρ^P has at most one negative eigenvalue.
- The absolute value of the negative eigenvalue is called *negativity*. It is a measure of entanglement.

7.13 Consistency of the definitions of entanglements for pure and mixed states

Let us now show that the notion of entanglement as the complement of the set of separable states generalizes the notion of entanglement for pure states.

For pure states we defined entanglement through Schmidt decomposition:

$$|\Psi\rangle = \sum_{j=1}^{M} \lambda_j |j\rangle \otimes |j\rangle, \quad \lambda_j > 0$$

$|\Psi\rangle$ is entangled if $M \geq 2$.

The density matrix associated with the state $|\Psi\rangle$ is

$$|\Psi\rangle\langle\Psi| = \sum_{j,k} \lambda_j\lambda_k |j\rangle\langle k| \otimes |j\rangle\langle k|$$

Its partial transpose is

$$(|\Psi\rangle\langle\Psi|)^P = \sum_{j,k} \lambda_j\lambda_k |j\rangle\langle k| \otimes |k\rangle\langle j|$$

It is not too difficult to guess the eigenvectors of $(|\Psi\rangle\langle\Psi|)^P$.

- There are M eigenvectors, namely, $|m\rangle \otimes |m\rangle$ with positive eigenvalues $\lambda_m^2 > 0$
- There are $\binom{M}{2}$ eigenvectors[3]with eigenvalues $\pm\lambda_n\lambda_m$, namely:

$$|\phi_\pm\rangle = |n\rangle \otimes |m\rangle \pm |m\rangle \otimes |n\rangle$$

Indeed

$$\begin{aligned}(|\Psi\rangle\langle\Psi|)^P |\phi_\pm\rangle &= \sum_{j,k} \lambda_j\lambda_k\big(\delta_{kn}\delta_{jm} \pm \delta_{km}\delta_{jn}\big)|j\rangle \otimes |k\rangle \\ &= \lambda_n\lambda_m \sum_{j,k} \big(\delta_{kn}\delta_{jm} \pm \delta_{km}\delta_{jn}\big)|j\rangle \otimes |k\rangle \\ &= \pm\lambda_m\lambda_n|\phi_\pm\rangle\end{aligned}$$

This shows that pure states which are entangled by Schmidt decomposition, are also entangled by the Peres test. The notion of entanglement for mixed states therefore generalizes the notion of entanglement for pure states.

7.14 Entangled photons from a quantum dot

In semiconductors[4] the conductance band is made form atomic s orbitals and the valence band is made of atomic p orbitals. As a consequences, there is one type of electron, with spin $\pm 1/2$, denoted $|\uparrow\rangle, |\downarrow\rangle$ and two types of holes: Those with spin $\pm 1/2$, called light, and $\pm 3/2$, called heavy. We denote the heavy holes $|\Uparrow\rangle, |\Downarrow\rangle$[5].

A dot may be populated by electrons and holes. The ground state is an empty dot. The excited states associated an electron-hole pairs are called excitons. An exciton level is called bright if the the electron-hole pair can recombine effectively to emit a photon and leave the dot empty. This happens if the exciton has unit angular momentum. The two bright exciton levels are made with heavy holes [6]:

$$|\uparrow\Downarrow\rangle \pm |\downarrow\Uparrow\rangle$$

which is a superposition of angular momentum ± 1 states.

Since a p orbital is like $x + iy$, the two superpositions have the geometry of two antennas one is x oriented and the other y oriented. As a consequence when the electron and hole recombine to emit a photon the radiation from the two levels are H and V polarized respectively. Consider first the left path: The

[3]Clearly, $\binom{1}{2} = 0$ since you can not choose 2 terms from a list with one term.

[4]The situation is reversed in topological insulators.

[5]$|\Uparrow\rangle = |p\rangle \otimes |\uparrow\rangle_h |\phi(r)\rangle$ where $|p\rangle$ is the atomic orbital and $|\phi(r)\rangle$ a slowly varying envelop. The light holes are more complicated.

[6]The $\pm$ from of the eigenstates is dictated by time reversal.

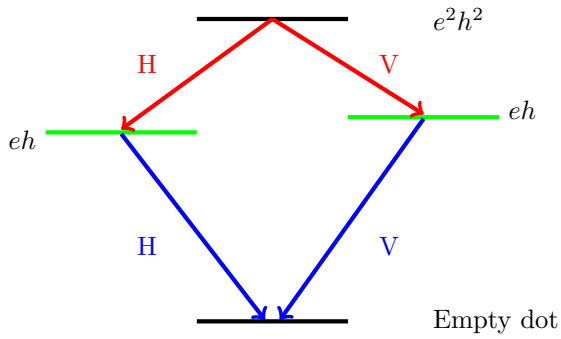


Figure 7.2: Level diagram in quantum dot. Two electron hole pairs make the top level (biexciton). When a pair recombines a (red) photon is emitted and the dot relaxes to the intermediate exciton with the remaining eh pair. This can happen in two paths ainces there are two intermediate exciton states. The two exciton levels (green) are not exactly degenerate and may be thought of as having the geometry of an antenna that is either x or y oriented. This implies that the the emitted photon is H polarized in one path and V polarized in the other. When the second pair recombines a second photon–blue–is emitted and the dot is empty. One decay path emits two horizontally polarized photons and one two vertically polarized.

wave function of the emitted photon pair with horizontal polarization is

$$\underbrace{|H,R_H\rangle}_{first}\otimes\underbrace{|H,B_H\rangle}_{second}=\underbrace{|H\rangle}_{Polarization}\otimes\underbrace{|R_H\rangle}_{color}\otimes\underbrace{|H\rangle}_{Polarization}\otimes\underbrace{|B_H\rangle}_{color}$$
$$=\underbrace{|H,H\rangle}_{polarization}\otimes\underbrace{|R_H,B_H\rangle}_{colors}$$

And similarly for the decay path on the right, but with H replaced by V. Thus, the emitted photon pairs are in an entangled state

$$\frac{|H,H\rangle\otimes|R_H,B_H\rangle+|V,V\rangle\otimes|R_V,B_V\rangle}{\sqrt{2}}$$

The state of the polarization qubits is obtained by tracing over the color. This gives the density matrix (in the basis HH,HV,VH,VV)

$$\rho=\frac{1}{2}\begin{pmatrix}1&0&0&c\\0&0&0&0\\0&0&0&0\\c^*&0&0&1\end{pmatrix},\quad c=\langle R_V,B_V|R_H,B_H\rangle$$

It represents an entangled state if $c\neq 0$.

NATURE|Vol 439|12 January 2006

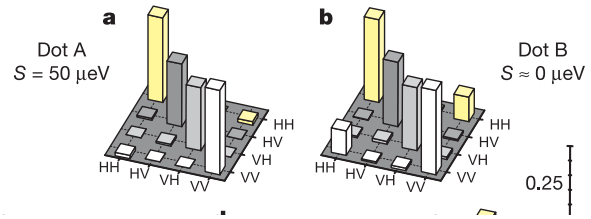


Figure 7.3: The shown two qubit tomography from a Nature article claims evidence for entanglement. The authors did not know about the Peres test for entanglement and invented their own bogus test which says that the non-zero corners of the density matrix imply entanglement. In this case it does not because of the large entries on the diagonal. The journal Nature is a business enterprise and not an non-for-profit organization. Do not build up high expectation with regard to honesty.

Exercise 7.12. *Two tomographies of a qubit pair (photon polarization) are shown below. One published in Nature and one in PRL. Both claim evidence for entanglement. Determine, using Perese test, which of the two claims is correct and which is wrong.*

If the colors give no information about the decay path $|c| = 1$ and the polarization is fully entangled. The case $c = 0$ is when the colors give complete "which path" information on the decay path. In this case the polarization is mixed but not entangled. $|c| = 1$ if the two exciton states were degenerate. In practice the degeneracy is split. If the split is small compared to the line width then $|c| \approx 1$. If the degeneracy lifting is large compared with the natural width, then $c \approx 0$. The colors give "which path" information if $c = 0$.

This is the description if the state was pure. In real systems the state of the system is never a pure state, and is contaminated my mixtures. This means that the 0 in the table are replaced by non zero entries. You can then use Peres to determine entanglement.

Exercise 7.13. *Two tomographies of a qubit pair (photon polarization) are shown below. One published in Nature and one, from Gershoni's lab, in PRL. Both claim evidence for entanglement. Determine, using Perese test, which of the two claims is correct and which is wrong.*

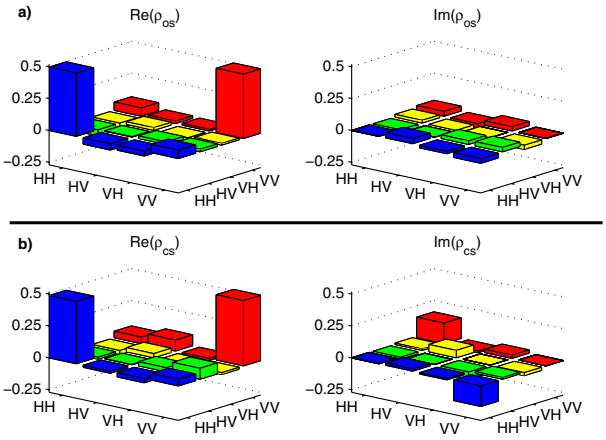


FIG. 3 (color online). The measured two photons' density matrix for photon pairs from a biexciton cascade: (a) [(b)] obtained with spectral window of 200 [25] μeV.

Figure 7.4: Two qubit tomography from a PRL (Akopian et. al.) with evidence for entanglement since the corner entries are larger than those on the diagonal.

7.15 The geometry of two qubits

The space of states of 2 qubits is 15 dimensional, and difficult to visualize. Consider the 3-dimensional cross section[7]

$$\rho = \frac{1}{4}\Big(\mathbb{1} + \sqrt{3}\big(xX\otimes X + yY\otimes Y + zZ\otimes Z\big)\Big)$$

Since

$$\text{Spectrum}(X\otimes X) = \pm 1$$

and

$$Z\otimes Z = -(X\otimes X)\,(Y\otimes Y)$$

the four eigenvalues of ρ are given by

$$\text{Spectrum}(\rho) = \frac{1}{4} + \frac{\sqrt{3}}{4}\Big(x\eta + y\tilde{\eta} - z\eta\tilde{\eta}\Big), \quad \eta,\tilde{\eta} = \pm 1$$

It follows that $\rho \geq 0$ provided

$$x\eta + y\tilde{\eta} - z\eta\tilde{\eta} \geq -\frac{1}{\sqrt{3}}$$

[7]The reason for the strange $\sqrt{3}$ normalization is consistency with Eq. **??**.

Each of the four conditions forces (x, y, z) to lie in a half-space. Since there are four conditions for $\eta = \pm 1, \tilde{\eta} = \pm 1$ the region in $\mathbb{R}^3$ that satisfies all 4 conditions is a tetrahedron. By symmetry, it is the regular tetrahedron.

Now, Peres test is both necessary and sufficient for two qubits. Since the Peres test flips the sign of Bob's Y we have

$$\rho^P = \frac{1}{4}\Big(\mathbb{1} + \sqrt{3}\big(xX \otimes X - yY \otimes Y + zZ \otimes Z\big)\Big)$$

It follows that $\rho^P \geq 0$ if

$$x\eta - y\tilde{\eta} - z\eta\tilde{\eta} \geq -\frac{1}{\sqrt{3}}$$

This gives 4 additional half-spaces. It follows that the set of separable states is the octahedron shown in the figure.

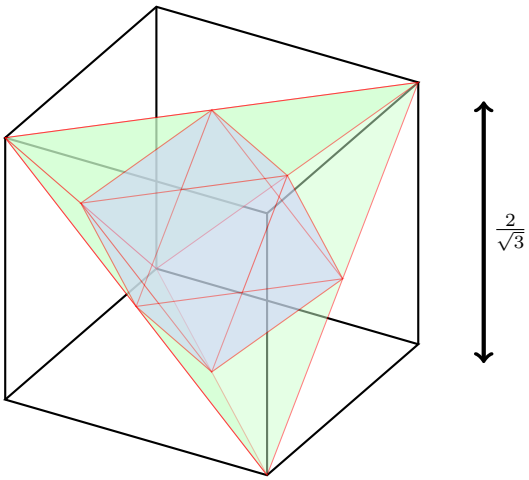


Figure 7.5: The figure shows the tetrahedron of states. The 4 maximally entangled Bell states at 4 of the 8 corners of a cube whose vertexes are $(\pm 1, \pm 1, \pm 1)/\sqrt{3}$. The separable states are the octahedron, which is the intersection of two tetrahedra. It is shown in blue.

7.16 Witnesses

Definition 7.14. *An observable W is called a witness if for every pure product state*

$$_A\langle\psi| \otimes {}_B\langle\phi|W|\psi\rangle_A \otimes |\phi\rangle_B \geq 0$$

It follows that if W is a witness then for any separable state ρ_s

$$Tr\left(\rho_s W\right) \geq 0$$

Hence if for some state

$$Tr\left(\rho_e W\right) < 0$$

then ρ_e is necessarily entangled. Any positive operator is a witness, but a boring

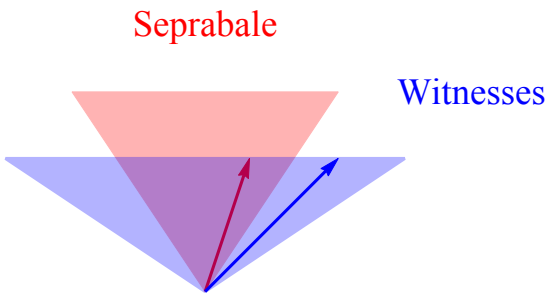


Figure 7.6: You may think of $Tr\ W\rho$ as scalar product between W and ρ. Witness have a positive scalar product with separable states, but have a neagative product with some, not all, entangled states.

one since it does not identify any entangled state. An witness is effective if it identifies some entangled states.

Example 7.15. *Swap is an effective witness*

Indeed, it is a witness since

$$_A\langle\psi| \otimes_B \langle\phi|Swap|\psi\rangle_A \otimes |\phi\rangle_B = \left({}_A\langle\psi| \otimes_B \langle\phi|\right)\left(|\phi\rangle_A \otimes |\psi\rangle_B\right) = |\langle\psi|\phi\rangle|^2 \geq 0$$

It is effective since it identifies the Bell singlet as entangled:

$$Swap\left(|01\rangle - |10\rangle\right) = -\left(|01\rangle - |10\rangle\right)$$

we have

$$\left(\langle 01| - \langle 10|\right) Swap\left(|01\rangle - |10\rangle\right) = -2$$

Exercise 7.16. *Show that the partial transpose of the projections on any Bell states is an effective witness.*

7.17 Whose wave function is it anyway?

Classical probability distributions do not describe the system, but describe our imperfect knowledge about it. If I know more than you do about a given system, then I would assign a different probability distribution. For example you assign $p(x)$ for the voters for Trump and Clinton, i.e. $x = \{Trump, Clinton\}$, and I assign $p(x, y)$ with $y = \{Male, Female\}$. I know more, and your $p(x)$ is a marginal of my $p(x, y)$. There is nothing wrong with one system with multiple properties being described by various probability distributions.

What about a quantum state. Is it OK for me and you to assign different states to a given system depending on our knowledge of it, or is the state a property of the system? It turns out that the answer is basically the same as in the classical case: It is OK provided we agree on the marginals.

Suppose Alice and Bob are separated by 1/2 light second and share $|\beta_2\rangle$ at $t = -1$. They agree beforehand that Alice will measure her qubit at $t = 0$. Alice finds $|0\rangle_A$. What states would Alice and Bob assign at $t = 1$?

- Bob knows that Alice measured her qubit but does not know what result she got. He assigns the states

  $$\rho_{AB} = \tfrac{1}{2}\left(|01\rangle\langle 01| + |10\rangle\langle 10|\right), \quad \rho_A = \rho_B = \tfrac{1}{2}\mathbb{1}$$

  Alice knows more than Bob and assigns

  $$\psi_{AB} = |01\rangle, \quad \rho_A = |0\rangle_A\langle 0|, \quad \rho_B = |1\rangle_A\langle 1|$$

  There is no problem with Alice and Bob assigning different states, since states carry only statistical information and both assignments would agree if they repeat the experiment many times.

- Alice tells Bob what she found. Both agree that the state is

  $$\psi_{AB} = |01\rangle, \quad \rho_A = |0\rangle_A\langle 0|, \quad \rho_B = |1\rangle_A\langle 1|$$

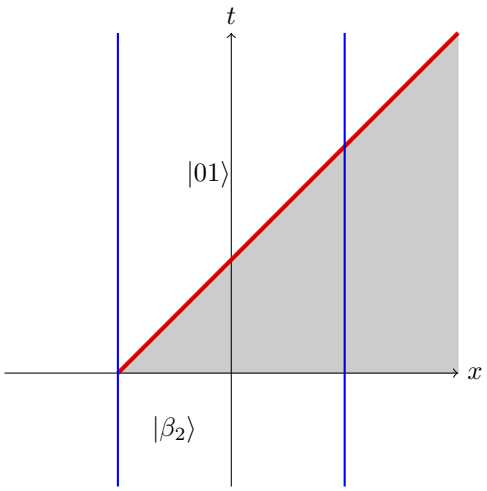


Figure 7.7: The blue line on the left represents Alice qubit. The blue line on the right Bob's. In the shaded triangle Alice and Bob assign different wave functions to the joint system.

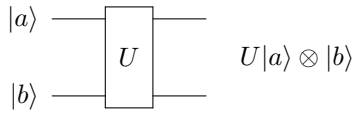
Chapter 8

Two qubits gates

A two qubit gate is a unitary U which acts on the computational basis by

$$U|a\rangle \otimes |b\rangle = \sum_{cd\in 0,1} \underbrace{U_{cd;ab}}_{4\times 4\text{matrix}} |c\rangle \otimes |d\rangle \tag{8.1}$$

Graphically



(8.2)

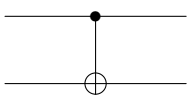
8.1 CNOT

This is the basic two qubit gate. It acts on the second qubit as NOT (i.e. X) conditioned on the first qubit being $|1\rangle$. It is defined by

$$CNOT|a\rangle \otimes |b\rangle = |a\rangle \otimes |a \oplus b\rangle, \quad a, b \in 0, 1 \tag{8.3}$$

Alternate representations of CNOT:

$$\begin{aligned} CNOT &= |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes X \\ &= \begin{pmatrix} \mathbb{1} & 0 \\ 0 & X \end{pmatrix} \\ &= \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 0 & 1 \\ & & 1 & 0 \end{pmatrix} \end{aligned}$$

$=$ 

Evidently:

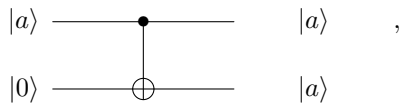
$$(CNOT)^2 = \mathbb{1}, \quad CNOT^* = CNOT$$

Building a practical and reliable CNOT gate is a challenge of quantum engineering. Building a CNOT with ion traps was one of the major achievements of D. Wineland, who got the Nobel in 2012. He used a scheme of Cirac and Zoller, who got the Wolf prize in 2013 for it.

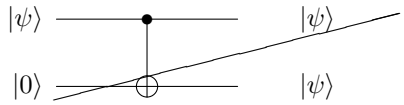
8.2 CNOT: A Classical Xerox machine

CNOT is a (classical) Xerox machine: It copies the states in the first register given in the computational basis, to the empty scratch-pad on the second register

$$CNOT|00\rangle = |00\rangle, \quad CNOT|10\rangle = |11\rangle$$

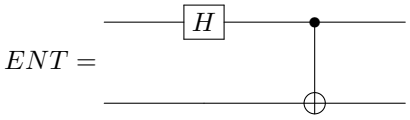


with $a \in \{0, 1\}$. However, as we shall see there is no quantum Xerox machines that copies general qubit states, i.e. superpositions, $|\psi\rangle$:



8.3 Entangling gate

The gate



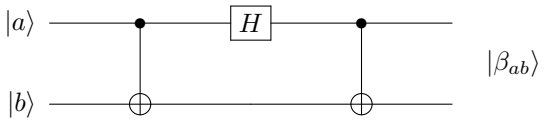
maps the computational basis to (a permutation of) the Bell basis. Indeed

$$\begin{aligned} |ab\rangle &\xrightarrow{H} \frac{1}{\sqrt{2}} \sum_c (-)^{ac} |c\rangle \otimes |b\rangle \\ &\xrightarrow{CNOT} \frac{1}{\sqrt{2}} \sum_c (-)^{ac} |c\rangle \otimes |c \oplus b\rangle \\ &= \frac{|0\rangle \otimes |b\rangle + (-)^a |1\rangle \otimes |b \oplus 1\rangle}{\sqrt{2}} \end{aligned}$$

The right hand side is manifestly in Schmidt form, so is maximally entangled. Since the basis is the computational basis, it is the Bell basis. Explicitly,

$$|00\rangle \mapsto |\beta_0\rangle, \quad |01\rangle \mapsto |\beta_1\rangle, \quad |10\rangle \mapsto |\beta_3\rangle, \quad |11\rangle \mapsto |\beta_2\rangle,$$

The order is spoiled in the last two entries. This is not a big deal and can be fixed easily: Since $CNOT$ interchanges $10 \leftrightarrow 11$ the improved circuit



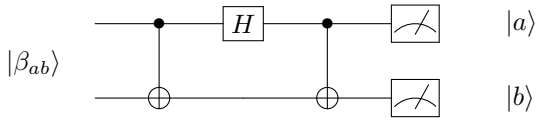
maps the computational basis to the Bell basis, preserving the order: $|ab\rangle \mapsto |\beta_{ab}\rangle$.

Some other ways to write ENT

$$ENT = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{1} & \mathbb{1} \\ X & -X \end{pmatrix}$$
$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{pmatrix}$$

8.4 Bell states detector and post-selection

Suppose someone sells you a black box that he claims emits $|\beta_0\rangle$ on demand. But, he charges you for every state you get from the box. You could verify his claim by tomography, but this will cost you a lot. However, with CNOT gates and Hadamard you can make a single test, using the circuit:



If the claim is right, you should get the output $|00\rangle$ with no error.

8.5 Swap

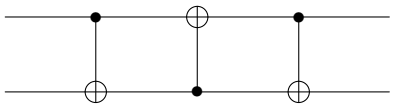
Swap interchanges the states of the two qubits.

$$Swap|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$$

There are two ways to do that: Simply interchange the physical two qubit:

A second way to view swapping is not as a physical interchange, but an interchange of quantum state. This offers greater flexibility because it allows Alice and Bob to have different carriers for the qubit, for example, Alice qubit could be a photon and Bob's could be a nuclear spin.

The swapping circuit is



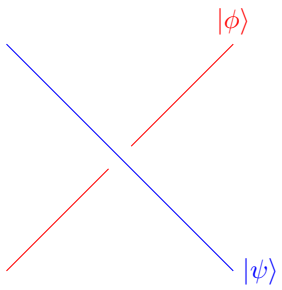


Figure 8.1: Physical interchange of qubits.

This is seen by applying Eq. 8.3

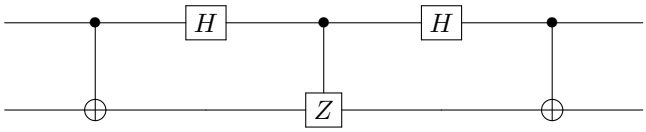
$$\begin{aligned} CNOT_1\ CNOT_2\ CNOT_1\ |a\rangle \otimes |b\rangle &= CNOT_1\ CNOT_2\ |a\rangle \otimes |b+a\rangle \\ &= CNOT_1\ |a+b+a\rangle \otimes |b+a\rangle \\ &= |b+2a\rangle \otimes |2b+3a\rangle \\ &= |b\rangle \otimes |a\rangle \end{aligned}$$

and the fact that for binaries $2a = 0, 3a = a$

As an operator in the computational basis

$$\begin{aligned} Swap &= \sum_{a,b\in 0,1} |a\rangle\langle b| \otimes |b\rangle\langle a| \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

The circuit



swaps the states of the two qubits with no physical interchange if the media. Alice qubit may be a nuclear spin and Bob's a photon. After swap they still posses the same physical object but the quantum state has been swapped.

8.6 C(Z)

$C(Z)$ is

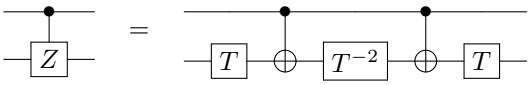
$$C(Z)|ab\rangle = (-)^{ab}|ab\rangle \Longleftrightarrow C(Z) = \frac{\mathbb{1}+Z}{2} \otimes \mathbb{1} + \frac{\mathbb{1}-Z}{2} \otimes Z \qquad (8.4)$$

You might first think that being a gauge transformation this can not be an interesting gate. However, from the fact that $HZ = XH$ we have

$$C(Z) = \big(\mathbb{1} \otimes H\big) C(X) \big(\mathbb{1} \otimes H\big)$$

so if you can construct $C(Z)$ and Hadamard you have got $CNOT$, and vice versa.

Exercise 8.1. *Show that with* $T^4 = Z$



8.6.1 C(Z) and controlled quantum evolutions

It follows from Eq. 8.4 that

$$2\,C(Z) - \mathbb{1} \otimes \mathbb{1} = \underbrace{\mathbb{1} \otimes Z + Z \otimes \mathbb{1}}_{\text{Zeeman term}} - \underbrace{Z \otimes Z}_{\text{spin-spin interaction}}$$

The terms on the right can be interpreted as interactions with physical interpretation: The first two terms on the right describe interaction of the a spin 1/2 with an external magnetic field. The third term describes mutual spin-spin interaction. To physically realize such a Hamiltonian you need to tune the magnetic field so that its strength matchers the spin-spin interaction.

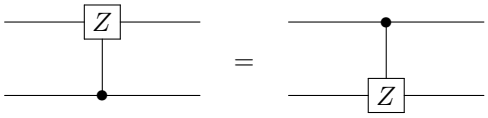
Using the fact that $(C(Z))^2 = \mathbb{1}$

$$e^{-it\mathbf{H}} = e^{it}(\mathbb{1}\cos 2t - iC(Z)\sin 2t)$$

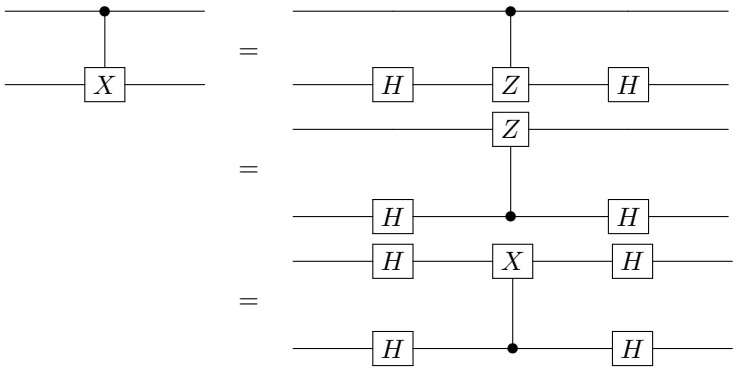
If you can control the the system so that the Hamiltonian is only effective for $0 \le t \le \pi/4$ the Hamiltonian evolution gives a $C(Z)$ gate up to an overall phase. C(Z) can and has been implemented in NMR and in Ion traps. .

8.7 Conditioning

In quantum gates the conditioning bit depends can sometimes be shifted. For example, from the definition of $C(Z)$ Eq. 8.4, it is obvious that:



The control point, has been shifted around. Now, using $HX = ZH$ we immediately get for CNOT



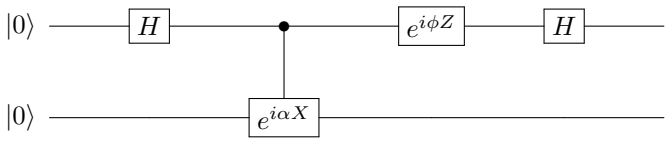
Thus, we have interchanged the conditioning bit from the first to the second at the price of making unitary transformations H, in this case, on the input and on the output.

8.8 "Which path detector"

In a Mach-Zehnder interferometer you measure the top qubit (as usual, in the computational basis), and want to use this information to learn about the control phase ϕ. Maximal visibility means that by changing ϕ the probability $P(0)$ can be changed from 0 to 1. If, in contrast, $P(\phi)$ is independent of ϕ the interference is lost.

It is a basic principle that if you try to monitor which path the quantum particle took you destroy the interference. The detection reduces a superposition to a mixture. Lets see how this works out in a quantum circuit.

We append to the Mach-Zehnder interferometer a second qubit that operates like "which path" detector. This is represented by the circuit



The first Hadamard gate creates a superposition in the computational basis. ϕ is the "difference in optical lengths" between the two computational basis states. The control gate is the "which path" detector. α measures how effective is the second qubit in monitoring the path: When $\alpha = 0$ the guard is asleep. When $\alpha = \pi/2$ the guard is fully alert and raises a flag if the first qubit uses the path $|1\rangle$. The control gate in this case is CNOT.

Disregarding normalization:

$$|0\rangle \otimes |0\rangle \underbrace{\mapsto}_{H} |+\rangle \otimes |0\rangle \underbrace{\mapsto}_{Control} |0\rangle \otimes |0\rangle + |1\rangle \otimes e^{i\alpha X}|0\rangle$$
$$\underbrace{\mapsto}_{\text{Phase gate}} e^{i\phi}|0\rangle \otimes |0\rangle + e^{-i\phi}|1\rangle \otimes e^{i\alpha X}|0\rangle$$
$$\underbrace{\mapsto}_{H} e^{i\phi}|+\rangle \otimes |0\rangle + e^{-i\phi}|-\rangle \otimes e^{i\alpha X}|0\rangle$$

When $\alpha = 0$, the guard is asleep. The (correctly normalized) output is a pure product state:

$$\frac{1}{\sqrt{2}}\Big(e^{i\phi}|+\rangle + e^{-i\phi}|-\rangle\Big) \otimes |0\rangle$$

The probability amplitude of finding the first qubit in $|0\rangle$ is $\cos\phi$. It is sensitive to ϕ, which means that we can see interference.

When $\alpha = \pi/2$ the guard is alert, it reacts to the state of the top qubit. The output state is

$$\frac{1}{\sqrt{2}}\Big(e^{i\phi}|+\rangle \otimes |0\rangle + ie^{-i\phi}|-\rangle \otimes |1\rangle\Big)$$

The probability to find the first qubit in the state $|0\rangle$ is the norm of the vector obtained by the projection of the first qubit to $|0\rangle$:

$$\frac{1}{2}\Big(e^{i\phi}|0\rangle \otimes |0\rangle + ie^{-i\phi}|0\rangle \otimes |1\rangle\Big) = |0\rangle \otimes \frac{e^{i\phi}|0\rangle + ie^{-i\phi}|1\rangle}{2}$$

Since the vector is a superposition of two orthogonal vectors its norm, by Pythagoras is $1/2$ independent of ϕ. This means that if you want to get information on the path (measured by second qubit) you loose the interference pattern in the first. The interference has been erased.

Exercise 8.2. *Derive the same result by computing instead ρ_A obtained by tracing over the second qubit.*

8.9 $C(U)$ from single qubit unitaries and CNOT

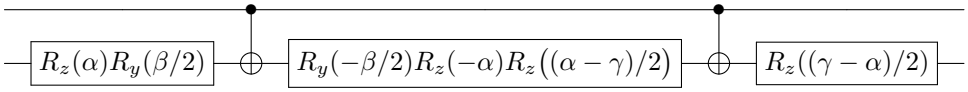
Any $C(U)$ gate can be built from CNOT and single qubit unitaries. Recall first that any U can be written as three Euler rotations

$$U = R_z(\alpha)\, R_y(\beta)\, R_z(\gamma)$$

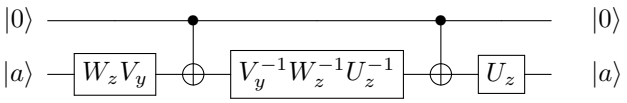
Define

$$W_z = R_z(\alpha), \quad V_y = R_y(\beta/2), \quad U_z = R_z((\gamma - \alpha)/2)$$

We claim that the circuit



To see this note first that if the control bit is 0 the circuit is the the identity. This follows from



where

$$W_z = R_z(\alpha), \quad V_y = R_y(\beta/2), \quad U_z = R_z((\gamma-\alpha)/2) \tag{8.5}$$

To see that it applies U on the second qubit when the control is 1 write

$$\begin{aligned} U &= W_zV_y\,X\,\big(V_y^{-1}W_z^{-1}U_z^{-1}\big)X\,U_z \\ &= W_zV_y\,\big(X\,V_y^{-1}X\big)\big(XW_z^{-1}X\big)\big(\big(XU_z^{-1}X\big)\,U_z \end{aligned} \tag{8.6}$$

Now X is a rotation by π around the x axis: flips the y and z axis. This means

$$U_z = XU_z^{-1}\,X, \quad V_y = XV_y^{-1}\,X$$

Hence

$$\begin{aligned} U &= W_zV_y\,\big(X\,V_y^{-1}X\big)\big(XW_z^{-1}X\big)\big(\big(XU_z^{-1}X\big)\,U_z \\ &= W_zV_y^2W_zU_z^2 \\ &= R_z(\alpha)R_y(\beta)R_z(\gamma) \end{aligned} \tag{8.7}$$

We have thus shown that CNOT, H and T are universal for 2 qubits. It turns out that they are universal for n-qubits too, but I will not show that.

Chapter 9

Hilbert space is big

Three things make quantum mechanics different from classical physics

- The theory is fundamentally probabilistic: Complete knowledge does not allow to predict everything.
- Superpositions: Cats can be both dead and alive.
- The Hilbert space is big: It grows exponentially with the number of qubits n.

In classical waves we have superpositions without probability and with no growth of the configuration space (it stays 3 dimensional); In classical statistical physics we have probability and exponentially many configurations but no superpositions and with n classical bits we can count to 2^n but have no superpositions and no probability.

9.1 n bits

With n bits you can count from $0, 1, \ldots, N-1$. Here and throughout

$$N = 2^n \tag{9.1}$$

Different representations of n bits:

- Arithmetically: The integers $\{0, \ldots, N-1\}$: In binary notation

$$a_n a_{n-1} \ldots a_1$$

- Geometrically: The vertexes of the unit cube in n dimensions:

$$(a_0, \ldots, a_n),\ a_j \in \{0, 1\}$$

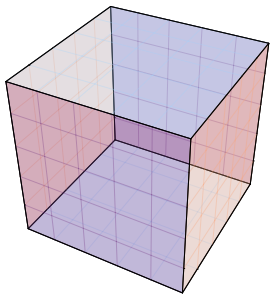


Figure 9.1: The unit cube in n dimensions has $N = 2^n$ vertexes corresponding to the N numbers $(0, 0, \ldots, 0)$ to $(1, 1, \ldots 1)$.

- Combinatorially: The number of different subsets of the set of n elements. For example with $n = 2$

$$\{\}, \quad \{0\}, \quad \{1\}, \quad \{01\}$$

Remark 9.1. *The space of functions is monstrously large: There are 4 functions of a single bit*

$$\underbrace{f_0(a) = 0, \quad f_1(a) = 1}_{even}, \quad \underbrace{f_3(a) = a, \quad f_4(a) = 1 \oplus a}_{odd}$$

where $\oplus$ is bit addition mod 2. In general, there are 2^N different functions of n arguments, each taking values in $\{0, 1\}$. For $n = 8$ the number is comparable to the number of atoms in the visible universe.

Remark 9.2. *The unit cube in n dimensions is increasingly anisotropic when n is large: The diagonal has length $\sqrt{n}$.*

Exercise 9.3. *Show that the number of d dimensional (cubic) faces of the unit cube in n dimensions built on the vertex $(0, 0, \ldots, 0)$ is $\binom{n}{d}$. (Hint: Check this for the square and the cube and generalize.)*

Exercise 9.4. *Show that the total number of $d \leq n$ dimensional cubes in the unit cube in n dimensions is $2^{n-d}\binom{n}{d}$.*

Exercise 9.5. *The Euler characteristics is the alternating sum $\chi = Vertices - Edges + Faces - \ldots$. Show that $\chi(cube) = 1$.*

9.2 Classical systems can be efficiently simulated

The state of n classical particles is uniquely specified by a point in phase space: $\mathbb{R}^{6n}$ i.e. by the $6n$ (generalized) coordinates

$$(q_1, \ldots, q_{6n})$$

The number of coordinates grows linearly in n.

# Coordinates	$O(n)$
# Differential equations	$O(n)$
# Terms in each equation	$O(n^2)$
# Total number of terms	$O(n^3)$

The evolution of a classical system of n particles is governed by $6n$ ordinary differential equations. The common physical forces are pair interactions and hence the number of terms in each equation is $O(n^2)$.

This means that the complexity of simulating a classical system grows polynomially with the number of particles, e.g. $O(n^3)$. This is why classical systems can be efficiently simulated on classical computers.

Example 9.6. *With a 10 Giga bytes of RAM memory you can simulated* 10^4 *particles, assuming that the complexity scales like* $O(n^3)$ *and the all constants (e.g. the number of bits you use to approximate real numbers) are* $O(1)$

9.3 Hilbert space blows up exponentially with n

The state of n qubits is fully specified by a vector in Hilbert space whose dimension is

$$\dim \mathcal{H} = N = 2^n$$

The dimension of the Hilbert pace grows exponentially faster than the dimension of classical phase space.

Pure states of n qubits are represented by $N = 2^n$ complex amplitudes:

$$|\Psi\rangle = \sum_{a_j \in 0,1} \underbrace{\psi_{a_1 \dots a_n}}_{\text{amplitude}} \underbrace{|a_1, \dots, a_n\rangle}_{\text{computational basis}}$$

where the N amplitudes can be written in two alternative notations:

$$\psi_{00\dots00} = \psi_0, \quad \psi_{00\dots01} = \psi_1, \quad \psi_{11\dots11} = \psi_{N-1}$$

The state $|\Psi\rangle$ evolves by Schrödinger equation

$$i\frac{d|\Psi\rangle}{dt} = H|\Psi\rangle$$

where H is an $N \times N$ complex matrix.

# Coordinates	$O(N)$
# Differential equation	1
# Terms in equation	$O(N^2)$

This implies that a quantum system of n qubits *can not* be simulated efficiently on a classical computer. The number of bits you need to simulate n qubits grows exponentially with the input.

This was Feynman's original motivation for proposing to build quantum computers: Unlike bits, qubits simulate qubits efficiently.

Example 9.7. *The number of PC world-wide is estimated to be* 10^9*, if each PC has* 10 *Gigabytes of RAM, the total RAM world wide is* $10^{19} \approx 2^{60}$ *Bytes. This will allow you to represent a single wave function of not quite 60 qubits. (Assuming all constants, including the number of bits used to approximat a real number, to be* $O(1)$*.) If the qubits are spin in a 3 dimensional lattice, it is the state of not quite* $4 \times 4 \times 4$ *spins. Not a very large lattice.*

9.4 Geometry of pure states

The space of distinct pure state is the complex projective space $CP(N-1)$

$$\mathbb{C}^N/\mathbb{C}$$

where we identify two vectors in the Hilbert space that differ by normalization and overall phase. $CP(N-1)$ is a compact set whose dimension is

$$\dim CP(N-1) = 2(N-1)$$

# qubits	$\dim CP(N-1)$
1	2
2	6
3	14

9.5 Geometry of states

Any mixed state can be written as a convex combination of pure states[1]:

$$\rho = \sum p_j |\psi_j\rangle\langle\psi_j|, \quad p_j \geq 0, \quad \sum p_j = 1$$

Since $|\psi_j\rangle\langle\psi_j|$ is a projection, it is a positive matrix. And since p_j are probabilities they are all positive numbers. Hence $p_j|\psi_j\rangle\langle\psi_j|$ is a positive matrix. A sum of positive matrices is a positive matrix so we have shown that

$$\rho \geq 0$$

[1] One may take this as a definition of pure states.

Moreover, by linearity

$$\begin{aligned} Tr\,\rho &= Tr\Big(\sum p_j|\psi_j\rangle\langle\psi_j|\Big) \\ &= \sum p_j Tr\left(|\psi_j\rangle\langle\psi_j|\right) \\ &= \sum p_j\langle\psi_j|\psi_j\rangle \\ &= \sum p_j \\ &= 1 \end{aligned}$$

The density matrix is therefor an $N \times N$ positive matrix with unit trace. Geometrically, the space of density matrices is a convex body whose dimension is

$$N^2 - 1$$

The pure states as its extreme points, i.e. they all lie on the boundary.

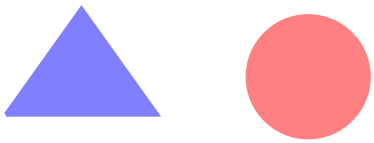


Figure 9.2: Convex sets and their extreme points. For the disk all the boundary, the circle, is made of extreme points. For the triangle the extreme points are the three vertexes. The space of states is intermediate: The extreme points are a set of relatively low dimensions. Only for a single qubit all the boundary corresponds to pure states.

# qubits	$\dim CP(N-1)$	$\dim \rho$
1	2	3
2	6	15
3	14	63

The extreme points of the set of states are the pure states. (Extreme points are the points that can not be represented as a weighted sum of other points.) Since the pure states are a $2(N-1)$ dimensional set, when n is large, they are an exponentially small part of the boundary of the set of states which is N^2-2 dimensional.

9.6 The Pauli basis

Let $\sigma_\mu, \quad \mu \in \{0,1,2,3\}$ be the Pauli matrices ($\sigma_0 = \mathbb{1}$). Write

$$\boldsymbol{\sigma}_\alpha = \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n}, \quad \alpha_j \in \{0,\ldots,3\} \tag{9.2}$$

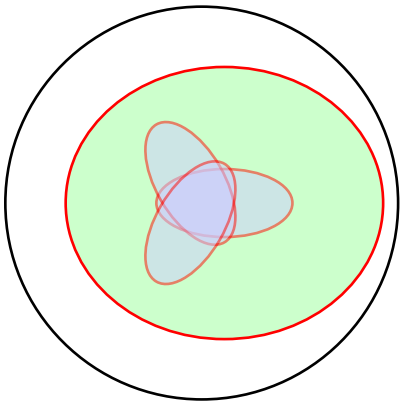


Figure 9.3: A schematic diagram of the space of states of three qubits. The black circle is the bounding sphere of unit radius where pure states lie. The greenish ellipse represents the space of states D_3. Since the pure states are a relatively tiny set a typical cross section will not reach the unit circle. The three bluish ellipses in represent the three bipartite separable states. Their intersection is the fully separable set of states.

for the n tensor product. The N matrices $\boldsymbol{\sigma}_\alpha$ provide a basis for the $N \times N$ Hermitian matrices over the reals.

$\boldsymbol{\sigma}_\alpha$ are elements of a group (Pauli group), are all of order 2 and either commute or anti-commute:

$$\boldsymbol{\sigma}_\alpha^2 = \mathbb{1}, \quad \boldsymbol{\sigma}_\alpha \boldsymbol{\sigma}_\beta = \pm \boldsymbol{\sigma}_\beta \boldsymbol{\sigma}_\alpha, \tag{9.3}$$

The basis is orthogonal with scalar product

$$Tr\left(\boldsymbol{\sigma}_\alpha \boldsymbol{\sigma}_\beta\right) = \delta_{\alpha\beta} N \Longrightarrow Tr\left(\boldsymbol{\sigma}_\alpha\right) = \delta_{\alpha,0} N, \tag{9.4}$$

9.7 Geometry of states of n qubits

We shall denote by D_n the space of trace-normalized states, represented in $\mathbb{R}^{N^2-1}$:

$$D_n = \left\{ \mathbf{r} \; | \mathbf{r} \in \mathbb{R}^{N^2-1}, \quad \rho \geq 0 \right\} \tag{9.5}$$

where $\mathbf{r}$ and ρ are related by

$$\rho = \frac{\boldsymbol{\sigma}_0}{N} + \frac{\sqrt{N-1}}{N} \sum_{\alpha=1}^{N^2-1} r_\alpha \boldsymbol{\sigma}_\alpha, \tag{9.6}$$

For $n = 1$ this is the Bloch ball. Hence D_n is the higher dimensional analog of the Bloch ball.

The Hilbert space distance is proportional to the Euclidean distance in $\mathbb{R}^{N^2-1}$:

$$\begin{aligned} N \ \mathrm{dist}^2(\rho,\rho') &= N\,Tr\,(\rho-\rho')^2 \\ &= (N-1)(\mathbf{r}-\mathbf{r}')^2 \\ &= (N-1)\ \mathrm{dist}^2(\mathbf{r},\mathbf{r}') \end{aligned} \tag{9.7}$$

Similarly, the scalar products are related by

$$N\,Tr\,(\rho\rho') = 1 + (N-1)(\mathbf{r}\cdot\mathbf{r}') \tag{9.8}$$

9.8 Qualitative features of D_n

The Bloch ball is not always a good guide to the geometry of the space of states D_n in high dimensions. D_n is not a ball, in general. It is not even reflection symmetric, and it is not true that the pure states are the boundary of D_n.

Here is what is true:

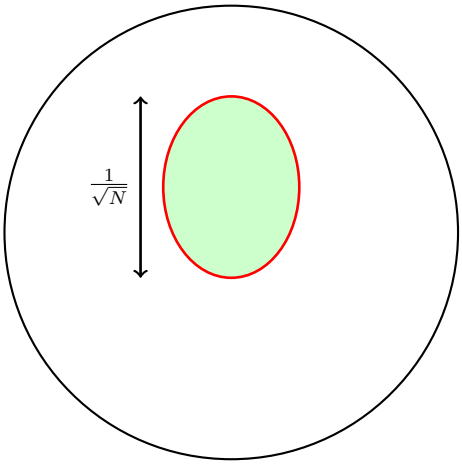


Figure 9.4: The figure shows a schematic 2-D section of the space of states in $\mathbb{R}^{N^2-1}$. The section goes through the origin (the fully mixed state). The black circle the intersection with the unit sphere. This is where all the pure states lie. Since the pure states are a set of relatively low dimension a generic 2-D section will miss all pure states. The intersection with D_n is the green ellipse. It will be close to spherical and small; The diameter is $O(1/\sqrt{N})$.

1. D_n is a convex body.
2. $\dim D_n = N^2 - 1$.
3. $0 \in D_n$ since the origin is a state: the fully mixed state.

4. {Pure states} $\subseteq S_1$ where S_1 is the unit sphere in $\mathbb{R}^{N^2-1}$ centered at the origin.

   To see this recall that pure state satisfy $\rho^2 = \rho$. Since $Tr\,\rho^2 = Tr\,\rho = 1$ by Eq. 9.8 the pure states satisfy

   $$N = 1 + (N-1)(\mathbf{r}\cdot\mathbf{r})$$

   which is the unit sphere.

5. $D_n \subseteq B_1$ where B_1 is the unit ball centered at the origin.

   This is because the extreme points, the pure states, all lie on the unit sphere.

6. The diameter of D_n, i.e. the largest distance between two points that belong to D_n, is approximately $\sqrt{2}$ for n large, and more precisely

   $$\text{diameter}\,(D_n) = \sqrt{\frac{2N}{N-1}}$$

   This follows from Eq. (9.7) and

   $$Tr(\rho-\rho')^2 = Tr\rho^2 + Tr\rho'^2 - 2Tr\rho\rho' \leq 2(1 - Tr\rho\rho') \leq 2$$

   The two inequalities are saturated for two pure orthogonal states.

7. The set of pure states is not symmetric under inversion $\mathbf{r} \mapsto -\mathbf{r}$ for $n \geq 2$.

   This follows from the fact that diameter$(D_n) < 2$ for $n \geq 2$ and hence antipodal points on the unit sphere can not both be states.

8. D_n *is not* inversion symmetric unless $n = 1$.

   This follows from the previous assertion.

9. D_n is reflection symmetric under flipping $\sigma_y \mapsto -\sigma_y$.

   This follows from the fact that if ρ is a state so is ρ^t.

10. The angle between two orthogonal pure states is

    $$\cos\theta = -\frac{1}{N-1}$$

    This follows from $Tr\,\rho\rho' = 0$ for orthogonal pure states, and Eq. 9.7.

    For a qubit, where $N = 2$, the angle is π and orthogonal states are antipodal, when $N \to \infty$ the angle is $\pi/2$.

11. Although the shape of D_n is complicated, one can compute the average raduis:

    $$\langle \mathbf{r}^2 \rangle = \frac{\int_{\mathcal{D}} \mathbf{r}^2 d\mathbf{r}}{\int_{\mathcal{D}} d\mathbf{r}} = \frac{N+1}{N^2+1} \tag{9.9}$$

    I do not show this computation.

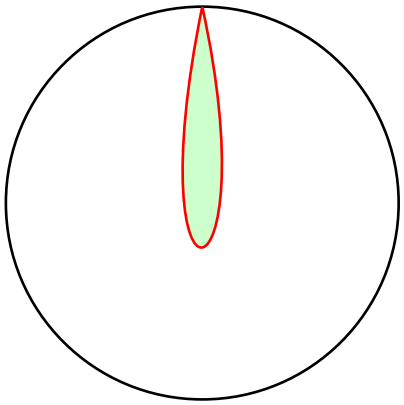


Figure 9.5: A typical cross section will miss all the pure states and so will not touch the unit sphere. The figure shows a cross section that is constrained to hit a pure state. It looks a bit like a needle.

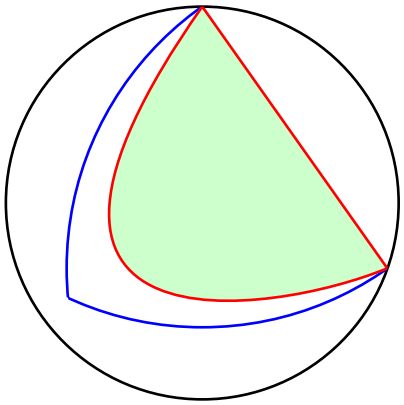


Figure 9.6: The figure shows a schematic section of D_n, $n \geq 2$, shown in light green, which goes through two generic orthogonal pure states. The black circle is the sphere of radius 1 which is the locus of pure states. The straight (red) line connecting the pure states lies on the boundary of $\mathcal{D}_n$ since the mixture of two pure states does not have a full rank if $n \geq 2$. The two (blue) arcs delineate the region that is at distance $\sqrt{2}$ from the two pure states.

9.9 2-D Cross sections

A two dimensional cross section of D_n along the Pauli coordinate planes corresponds to the matrices $\rho(x, y)$:

$$N\rho(x, y) = \boldsymbol{\sigma}_0 + \sqrt{N-1}(x\boldsymbol{\sigma}_\alpha + y\boldsymbol{\sigma}_\beta), \quad \alpha \neq \beta \tag{9.10}$$

By Eq. (9.3) $\boldsymbol{\sigma}_{\alpha,\beta}$ either commute or anti-commute.

- $\{\boldsymbol{\sigma}_\alpha, \boldsymbol{\sigma}_\beta\} = 0$: Since

$$(x\boldsymbol{\sigma}_\alpha + y\boldsymbol{\sigma}_\beta)^2 = (x^2 + y^2)\mathbb{1}, \quad Tr(x\boldsymbol{\sigma}_\alpha + y\boldsymbol{\sigma}_\beta) = 0$$

It follows that

$$\text{Spectrum } (x\boldsymbol{\sigma}_\alpha + y\boldsymbol{\sigma}_\beta) = \pm\sqrt{x^2 + y^2}$$

and $\rho(x, y) \geq 0$ provided

$$x^2 + y^2 \leq \frac{1}{N-1} \tag{9.11}$$

The cross section is a disk which is exponentially small as n is large.

- $[\boldsymbol{\sigma}_\alpha, \boldsymbol{\sigma}_\beta] = 0$ (This requires $n \geq 2$):
Since the matrices commute they have common eigenvectors and

$$\text{Spectrum}(x\boldsymbol{\sigma}_\alpha + y\boldsymbol{\sigma}_\beta) = \{\pm x \pm y, \pm x \mp y\}$$

It follows that $\rho \geq 0$ if

$$\pm x \pm y < \frac{1}{\sqrt{N-1}}, \quad \pm x \mp y < \frac{1}{\sqrt{N-1}} \tag{9.12}$$

The cross section is a square which is exponentially small when n is large .

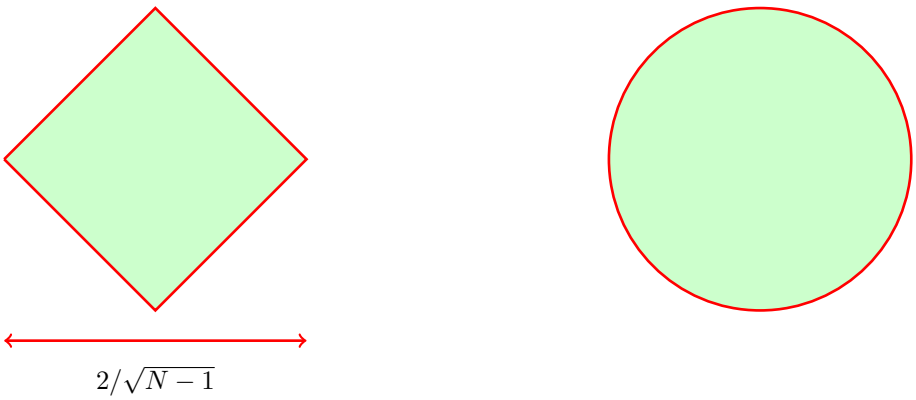


Figure 9.7: Any two dimensional cross sections of the space of states of n qubits, D_n, along the plane of Pauli coordinates, $\boldsymbol{\sigma}_\alpha - \boldsymbol{\sigma}_\beta$ is either a tiny square of a tiny disk both of size $1/\sqrt{N-1}$.

9.10 Clifford algebras

Clifford matrices γ_j are defined by

$$\{\gamma_j, \gamma_k\} = 2\delta_{jk}$$

The Pauli matrices σ_j are a two dimensional example. Dirac γ_μ matrices are a four dimensional example.

One can construct high dimensional Clifford matrices by the following iterative procedure: Starting with the Pauli matrices construct the tensor product

$$\gamma_j = \boldsymbol{\sigma}_{1j} = \sigma_1 \otimes \sigma_j, \quad j \in \{1,2,3\}, \qquad \gamma_4 = \boldsymbol{\sigma}_{30} = \sigma_3 \otimes \mathbb{1}$$

One readily check that γ_j together with $\gamma_5 = \gamma_1 \ldots \gamma_4$ are all Clifford.

The Clifford matrices define an algebra: You can add and multiply linear combinations of them.

Among $N^2 - 1$ matrices $\boldsymbol{\sigma}_\alpha$ there are m Clifford matrices γ_j where

$$m = \begin{cases} 2n & n \text{ even} \\ 2n+1 & n \text{ odd} \end{cases}$$

9.11 Clifford cross sections are balls

Consider an m dimensional cross section of D_n of the form

$$N\rho(\mathbf{r}) = \boldsymbol{\sigma}_0 + \sqrt{N-1}\sum_\alpha r_\alpha \boldsymbol{\sigma}_\alpha \tag{9.13}$$

where $\boldsymbol{\sigma}_\alpha$ are Clifford matrices. Note that when n is large, this is a relatively "low dimensional" cross section of a much higher dimensional body D_n.

For Clifford matrices

$$\Big(\sum r_\alpha \boldsymbol{\sigma}_\alpha\Big)^2 = \mathbf{r}^2$$

It follows that the density matrix ρ is positive if

$$0 \le N\rho = \mathbb{1} + \sqrt{N-1}\sum_{\alpha=1}^{m} r_\alpha \boldsymbol{\sigma}_\alpha, \quad r_\alpha \in \mathbb{R} \tag{9.14}$$

This is the case provided

$$\mathbf{r}^2 \le \frac{1}{N-1}$$

Note the agreement with the average radius Eq. 9.9 to leading order in N.

This result is closely related to a what mathematicians call measure concentration and the Dvoretzy-Milman theorem.

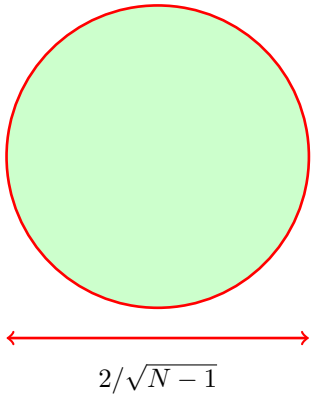


Figure 9.8: An $O(n)$ dimensional cross section of D_n along the direction of the Clifford matrices in $\boldsymbol{\sigma}_\alpha$ is a tiny ball of radius $1/\sqrt{N}$. This is, in a sense, the generic behavior.

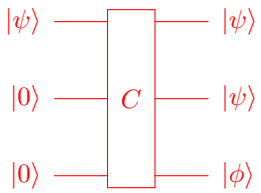
Chapter 10

Quantum tricks: II

10.1 No cloning

An unknown quantum state $|\psi\rangle$ can not be cloned (copied). In fact, one can not even clone three states

$$|0\rangle, \quad |1\rangle, \quad |\psi\rangle = \cos\psi|0\rangle + \sin\psi|1\rangle, \quad \psi \neq 0, \pm\pi/2, \pi$$



Proof: Assume you can then

$$C \underbrace{|\psi\rangle}_{original} \otimes \underbrace{|0\rangle}_{blank\ register} \otimes \underbrace{|0\rangle}_{internal\ state} = \underbrace{|\psi\rangle \otimes |\psi\rangle}_{two\ copies} \otimes \underbrace{|\phi_\psi\rangle}_{fax\ final\ state}$$

We allow the final state of the copier to depend on the state to be copied.

By linearity:

$$\begin{aligned} \cos\psi C|000\rangle + \sin\psi C|100\rangle &= \underbrace{\cos\psi|00\phi_0\rangle + \sin\psi|11\phi_1\rangle}_{by\ linearilty} \\ &= \underbrace{|\psi,\psi,\phi_\psi\rangle}_{by\ ansatz} \\ &= \cos\psi|0\psi\phi_\psi\rangle + \sin\psi|1\psi\phi_\psi\rangle \end{aligned}$$

Compare the rhs of the first line and the third line by projection on the computational basis of the first qubit:

$$\cancel{\cos\psi}(|0\phi_0\rangle - |\psi\phi_\psi\rangle) = 0, \quad \cancel{\sin\psi}(|1\phi_1\rangle - |\psi\phi_\psi\rangle) = 0$$

since, by assumption $\cos\psi, \sin\psi \neq 0$. Now, comparing the next factor gives a contradiction:

$$|0\rangle = |\psi\rangle = |1\rangle$$

We conclude that we can only copy the computational basis, and can make only classical fax machine (Section 8.2).

Exercise 10.1. *If the state $|\psi\rangle$ is known, it can be cloned. Explain.*

10.2 Cloning allows for superluminal signaling

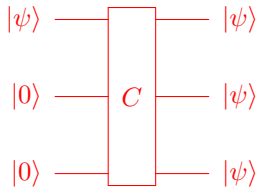
Another way to see that cloning should be impossible is to observe that **if one could clone** states in two non-orthogonal basis, say the computational Z basis and in the X basis, then one could signal instantaneously. Quantum mechanics would then be in conflict with the principle that the propagation speed of signals is finite.

Protocol (assuming cloning):

- Alice and Bob share a singlet.
- Alice measures her qubit in the X basis if she wants to transmit 0 and in the Z basis otherwise.
- Alice and Bob agree that Alice will measure at 08:00
- Bob knows that his qubit is in one of the 4 possible states at 08:01

$$|0\rangle, \quad |1\rangle, \quad |+\rangle, \quad |-\rangle$$

- Bob clones his state



- Bob makes a tomography of the ensemble $|\psi\rangle$ on the right. If he finds $|\pm\rangle$ he got the bit 0 and if he finds $|0\rangle$ or $|1\rangle$ he got the bit 1.

This requires Bob to clone in two non-orthogonal bases X and Z which fortunately is not possible.

Remark 10.2. *Note that in order for Bob to do tomography it is important that his qubits are in a product state of identical states:*

$$\underbrace{|\psi\rangle \otimes |\psi\rangle \otimes \cdots \otimes |\psi\rangle}_{3n \text{ copies}} = \underbrace{|\psi\rangle \otimes \cdots \otimes |\psi\rangle}_{n \text{ copies}} \otimes \underbrace{|\psi\rangle \otimes \cdots \otimes |\psi\rangle}_{n \text{ copies}} \otimes \underbrace{|\psi\rangle \otimes \cdots \otimes |\psi\rangle}_{n \text{ copies}}$$

Bob can then measure the expectation of X for the first n qubit, Y for the next n qubits and Z for the remaining n to get a tomography of $|\psi\rangle$.

10.3 Function gates

A function

$$f : x \in \mathbb{Z}_2^n \mapsto f(x) \in \mathbb{Z}_2$$

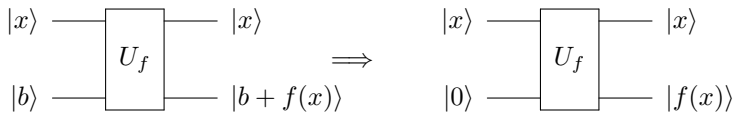
is, in general, not invertible. How can we represent such functions by unitary gates, which are, of course, invertible. We do that with redundancy: add an extra memory bit to the output:

$$U_f|x\rangle \otimes |b\rangle = |x\rangle \otimes |b \oplus f(x)\rangle$$

To evaluate the function on the input a we feed in blank $b = 0$:

$$U_f|x\rangle \otimes |0\rangle = |x\rangle \otimes |f(x)\rangle$$

As a circuit:



U_f is its own inverse:

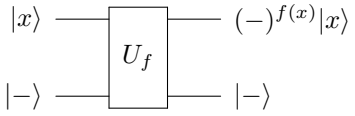
$$U_f U_f = \mathbb{1}$$

10.4 Phase kickback

In the computational basis a function gate appears to act on the bottom qubit and leave the top register idle. This is actually a basis dependent property–If the second qubit is in the $|-\rangle$ basis and the first register is in the computational basis then the result is an overall phase which we can assign as we please:

$$U_f|x\rangle \otimes |-\rangle = |x\rangle \otimes \frac{|f(x)\rangle - |1 \oplus f(x)\rangle}{\sqrt{2}} = (-)^{f(x)}|x\rangle \otimes |-\rangle$$

It is conventional to append the sign to the first register



This is phase kickback. We shall presently see how we can make good use of this.

10.5 Deutsch algorithm

The Deutsch algorithm (1992) is perhaps the simplest algorithm that shows that a quantum machine can be more powerful than the corresponding classical machine.

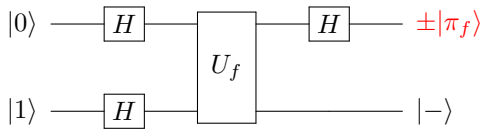
We say that $f : \{0,1\} \to \{0,1\}$ is even if $f(0) = f(1)$ and odd otherwise. We denote this by π_f, the parity of f:

$$\pi_f = f(0) \oplus f(1) = \begin{cases} 1 & f \text{ odd} \\ 0 & f \text{ even} \end{cases}$$

The Deutsch task:

- You are given the function gate U_f
- You do not know what f is.
- You are requested to determine the parity π_f with minimal queries of U_f

Classically, to determine π_f you need to compute f twice. But with a quantum gate one query suffices. The claim is that



(10.1)

The top qubit gives the parity of the function f with one query.

Here is why:

$$\begin{aligned} |0\rangle \otimes |1\rangle &\xrightarrow{H\otimes H} |+\rangle \otimes |-\rangle = \left(\frac{1}{\sqrt{2}} \sum_{a\in 0,1} |a\rangle\right) \otimes |-\rangle \\ &\xrightarrow{U} \left(\frac{1}{\sqrt{2}} \sum_{a\in 0,1} (-)^{f(a)} |a\rangle\right) \otimes |-\rangle \\ &\xrightarrow{H\otimes \mathbb{1}} \frac{1}{2} \left(\sum_{a,b\in 0,1} (-)^{f(a)+ab} |b\rangle\right) \otimes |-\rangle \end{aligned}$$

We made use of phase kickback in the second line. The brackets in the last line is

$$\frac{1}{2} \sum_{a,b\in 0,1} (-)^{f(a)+ab} |b\rangle = \underbrace{\left(\sum_{a\in 0,1} (-)^{f(a)}\right)}_{\pm\pi_f \oplus 1} |0\rangle + \underbrace{\left(\sum_{a\in 0,1} (-)^{f(a)+a}\right)}_{\pm\pi_f} |1\rangle \tag{10.2}$$

Note that if $f(a)$ is odd then $f(a) \oplus 1$ is even and vice versa. Hence

$$\underbrace{\sum_{a\in 0,1} (-)^{f(a)} = 0}_{\text{f odd}}, \quad \underbrace{\sum_{a\in 0,1} (-)^{f(a)+a} = 0}_{\text{f even}}$$

So, finally,

$$\frac{1}{2}\sum_{a,b\in 0,1}(-)^{f(a)+ab}|b\rangle = \pm|\pi_f\rangle \tag{10.3}$$

A measurement of the top qubit in the computational basis gives the parity of f: If $a = 0$ then f is even and if $a = 1$ then f is odd.

This algorithm makes use of QM in several ways:

- Quantum parallelism: The first Hadamard gate allows to evaluate f for the $a = 0, 1$ simultaneously.
- Interference: The second Hadamard makes constructive interference to produce the correct answer with enhanced probability.
- Collapse: Measurement in the computational basis reveals the parity with probability 1.

Remark 10.3. . *There is an extension of this trick known as Deutsch-Josza algorithm where the gain is not 2 but* 2^n. *Consider a function*

$$f : \mathbb{Z}_2^n \to \mathbb{Z}_2$$

f *is promised to be either balanced or constant, where*

$$\sum_0^{N-1} f(j) = \begin{cases} \pm N & constant \\ 0 & balanced \end{cases}, \quad N = 2^n$$

To determine if f *is balanced of constant you need to evaluate it at least* $(N/2)+1$ *times. With a quantum circuit you can evaluate it once. You gain exponentially in the number of bits. The gates and proof is essentially the same.*

10.6 Teleportation

QM does not allow for cloning an unknown quantum state. It does, however, allow for teleportation: The transfer of the unknown quantum state $|\psi\rangle$ from Alice to Bob. It is like a quantum fax except that the original is destroyed so no cloning is still respected.

Unlike teleportation of science fiction, quantum teleportation is not the transfer of material quantum qubits it is the transfer of the "information" encoded in the wave function.

Teleportation protocol:

- Alice has a qubit in an unknown state $|\psi\rangle$ which she wishes to teleport to Bob
- As a resource Alice and Bob need to share the Bell state $|\beta_\mu\rangle$ (so Alice has in total two qubits)

- Alice measures her pair of qubits in the Bell states.
- Alice calls Bob and tells him which Bell state she found, say $|\beta_\nu\rangle$.
- Bob puts his qubit through a single, unitary, qubit gate $U_{\mu\nu} = \sigma_\mu^t \sigma_\nu$.
- The output of the gate changes Alice preparation of Bobs qubit to $|\psi\rangle$

Remark 10.4.

- *No quantum particles are exchanged.*
- *Two bits of classical information are exchanged: The number* $\nu \in \{0, 1, 2, 3\}$
- *Asher Peres said that what is being teleported is the soul of the quantum particle not its body.*

Theorem 10.5 (Teleportation identity due to O. Kenneth)**.**

$$|\psi\rangle_A \otimes |\beta_\mu\rangle_{AB} = \frac{1}{2} \sum_\nu |\beta_\nu\rangle_{AA} \otimes |U_{\mu\nu}\psi\rangle_B, \quad U_{\mu\nu} = (\sigma_\mu)^t \sigma_\nu \tag{10.4}$$

If Alice measures her two qubits in the Bell basis she will find one of the bell states, say β_ν and thus remotely prepare Bob's qubit

$$|\psi\rangle_A \otimes |\beta_\mu\rangle_{AB} \xrightarrow[\text{Alice measures}]{} |\beta_\nu\rangle_{AA} \otimes |U_{\mu\nu}\psi\rangle_B \xrightarrow[\text{remote preparation}]{} |U_{\mu\nu}\psi\rangle_B$$

Remark 10.6.

- *Before Alice inform Bob of the results of her measurement, Bob does not know if his qubit is* $|\psi\rangle$ *or one of 4 unitary transformations of it:* $U_{\mu\nu}|\psi\rangle$
- *There is no superluminal propagation of quantum information. Alice needs to make a phone call*
- *The qubits did not move, only the quantum state was transfered.*
- *Since the outcome of Alice measurement is* $\nu \in 0, 1, 2, 3$*, the information she needs to sends Bob is equivalent to 2 bits*
- *There are 4 normalized and orthogonal states on the right of Eq. 10.4 and one on the left. This explains the factor 2.*
- $U_{\mu\nu}$ *is unitary. The specific form is not terribly important, what is important is that Bob knows* μ*. the Bell state they initially share, and he needs to know* ν*. He get this from Alice.*
- *Bob gets* $\nu \in \{0, 1, 2, 3\}$ *which is equivalent to* two *binary digits from Alice.*

It is enough to show the identify for β_0 and for the $|a\rangle$ in the computational basis, i.e.:

$$2|a\rangle_A \otimes |\beta_0\rangle_{AB} = \sum_\nu |\beta_\nu\rangle_{AA} \otimes |\sigma_\nu a\rangle_B \tag{10.5}$$

The case of general ψ follows by linearity,

$$2|\psi\rangle_A \otimes |\beta_0\rangle_{AB} = \sum_{\nu=0}^{3} |\beta_\nu\rangle_{AA} \otimes \sigma_\nu|\psi\rangle_B$$

and for general Bell state by Bob local operations $\mathbb{1} \otimes \sigma_\mu^t$ that permute the Bell states.

It remains to prove the identity for $\mu = 0$ and the computational basis

$$\begin{aligned}\sqrt{2}|a\rangle_A \otimes |\beta_0\rangle_{AB} &= \sum_{b\in 0,1} |a\rangle_A \otimes |bb\rangle_{AB} \\ &= \sum_{b\in 0,1} |ab\rangle_{AA} \otimes |b\rangle_B\end{aligned}$$

Now express Alice pair in the Bell basis using Exercise 7.4

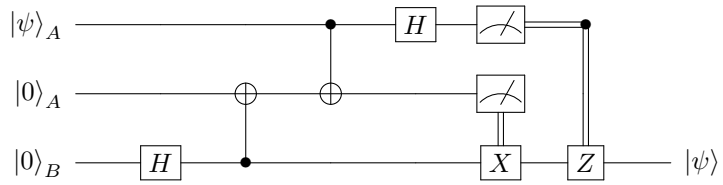
$$\sqrt{2}|ab\rangle = \sum_{\nu=0}^{3} (\sigma_\nu)_{ba}|\beta_\nu\rangle$$

to get

$$\begin{aligned}2|a\rangle_A \otimes |\beta_0\rangle_{AB} &= \sum_{b=0}^{1}\sum_{\nu=0}^{3} (\sigma_\nu)_{ba}|\beta_\nu\rangle_{AA} \otimes |b\rangle_B \\ &= \sum_{\nu=0}^{3} |\beta_\nu\rangle_{AA} \otimes \sum_{b=0}^{1} (\sigma_\nu)_{ba}|b\rangle_B \\ &= \sum_{\nu=0}^{3} |\beta_\nu\rangle_{AA} \otimes \sigma_\nu|a\rangle_B\end{aligned}$$

In the last line I used Eq. 2.1.

A circuit that implements teleportation is



- The two upper wires belong to Alice. The lowest wire is Bob's.
- The first two gates create the Bell state $|\beta_0\rangle$ that Alice and Bob share

- The next two gates are the Bell analyzer of Alice. They transform the four Bell states to the computational basis.
- The meters represent Alice measurement of her two qubits in the computational basis.
- The double lines are the classical (phone) channels where Alice tells Bob what she finds
- The results condition the gates that Bob needs to apply to retrieve $|\psi\rangle$

Exercise 10.7. *Suppose that the Bell state shared by Alice and Bob is contaminated with the identity, i.e.*

$$|\beta_0\rangle\langle\beta_0| \mapsto p|\beta_0\rangle\langle\beta_0| + (1-p)\frac{\mathbb{1}}{4}, \quad 0 < p < 1$$

Show that the teleported state is contaminated with the identity

$$|\psi\rangle\langle\psi| \mapsto p|\psi\rangle\langle\psi| + (1-p)\frac{\mathbb{1}}{2}, \quad 0 < p < 1$$

(Hint: Use linearity and the fact that $\sum |\beta_\mu\rangle\langle\beta_\mu| = \mathbb{1}$*.)*

Exercise 10.8. *Show the teleportation identity for qdits*

$$|\psi\rangle \otimes |\beta_{00}\rangle = \frac{1}{d}\sum_{j,k=1}^{d} S^k \otimes T^j \otimes U_{jk}|\beta_{00}\rangle \otimes |\psi\rangle, \quad U_{jk} = T^j S^{-k}$$

where

$$|\beta_{00}\rangle = \frac{1}{\sqrt{d}}\sum_{n=1}^{d} |n\rangle \otimes |n\rangle$$

10.7 Entanglement transfer

Suppose Alice and Bob share an entangled pair and Bob and Charlie share a (different) pair, e.g.

$$|\beta_0\rangle_{AB} \otimes |\beta_0\rangle_{BC}$$

Alice and Charlie can use Bobs services to get entangled. The basic identity of entanglement transfer is:

$$|\beta_0\rangle_{AB} \otimes |\beta_0\rangle_{BC} = \frac{1}{2}\sum_{\mu} |\beta_\mu\rangle_{BB} \otimes |\beta_\mu\rangle_{CA} \tag{10.6}$$

This means that if Bob measures his pair of qubits in the Bell basis, he prepares a Bell state for himself and also for Alice and Charlie. The price he pays is that he is not longer entangled with anybody. Entanglement is monogamous.

To show this start with

$$\begin{aligned}
|\beta_0\rangle_{AB} \otimes |\beta_0\rangle_{BC} &= \frac{1}{2} \sum_{a,b\in 0,1} |aa\rangle_{AB} \otimes |bb\rangle_{BC} \\
&= \frac{1}{2} \sum_{a,b\in 0,1} |a\rangle_A \otimes |ab\rangle_{BB} \otimes |b\rangle_C \\
&= \frac{1}{\sqrt{8}} \sum_{a,b,\mu} |a\rangle_A \otimes (\sigma_\mu)_{ba} |\beta_\mu\rangle_{BB} \otimes |b\rangle_C \\
&= \frac{1}{\sqrt{8}} \sum_{a,b,\mu} |a\rangle_A \otimes |\beta_\mu\rangle_{BB} \otimes (\sigma_\mu)_{ba} |b\rangle_C \qquad (10.7)
\end{aligned}$$

Now simply reorder the tensor product putting Bob upfront

$$\begin{aligned}
|\beta_0\rangle_{AB} \otimes |\beta_0\rangle_{BC} &= \frac{1}{2\sqrt{2}} \sum_\mu |\beta_\mu\rangle_{BB} \sum_{ab} \sigma^\mu_{ba} |b\rangle_C \otimes |a\rangle_A \\
&= \frac{1}{2} \sum_\mu |\beta_\mu\rangle_{BB} \otimes |\beta_\mu\rangle_{CA}
\end{aligned}$$

Exercise 10.9. *Repeat the exercise in the case that the initial state is*

$$|\beta_\mu\rangle_{AB} \otimes |\beta_\nu\rangle_{BC}$$

10.8 Monogamy of entanglement

In applications Alice and Bob may want to be sure that their qubits are not entangled with the eavesdropper Eve.

Suppose they were. Let $|\Psi\rangle_{ABE}$ denote the 3-parties state. Then

$$\rho_{AB} = Tr_E |\Psi\rangle_{ABE} \langle\Psi|, \quad \rho_E = Tr_{AB} |\Psi\rangle_{ABE} \langle\Psi|$$

If Alice and Bob were entangled with Eve then by the Schmidt decomposition ρ_{AB} must be mixed. It follows that if Alice and Bob can be sure that they prepared and possess a (pure) Bell state, this gives them a guarantee they are not entangled with Eve. This is (a special case of) monogamy.

Exercise 10.10. *Suppose Alice and Bob each have 2 qubits. Show that the 4 states*

$$\frac{1}{2} \sum \omega^{2a+b} |ab\rangle \otimes |ab\rangle, \quad \omega \in \pm 1, \pm i$$

are maximally entangled and orthogonal. Construct the corresponding Bell states, i.e. 16 maximally entangled orthogonal states.

10.9 Schrödinger cat: Fragile entanglement

Schrödinger's cat is a putative macroscopic object which is put into superposition of being both dead and alive. We do not encounter macroscopic objects in superposition. Why?

Consider the state

$$|\psi\rangle = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}$$

If the number of qubits, n, is large you can view this as a superposition of two macroscopic states.

This entanglement is fragile in the following sense: Suppose one qubit, say the last one, gets lost. Then remaining qubits are described by

$$\rho_A = Tr_{\text{last qubit}}|\psi\rangle\langle\psi| = \frac{\big(|0\rangle\langle 0|\big)^{\otimes(n-1)} + \big(|1\rangle\langle 1|\big)^{\otimes(n-1)}}{2}$$

This is a separable state which describes a cat that is *either* dead *or* alive.

In the famous Schrödinger cat experiment, it is enough for a single photon to be lost so as to converted the quantum super-position of dead *and* alive can to classical mixtures of dead *or* alive cat.

10.10 Classical vs quantum computers

A classical computer is a *finite* machine that accepts a program π and input $x \in \mathbb{Z}$ and outputs $y \in \mathbb{Z}$:

$$\pi : x \mapsto y, \quad x, y \in \mathbb{Z}$$

A computer is universal if any program that runs on some finite machine can be translated to an equivalent program, which I shall still call π, on the universal computer. Since π is a *finite* sequence of bits it can be identified with an integer in $\pi \in \mathbb{Z}$.

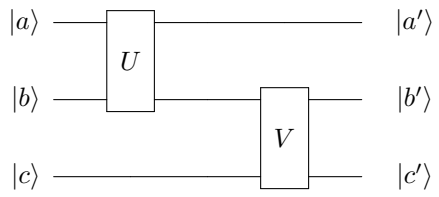
Remark 10.11. *A deep result of Turing is that not all functions can be realized as programs on a universal computer. For example, there is no program that will accept any program as input and will determine if the program halts or not.*

Turing machine is a model of a universal computer:

$$T : (\pi, x) \mapsto y, \quad \pi, x, y \in \mathbb{Z}$$

Turing did not require machines to be reversible, but it turns out that there is no loss in requiring reversibility. We may then represent a classical Turing

machine as a QM circuit:

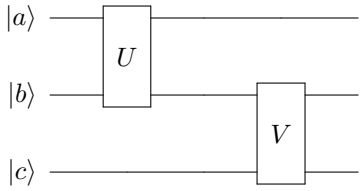


The circuit must have the following special properties in order that it represents a classical Turing machine:

- Finiteness: It must be a collection of finite number of quantum gates each operating on a finite number of qubits.
- Classical: The machine takes any *input* in the computational basis to an output that is also in the computational basis. It does not generate superpositions.

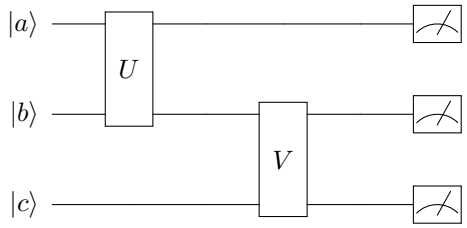
A quantum computer is almost the same thing except that we drop the condition that the output y is a state in the computational basis. y may be any state in the Hilbert space. The gates may create superpositions.

This makes it clear that a classical computer is a very special case of a quantum computer. In particular, a quantum computer *can not do less* than a classical computer.



$$|\psi\rangle = \sum \psi_{a'b'c'} |a'\rangle \otimes |b'\rangle \otimes |c'\rangle$$

In general, a quantum computer will not not give you a definite answer. The output y is a superposition. If you measure the output, always in the computational basis, you will find $|a'b'c'\rangle$ with probability $|\psi_{a'b'c'}|^2$.



The art of quantum computing is to design algorithms so that through interference $|\psi_{a'b'c'}|^2 \approx 0$ for most wrong/bad values, and $|\psi_{a''b''c''}|^2 \approx 1$ for the good/right values. We shall see an example for that in the next section.

Chapter 11

Quantum correlations

11.1 Hidden variables

Classical physics is deterministic. QM is not. Hidden variables is an attempt to embed QM within a deterministic setting. Namely, *hidden-variables* is the attempt to claim that the wave function gives an incomplete description of the system. There are hidden variables that we have not yet learned how to detect and measure and if we knew them, we would also know deterministically how the system evolves. The statistical aspects of QM therefore reflect our ignorance of these variables.

11.2 Counterfactual

Suppose you could make one of two measurements: X or Y. You choose one, say X. What about the measurements you did not perform? In classical physics we assume that also Y has a value, but it is unknown. Assigning values to the measurements you did not make is called *counterfactual*.

In standard QM X and Y are represented by matrices and the measurement does not reveal pre-existing values. Rather, facts emerge from random picking from the eigenvalues of the matrix. In particular, Y has no value if you chose to measure X.

11.2.1 Counterfactual spins

It is actually not possible to assign values for non-commuting observables consistent with QM.

In QM the commutations of angular momentum

$$[J_x, J_y] = iJ_z \tag{11.1}$$

forces the relations

$$2j \in \{0, 1, \ldots, \infty\}, \quad J_k \in \{-j, \ldots, j\}, \quad J \cdot J = j(j+1) \tag{11.2}$$

Since $[J_k, J\cdot J]=0$ the pair can have simultaneous values.

Is it possible to assign values also to the remaining components J_m?

Consider $j=3/2$. Then if J_1, J_2, J_3 all have values, by Eq. 11.2, they all have to satisfy:

$$\underbrace{J_x, J_y, J_z \in \{\pm 3/2, \pm 1/2\}}_{values}, \quad \underbrace{J\cdot J = J_x^2+J_y^2+J_z^2 = \frac{3\times 5}{4}}_{sum\ rule}$$

But this is not possible:

$$J\cdot J \in \left\{\frac{3\times 9}{4}, \frac{2\times 9+1}{4}, \frac{2\times 1+9}{4}, \frac{3}{4}\right\} = \left\{\frac{27}{4}, \frac{19}{4}, \frac{11}{4}, \frac{3}{4}\right\} \not\ni \frac{15}{4}$$

Hence J_x, J_y, J_z can not be assigned values consistent with Eq. 11.2.

Exercise 11.1. *What about $j=1/2$ and $j=1$?*

11.3 The GHZ game

QM does not attempt to assign values to counterfactuals while classical physics does. This gives QM more freedom and can lead to correlations that are classically impossible[1].

To see this consider the following cooperative game that Alice, Bob and Charlie play against the house. The rules are:

- They are allowed to make preparations and decide on a common strategy.
- Once the game starts, they are not allowed to communicate.
- Each participant is asked one of two questions: Question X or a question Y from a pool of 4 questions

$$\begin{aligned} Q_1 &= X_A X_B X_C; \quad Q_2 = X_A Y_B Y_C; \\ Q_3 &= Y_A X_B Y_C; \quad Q_4 = Y_A Y_B X_C \end{aligned} \tag{11.3}$$

- The legal response of Alice Bob and Charlie to the question is ± 1 and the answer of the team is the product of their answers.
- The participants win if the products of their responses satisfy

$$q_1 = -1; \quad q_2 = q_3 = q_4 = 1 \tag{11.4}$$

- The participants do not know what questions the other participants were asked, and what they answered.

[1] This is also called quantum pseud-telepathy

11.3.1 No classical strategy can always win

A classical strategy assigns to $\{X_{A,B,C}, Y_{A,B,C}\}$ numerical values in ± 1. It then determines the answers to all 4 questions (even though in a given round only one question is asked). There is no classical strategy that wins for all 4 questions. The 4 equations 11.4 with 6 numerical unknowns $\{x_j, y_j\} \in \pm 1$ admit no solution. To see this, consider the table:

Alice	Bob	Charlie	Win
x_A	x_B	x_C	-1
x_A	y_B	y_C	1
y_A	x_B	y_C	1
y_A	y_B	x_C	1
1	1	1	± 1

The product of all the entries in the table is $+1$ as you can see by first multiplying columns. However, to win all the times the product of all entries must be -1 as you can see by first multiplying rows.

It follows that there is no classical strategy that wins every time.

Exercise 11.2. *Show that a random flipping between strategies can not outpreform the best strategy.*

11.4 A Bell inequality

If we assign values to numerical values to the questions $X_{A,B,C}$ and $Y_{A,B,C}$ that take values in ± 1 then the the answers to the 4 questions are determined and satisfy

$$q_1 q_2 q_3 q_4 = 1 \tag{11.5}$$

which violates the product of the last column.

QM does not attempt to assign numerical values to the questions. Instead it assigns matrices, and more explicitly, the corresponding Pauli matrices. Alice, Bob and Charlie are then allowed to measure the operator corresponding to the question on a 3 qubits state the prepared ahead of time. Since the spectrum of Pauli matrices is ± 1, this means that if Alice, Bob and Carlie make the corresponding measurement it always yields a legitimate answer ± 1.

The 4 questions are now represented by four 8×8 matrices:

$$Q_1 = X_A \otimes X_B \otimes X_C, \quad Q_2 = X_A \otimes Y_B \otimes Y_C,$$
$$Q_3 = Y_A \otimes Y_B \otimes X_C, \quad Q_4 = Y_A \otimes X_B \otimes Y_C$$

Note that the four questions are mutually commuting

$$[Q_j, Q_k] = 0$$

and satisfy the operator equality

$$Q_1 Q_2 Q_3 Q_4 = -\mathbb{1} \tag{11.6}$$

This means that the right hand column of the table is automatically satisfied (and the order does not matter).

I shall call the conflict, expressed by the fact that 4 quantum measurements necessarily gives $-\mathbb{1}$ while the classical assignment necessarily satisfied $+1$ a *Bell inequality*.

This is the conflict between quantum mechanics and any classical theory (such as hidden variables) which assigns (commuting) values also to questions that have not been asked.

11.4.1 The GHZ state

There are 4 commuting questions Q_μ with $Q_\mu^2 = \mathbb{1}$. This means that

$$Spect(Q_\mu) = \{\pm 1\}$$

The 4 questions are related by one relation, Eq.11.6. So three questions are independent. Since Q_μ are 8×8 commuting matrices we can find the eight joint eigenvalues corresponding to all all ± 1 eigenvalues to three questions. In particular, it follows that there is an eigenvector, which we call $|GHZ\rangle$ named after Greenberger-Horn-Zeilinger, such that

$$\begin{aligned} Q_1|GHZ\rangle &= -|GHZ\rangle, \\ Q_2|GHZ\rangle = Q_3|GHZ\rangle &= Q_4|GHZ\rangle = |GHZ\rangle \end{aligned} \tag{11.7}$$

Using

$$Z|a\rangle = (-)^a|a\rangle, \quad X|a\rangle = |a+1\rangle, \quad Y|a\rangle = iXZ|a\rangle = i(-)^a|a+1\rangle$$

one verifies that

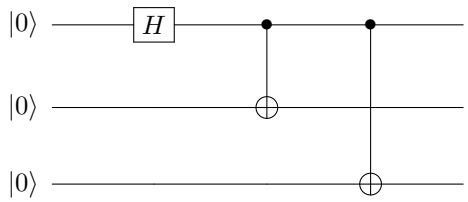
$$|GHZ\rangle = \frac{|000\rangle - |111\rangle}{\sqrt{2}}$$

The four questions have deterministic values in the GHZ state.

This happens without assigning sure values to (the four) questions that have not been asked, such as

$$X_A \otimes X_B \otimes Y_C, \quad \ldots, \quad Y_A \otimes Y_B \otimes Y_C$$

Exercise 11.3. *Show that the circuit*



generatess GHZ.

Exercise 11.4. *Find the state that will answer wrongly all 4 questions.*

11.4.2 Winning the game with GHZ

Alice, Bob and Charlie prepare a GHZ state (a fresh one for every round of questions): Alice takes with her, to her separate room, the first qubit, Bob the second and Charlie the third. Upon being asked a question x or y each measures the corresponding observable X or Y. The result of each measurement is, of course, ± 1 and is random.

ABC are guaranteed to win by the fact that GHZ is an eigenstate of the four questions by Eq. 11.7.

It is important to observe that the answers Alice, Bob and Charlie give are random. For a given question, there are 4 possible winning responses (and four losing ones). ABC will give all possible correct answers with equal probability.

Exercise 11.5. *Show that the probability that Alice answers + to question X in GHZ state is*

$$\left\langle GHZ \left| \frac{\mathbb{1} \pm X}{2} \otimes \mathbb{1} \otimes \mathbb{1} \right| GHZ \right\rangle = \frac{1}{2}$$

and similarly for question Y

Exercise 11.6. *Show that*

$$2|GHZ\rangle = |+\rangle \otimes |+\rangle \otimes |-\rangle + |+\rangle \otimes |-\rangle \otimes |+\rangle + |-\rangle \otimes |+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle \otimes |-\rangle$$

What do you learn from that about the possible answers to question Q_1?

Exercise 11.7. *Show that*

$$\langle GHZ|(\mathbb{1} + \alpha X) \otimes (\mathbb{1} + \beta X) \otimes (\mathbb{1} + \gamma X)|GHZ\rangle = 1 - \alpha\beta\gamma$$

and

$$\langle GHZ|(\mathbb{1} + \alpha X) \otimes (\mathbb{1} + \beta Y) \otimes (\mathbb{1} + \gamma Y)|GHZ\rangle = 1 + \alpha\beta\gamma$$

Exercise 11.8. *Mermin-Peres square. (Wikipedia)*

11.5 Sharing secrets with partners you mistrust

Suppose Alice wants to send Bob and Charlie a message, the bit b. Say part of a key to safe. She does not trust either Bob or Charlie and wants to send a message that can be deciphered only when both cooperate. Classically this can be done by Alice choosing randomly a and sending Bob $a+b$ and Charlie a and tell them to add their messages:

$$b \to \{b + a, a\} \to b + 2a = b$$

The weakness of this is that Alice needs a private channel to transmit the messages to each partner. She can not broadcast.

If ABC share a GHZ state Alice can transmit the secret by broadcasting.

Protocol:

- Alice measures X and finds $x = \pm 1$.
- She broadcast to Bob and Charlie to both measure X or Y according to:

$$\underbrace{X\delta_{x,-1} + Y\delta_{x,1}}_{\text{to transmit 0}}, \quad \underbrace{X\delta_{x,1} + Y\delta_{x,-1}}_{\text{to transmit 1}}$$

  (depending on what she found and the bit she wants to transmit).
- The secret is the parity of the product of Bob and Charlie's results.

11.6 The Quantum view of reality

Quantum mechanics departs from classical physics. Classical physics assumes that there is an objective reality which, in principle at least, can be discovered by the observer. In quantum mechanics, the act of observations prepares the system, in general, in a new state and it is not possible to tell what has been the state of the system before the measurement. The Copenhagen view of quantum mechanics is that there is no reality independent of observation. The position of a quantum particle takes its value once the position is measured. Prior to the observation the position is an operator and has no value. The uncertainty principle prevents the system to be in a well defined state of non-commuting observable. This strange view of reality was challenged by Einstein who asked: Do you believed that the moon is there only when you look?

11.6.1 Hidden variables

Hidden variables attempt to reconcile QM with the classical view of an objective reality (realism). QM is viewed as an effective theory of a more fundamental theory which is deterministic and realistic involving variables that we have not yet learned how to measure and control. These are the hidden variables. The probabilistic nature of QM then arises from our ignorance of these variables.

John Bell showed that the hypothesis of Hidden variables can be tested experimentally under appropriate assumptions. The test involves measurements of correlations of qubits in an entangled state.

11.6.2 Bell inequalities

The simplest Bell test is the CHSH inequality (the initials of those who found it). Consider an experiment with a pair of qubits illustrated in Fig. 11.1. The test involves a pair of qubits in each run and is repeated many times. Alice chooses to measure her qubits in one of two directions labeled by $\alpha = \pm 1$ and similarly Bob picks two directions labeled by $\beta = \pm 1$ to measure his qubits. In each run Alice finds $a_\alpha = \pm 1$ and Bob find $b_\beta = \pm 1$.

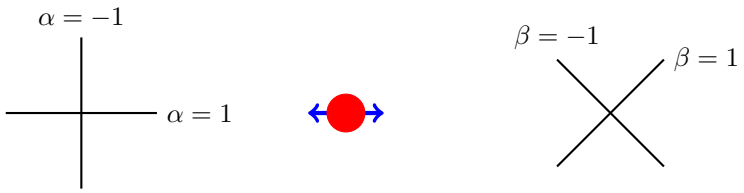


Figure 11.1: An entangled pair is generated at the red dot. Alice and Bob receive one qubit each. Alice chooses randomly one of two directions to measure her qubit and so does Bob. The experiment is repeated many times and statistics of correlations is made.

11.6.3 CHSH

The hidden variables view of the CHSH setting is as follows: The correlations between the measurements of Alice and Bob arise from the way the two qubits have been prepared at the source. The Hidden variables λ encode this information. We denote by $P(\lambda)$ their distribution.

Alice and Bob actually make 4 different experiments

$$\{\alpha, \beta\}, \quad \alpha, \beta = \pm 1 \tag{11.8}$$

In world that is realistic, a_1 and a_{-1} coexist even though only one of them is measured in any given experiment. In a world that is also deterministic, this means that there are 2^4 functions

$$a_j(\lambda, \alpha, \beta) \in \pm 1, \quad b_j(\lambda, \alpha, \beta) \in \pm 1 \tag{11.9}$$

that describe the reality underlying the 4 possible experiments.

Since $0 = b_1^2 - b_{-1}^2 = (b_1 + b_{-1})(b_1 - b_2 - 1)$ one term in the product must vanish. It follows that in all 4 experiments, for any value of the hidden λ

$$a_1(\lambda, \alpha, \beta)\big(b_1(\lambda, \alpha, \beta) + b_2(\lambda, \alpha, \beta)\big) + a_2(\lambda, \alpha, \beta)\big(b_1(\lambda, \alpha, \beta) - b_2(\lambda, \alpha, \beta)\big) = \pm 2$$

The equality can not be tested experimentally because it involves counterfactuals: In any given experiment only one term among the four is measured, namely

$$a_\alpha(\lambda, \alpha, \beta) b_\beta(\lambda, \alpha, \beta) = \pm 1 \tag{11.10}$$

If we would take the four terms from four different experiments then the ± 2 on rhs is replaced by $0, \pm 2, \pm 4$.

We can formally take the expectation with respect to λ, α, β to get the CHSH inequality

$$-2 \leq \mathbb{E}(a_1 b_1) + \mathbb{E}(a_1 b_2) + \mathbb{E}(a_2 b_1) - \mathbb{E}(a_2 b_2) \leq 2 \tag{11.11}$$

The trouble with this inequality is that the average is over both real and imagined measurements and it is not clear how to relate it to what one can actually measure.

To proceed we examine the functions a_j, b_k more closely. The functions a_j and b_k are supposed to take the value of actual measurement, had such a measurement been carried out. This means that a_j must take the same value for $\alpha = \pm j$ and similarly for b_k. This can be written as

$$a_j(\lambda, \alpha, \beta) = a_j(\lambda, \beta), \quad b_k(\lambda, \alpha, \beta) = b_k(\lambda, \alpha) \tag{11.12}$$

and hence

$$a_j(\lambda, \alpha, \beta) b_k(\lambda, \alpha, \beta) = a_j(\lambda, \beta) b_k(\lambda, \alpha) \tag{11.13}$$

The rhs still depends on $\alpha\beta$ and therefore the average involves both real and imagined measurements. To proceed we make the *physical* assumption that the theory is *local*. This means that the the measurment of Alice is independent of what Bob chooses to measure and vice versa. This means:

$$a_j(\lambda, \alpha, \beta) b_k(\lambda, \alpha, \beta) = a_j(\lambda, \beta) b_k(\lambda, \alpha) = a_j(\lambda) b_k(\lambda) \tag{11.14}$$

This guarantees that the observable take the same values in real and imagined experiments and we can then apply CHSH to the actual experimental correlations.

11.7 Tsirelson bound

Let T be the observable

$$T = A_1 \otimes (B_1 + B_2) + A_2 \otimes (B_1 - B_2) \tag{11.15}$$

where A_j and B_j satisfy $A_j^2 = B_k^2 = \mathbb{1}$. A computation gives

$$0 \leq T^2 = 4\ \mathbb{1} \otimes \mathbb{1} + i[A_0, A_1] \otimes i[B_0, B_1]$$

Since

$$-2 \leq i[A_1, A_2] \leq 2$$

it follows that

$$0 \leq T^2 \leq 8 \Longrightarrow -2\sqrt{2} \leq T \leq 2\sqrt{2} \tag{11.16}$$

11.7.1 Observables that Saturate Tsirelson bound

We can saturate the Tsirelson with:

$$A_1 = H, \quad A_2 = ZHZ, \quad B_1 = Z, \quad B_2 = X = HZH \tag{11.17}$$

A computation gives

$$T = \sqrt{2}\left(Z \otimes Z + X \otimes X\right) \tag{11.18}$$

Since $Z \otimes Z$ and $X \otimes X$ commute and their spectra are ± 1 and therefore

$$Spect(T) = \sqrt{2}\{\pm 2, 0\} \tag{11.19}$$

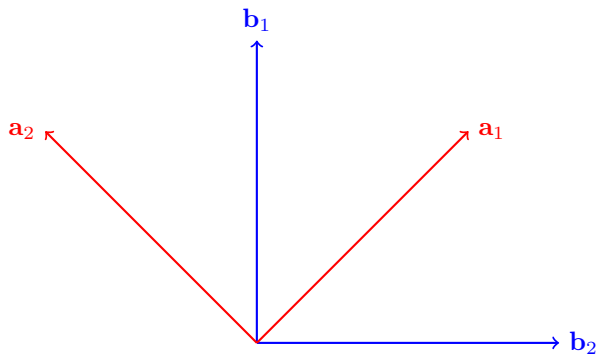


Figure 11.2: The angle between the unit vectors is such that $\mathbf{a}_j \cdot \mathbf{B}_j = \cos \pi/4$ except for $\mathbf{a_2} \cdot \mathbf{b_2} = -\cos \pi/4$. This saturates the Bell inequality since $\mathbf{a_1 b_2} + \mathbf{a_1 b_1} + \mathbf{a_2 b_1} - \mathbf{a_2 b_2} = 2\sqrt{2}$

saturates the bound Eq. 11.16.

Moreover, since the $Z \times Z$ and $X \times X$ are syndromes of the Bell states the maximal and minimal eigenvector of T must be Bell states. In fact

$$\begin{aligned} T|\beta_2\rangle &= -2\sqrt{2}|\beta_2\rangle, \quad |\beta_2\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}} \\ T|\beta_0\rangle &= 2\sqrt{2}|\beta_0\rangle, \quad |\beta_0\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \end{aligned} \tag{11.20}$$

They maximally violates the Bell inequality:

$$\langle\beta_0|T|\beta_0\rangle = -\langle\beta_2|T|\beta_2\rangle = 2\sqrt{2} \tag{11.21}$$

11.7.2 Geometric picture

For the Bell singlet

$$\langle\beta_2|\sigma_j \otimes \sigma_k|\beta_2\rangle = -\delta_{jk}$$

It follows that

$$\langle\beta_2|\mathbf{a} \cdot \boldsymbol{\sigma} \otimes \mathbf{b} \cdot \boldsymbol{\sigma}|\beta_2\rangle = -\mathbf{a} \cdot \mathbf{b} \tag{11.22}$$

and therefore QM is in conflict with hidden variables if we can find unit vectors so that

$$|\mathbf{a_1} \cdot (\mathbf{b_1} + \mathbf{b_2}) + \mathbf{a_2} \cdot (\mathbf{b_1} - \mathbf{b_2})| > 2$$

Observe that with $\mathbf{b_j}$ unit vectors, $\mathbf{b_1} \pm \mathbf{b_2}$ are orthogonal vectors whose maximal length $\sqrt{2}$.

To maximize the violation of Bell inequalities we then want $\mathbf{b}_1 \pm \mathbf{b}_2$ to be parallel to $\mathbf{a}_1$ and $\mathbf{a}_2$. This gives the vector in the figure 12.4 which saturate the Tsirelson bound.

Exercise 11.9. *Show that separable states satisfy the CHSH inequality.*

11.7.3 The CHSH Game

The rules of the game are

- Alice and Bob can not communicate but are allowed to share a Bell pair.
- Alice is asked a question $A \in \{0,1\}$ to which she answers $a \in \{0,1\}$
- Similarly, Bob is asked a question $B \in \{0,1\}$ to which he answers $b \in \{0,1\}$
- They win the game if the responses match the questions according to

$$a \oplus b \oplus 1 = A \cdot B \tag{11.23}$$

  In words Alice and Bob need to choose opposite bits except if both were asked $A = B = 1$.

You can not satisfy Eq. 11.23 in all cases. Suppose you could satisfy Eq. 11.23 then summation Mod 2 over all cases on the rhs gives:

$$\begin{aligned}
1 &= \sum_{A,B\in 0,1} A \cdot B \\
&= \sum_{A,B\in 0,1} a(A) \oplus b(B) \oplus 1 \\
&= \sum_{B\in 0,1} \big(a(0) \oplus b(B) \oplus 1\big) \oplus \big(a(1) \oplus b(B) \oplus 1\big) \\
&= \sum_{B\in 0,1} \big(a(0) \oplus a(1)\big) \\
&= 0
\end{aligned}$$

Alice and Bob must therefor fail on at least one question. They may, for example, agree that Alice always chooses 0 while Bob always chooses 1. If all questions are asked with equal probability their win-loss ratio is at best 3:1.

Exercise 11.10. *Give an argument why a random strategy will not help (Hint: Use convexity)*

If Alice and Bob share Bell pairs, they can improve their winning ratio to about 85%.

- Depending on $A \in 0,1$ Alice measures her qubit in the direction as in Fig. 11.2.
- The measurement determines Alice response
- Similarly for Bob.

By Eq. 7.6

$$Prob(\pm\mp|\mathbf{A},\mathbf{B},\beta_2)=\frac{1+\mathbf{A}\cdot\mathbf{B}}{4},\quad Prob(\pm\pm|\mathbf{A},\mathbf{B},\beta_2)=\frac{1-\mathbf{A}\cdot\mathbf{B}}{4}$$

so the probability that Alice and Bob give the same answer when they measure in the $\mathbf{A}$, $\mathbf{B}$ directions is

$$Prob(opposite|\mathbf{A},\mathbf{B},\beta_2)=\frac{1+\mathbf{A}\cdot\mathbf{B}}{2},\quad Prob(same|\mathbf{A},\mathbf{B},\beta_2)=\frac{1-\mathbf{A}\cdot\mathbf{B}}{2}$$

For the directions in Fig 11.2

$$Prob(opposite|\mathbf{A_j},\mathbf{B_k},\beta_2)=\frac{1+\cos\pi/4}{2},\quad j\cdot k=0$$

and

$$Prob(same|\mathbf{A_1},\mathbf{B_1},\beta_2)=\frac{1+\cos\pi/4}{2}$$

We see that for all 4 questions the winning probability is the same. Hence

Theorem 11.11. *If Alice and Bob share a Bell state, and follow the above recipe they win the CHSH game with probability*

$$\frac{1+\cos(\pi/4)}{2}\approx .85$$

This beats the classical optimum .75.

11.8 QM is non-signaling

Let

$$P(a,b|A,B)$$

denote the (conditional) probability that if Alice chooses to make the test A and Bob B she gets the result $a\in\pm1$ and he $b\in\pm1$. In the CHSH setting A and B are the two directions.

Definition 11.12. *The conditional probability $P(a,b|A,B)$ is called non-signaling if the marginal*

$$P(a|A,B)=\sum_b P(a,b|A,B)$$

is independent of B (and similarly for $(a,A)\leftrightarrow(b,B)$).

In words, by looking at her data, Alice gets no information on the experiment Bob choose to make.

Alice and Bob choose directions $\mathbf{a}$ and $\mathbf{b}$ to measure their qubits. The probabilities for a given event $(\pm,\pm)$ is given by the associated projection, namely:

$$P(\pm,\pm|\mathbf{a},\mathbf{b})=Tr\left(\underbrace{\frac{\mathbb{1}\pm\mathbf{a}\cdot\boldsymbol{\sigma}}{2}}_{projection}\otimes\underbrace{\frac{\mathbb{1}\pm\mathbf{b}\cdot\boldsymbol{\sigma}}{2}}_{projection}\rho\right)$$

with $a, b \in \pm 1$. Evidently

$$\sum_{\pm} P(\pm, \pm|\mathbf{a}, \mathbf{b}) = Tr\left(\frac{\mathbb{1}\pm\mathbf{a}\cdot\sigma}{2} \otimes \mathbb{1}\rho\right)$$

is independent of $\mathbf{b}$. Alice can not tell what Bob chose to measure by collecting only statistics of her qubit. We have therefore shown that projective measurements on 2 qubits are non-signaling.

11.9 Popescu Rohrlich box

Popescu and Rohrlich asked the question: Suppose you accept the principle that correlations do not allow (superluminal) signaling. Does this impose a constraint on the correlations? In other words, does the QM bound of $\pm 2\sqrt{2}$ on correlations follow from a "No signaling" principle. The answer turns out to be no. Moreover, correlations and signaling are not directly related.

The Popescu-Rohrlich box is an oracle that answer the four questions of the game in section 11.4 without mistake[2] :

$$a_1 b_1 = a_1 b_2 = a_2 b_1 = 1, \quad a_2 b_2 = -1 \tag{11.24}$$

In particular, Tsirelson inequality is violated:

$$a_1 b_1 + a_1 b_2 + a_2 b_1 - a_2 b_2 = 4$$

Does this imply signaling? To investigate this consider the table of probabilities

		B_1		B_2	
		1	-1	1	-1
A_1	1	p	0	q	0
	-1	0	1-p	0	1-q
A_2	1	p	0	0	q
	-1	0	1-p	1-q	0

By Eq. 11.23 it indeed maximally violates Tsirelson inequality

$$\langle a_1 b_1 \rangle + \langle a_1 b_2 \rangle + \langle a_2 b_1 \rangle - \langle a_2 b_2 \rangle = 4$$

It is signaling for $p \neq q$ but non-signaling if $p = q$. Rohrlich was then forced to conclude that the principle of non-signaling does not lead to the Tsirelson inequality and you can not derive QM from the principle of non-signaling. But, of course, QM is consistent with the principle of no-signaling as we have seen.

[2] There is no way to assign values to to all a_j and b_j simultaneously so that Eq. 11.24 holds, as can be seen by multiplying the four equalities.

Chapter 12

Grover search algorithm

12.1 Searching an ordered data base

In the old days we used to have telephone books. The book had many names, say one million, ordered lexicographically. Each name X is written in n bits, say $n = 20$, in one column and in the next column, is $Y(X)$, the phone number.

People searched telephone numbers using the method of *divide and conquer*: You'd need to turn about 20 pages to find a number [1]. You open the book in the middle, at the 2^{n-1} entry, and compare with the name X. to determines if X is in the first half of the phone book or the second half. Repeating the procedure you locate the number in $O(n)$ steps.

Figure 12.1: Divides and conquer: You isolate the red point in $\log_2 N$ queries.

12.2 Unstructured data base

If you have a phone number and want to find the name an ordinary phone book is not very useful. (It is best to call and ask, hoping they will tell you.) The answer is buried in the phone book. The book "knows" the answer, but it is not clear how to search it.

You can try and read the entire phone book. If you are lucky, you succeed in the first entry. If you are unlucky it will be the last. If you decide on making random queries of the phone book, the probability that in m queries you still guessed wrong is

$$(1 - 1/N)^m = \left((1 - 1/N)^N\right)^{m/N} \to e^{-m/N}$$

[1] $2^{10} = 1024 \Longrightarrow 2^{20} > 10^6$.

In either case, the search will have a finite success probability with $m = O(N)$. The fact that each query of the phone book is quick does not help because you need many queries.

12.2.1 Oracles and one way functions

You may think of the phone book as an oracle: It will give you quickly the number associated to the name, but will refuse to give you the name associated to a number. If you phrase the question right the oracle tells you an answer. A phone book is like a *one way function*

$$phone\ book : name \mapsto phone\ number$$

It is easy to compute the function and much more difficult to compute its inverse.

12.2.2 Complexity for kids

Complexity is the heart of computer science. It classifies problems according to how hard they are.

A problem is easy if a (classical) computer can give an answer in polynomial time in the the number of digits of the input, n. For example, adding two numbers is easy. So is multiplying them. Euclid algorithm for finding $GCD(p, q)$ is a more fancy example. All these problems are said to be in P.

NP is a class of hard problems. It is a little unfortunate that NP does not stand for Non-Polynomial, but rather for Nondeterministic polynomial. This means (roughly) that these are problems that can be solved in polynomial time given an "oracle". NP problems are the class of problems where finding a solution is hard (for this you need the oracle) but verifying the solution (given by the oracle) can be done in P.

Searching in unstructured data is NP.

12.2.3 Every problem is a search problem

Suppose you have a problem that has a solution that could be verified by a computer program. This means that there is a function δ so that for a proposal Y the function gives $\delta(Y) = True$ if Y is the solution and $\delta(Y) = False$ if it is not. Finding the solution is a search problem

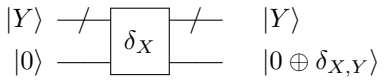
`Find` Y `so that` $\delta(Y) = True$

A stupid strategy is to examine all Y with n digits sequentially. This involves $O(N)$ operations. This is as bad as it gets.

Exercise 12.1. *Formulate the problem of searching the name associated with a phone number as a search problem.*

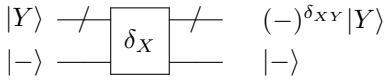
12.2.4 Quantum Oracle

The oracle is a unitary gate that we are allowed to query. It tells you if X is the solution to the problem and is represented by the circuit:

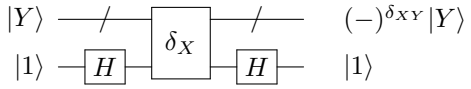


The flag in the second qubit. It flips if the input $Y = X$ right. It is promised that X is one of the basis vectors in the computational basis, e.g. $|a_1 \dots a_n\rangle, \quad a_j \in 0, 1$.

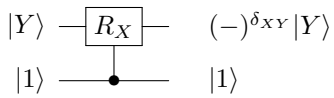
By the phase kickback trick



We can write the circuit also with the bottom qubit set to be $|1\rangle$:



All three gates, 2 Hadamard and the function gate act like the controlled reflection gate $C(R_X)$ with the control set to be on:



where

$$R_X = \mathbb{1} - 2|X\rangle\langle X|$$

This is the oracle we are allowed to use.

12.3 The search problem

You are given the oracle that performs the unitary gate operation

$$R_X = \mathbb{1} - 2|X\rangle\langle X|$$

for some unknown vector $|X\rangle$ is the computational basis. You are not told what X is. Your task is to find X.

With n qubits X could take any one of $N = 2^n$ values. You could try all entries $0, 1, \dots, N-1$ one by one, and look for the flag in the second qubit to wave. This will take you $O(N)$ queries of the gate to find X.

Grover made the remarkable discovery that with $O(\sqrt{N})$ operations you can find X with high probability.

This problem is of the same type as searching in an unstructured data base. Grover algorithm does not make hard problems (NP) easy (P), but gives a substantial gain nevertheless.

12.3.1 The Democratic superposition

When we do a quantum search we are given one piece of (important) information: Namely, the requisite item X is not any vector in the Hilbert space, but rather one of the basis vectors in the computational basis, i.e.

$$|X\rangle = |a_0 \dots a_n\rangle, \quad a_j \in 0, 1$$

Grover stars with the following simple observation. The democratic superposition

$$|D\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} |j\rangle$$

overlaps with X by

$$\langle X|D\rangle = \frac{1}{\sqrt{N}}$$

Grover makes use of this $\sqrt{N}$. Of course, if we measure we fall back to the classical result since the success probability is

$$Prob(D|X) = |\langle X|D\rangle|^2 = \frac{1}{N}$$

as in a classical search.

Grover idea is to try and increase the overlap by a series of $\sqrt{N}$ unitary rotations that turn Y towards X.

12.3.2 Reflections and rotations

Definition 12.2. *An orthogonal transformations R is called a rotation if* $\det R = 1$. *It is called a reflection if* $Eigenvalues(R) = \{-1, 1, \dots 1\}$. *The eigenvector associated with the eigenvalue -1 is the axis of reflection. Its orthogonal complement is the reflection hyperplane. In particular, for a reflection* $\det R = -1$ *and the product of two reflections is a rotation.*

Exercise 12.3. *What rotation is the product of the reflection of the X and Y axis in three dimensions.*

Consider the product of two reflections about the axis of two (different, real, unit) vectors X and Y:

$$R = R_Y R_X, \quad R_X = \mathbb{1} - 2\underbrace{|X\rangle\langle X|}_{P_X}, \quad R_Y = \mathbb{1} - 2\underbrace{|Y\rangle\langle Y|}_{P_Y}$$

where

$$\langle X|Y\rangle = \cos\theta \neq \pm 1$$

This must be a rotation. We want to determine what rotation it is.

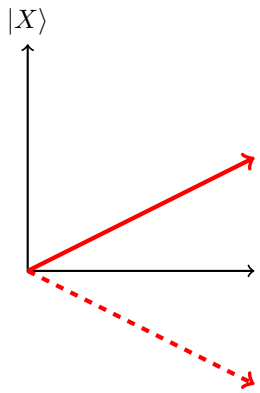


Figure 12.2: Reflection about the $|X\rangle$ direction.

$|X\rangle$ and $|Y\rangle$ span a 2-dimensional plane. An orthogonal basis in this plane is

$$|a_\pm\rangle = \frac{|X\rangle \pm |Y\rangle}{\sqrt{2(1 \pm \cos\theta)}}$$

The projection on this plane is given by

$$P = |a_+\rangle\langle a_+| + |a_-\rangle\langle a_-|,$$

It is clear that R acts trivially on $\mathbb{1} - P$ and must be a rotation of this plane.

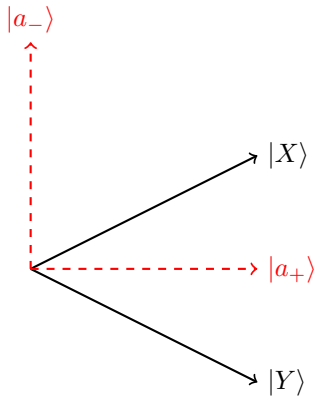


Figure 12.3: Construction of a basis in *Range* $X \oplus$ *Range* Y

The non-trivial part of R namely PRP can be viewed as a 2×2 rotation

matrix in two dimensions. Any such matrix can be written as

$$\begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$$

The angle ϕ can be computed in terms of θ by comparing the traces

$$2\cos\phi = Tr\left(PR_Y R_X P\right) \tag{12.1}$$

Using

$$\begin{aligned}
\langle a_\pm|R_Y R_X|a_\pm\rangle &= 1 - 2|\langle a_\pm|X\rangle|^2 - 2|\langle a_\pm|Y\rangle|^2 + 4\langle a_\pm|X\rangle\langle X|Y\rangle\langle Y|a_\pm\rangle \\
&= 1 - (1 \pm \cos\theta) - (1 \pm \cos\theta) \pm 2(1 \pm \cos\theta)\cos\theta \\
&= -1 + 2\cos^2\theta \\
&= \cos 2\theta
\end{aligned}$$

We see that

$$Tr\left(PR_Y R_X P\right) = 2\cos(2\theta)$$

and the angle of rotation is $\phi = 2\theta$. This is illustrated in the figure.

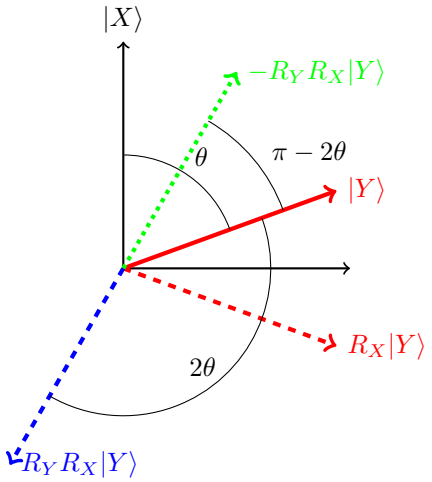


Figure 12.4: Two reflections make a rotation by twice the angle between the axes.

12.4 Grover box

We have seen that the product of two reflections is a rotation by 2θ. The trouble is that it is a rotation that does not rotate Y to X but rather Y away from X.

This can be easily fixed by an overall minus sign. So, lets define the Grover box by

$$-\boxed{G_X}- \;=\; -\boxed{R_X}-\boxed{-R_D}-$$

This gives a rotation by

$$\pi - 2\theta = 2\left(\frac{\pi}{2} - \theta\right)$$

rotating Y towards X.

Now take $Y = D$ to be the democratic superposition and X is the searched item

$$|D\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} |j\rangle$$

Here

$$\langle D|X\rangle = \frac{1}{\sqrt{N}} = \cos\theta$$

The democratic superposition is almost orthogonal to the searched item X if N is large: θ is very close to $\pi/2$

$$\theta \approx \frac{\pi}{2} - \frac{1}{\sqrt{N}} \Longrightarrow \frac{\pi}{2} - \theta \approx \frac{1}{\sqrt{N}}$$

The Grover gate then gives rotation that is close to

$$\pi - 2\theta \approx \frac{2}{\sqrt{N}}$$

from the direction of D towards the unknown X. It follows that

$$\left\lfloor \frac{\pi\sqrt{N}}{4} \right\rfloor$$

applications of Grover gate rotate the democratic state close to the unknown state $|X\rangle$, i.e. within an angle

$$O\left(\frac{1}{\sqrt{N}}\right)$$

The probability that measuring the output in the computational basis yields the correct answer is very close to 1:

$$Prob(success) = \left(1 - O\left(\frac{1}{\sqrt{N}}\right)\right)^2 \approx 1 - O\left(\frac{1}{\sqrt{N}}\right)$$

It is not important to make precisely $\frac{\pi\sqrt{N}}{4}$. For example, for success probability of say $p = 1/2$ we need to rotate by $\pi/4$ rather than $\pi/2$ and this requiters

$$\left\lfloor \frac{\pi\sqrt{N}}{8} \right\rfloor$$

applications of the gate.

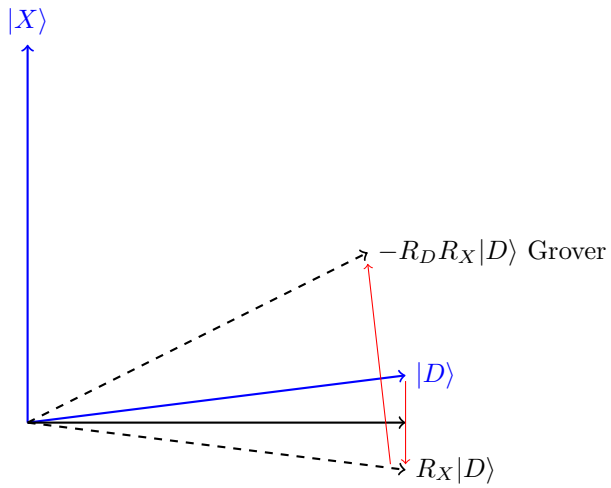
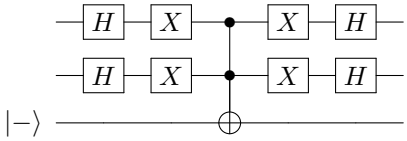


Figure 12.5: The two Grover reflections $-R_D RX$ rotate the initial democratic superposition towards the solution by an angle $O(2/\sqrt{N})$

Exercise 12.4. *Show that the circuit implements R_Y for two qubits.*



Remark 12.5 (No fixed point)**.** *Since R is a rotation it has no fixed point. A consequence of this is a not-so-nice feature of Grover, namely, that the solution $|X\rangle$ is not a fixed point of R: You should not apply R too many time; if you do you overshoot the solution. With $\frac{\pi\sqrt{N}}{2}$ you are again almost orthogonal to $|X\rangle$.*

Remark 12.6 (Multiple solutions)**.** *If the search problem has several solutions then take $|X\rangle$ to be the superposition of all of them.*

Exercise 12.7. *What happens if there is no solution?*

Remark 12.8. *If there are lots of solutions, $O(N)$, the problem is easy and you do not need any fancy algorithm.*

Exercise 12.9. *if you believe that $P \neq NP$ then, according to Bennett, Bernstein, Brassard and Vazirani Grover is optimal. If you explain this result to me you'll get bonus points.*

Chapter 13

RSA for pedestrians

13.1 Public key

RSA is a public key encryption system that allows anyone to encrypt a message, but only the owner of the key to decipher. It has a public key that allows anybody to encrypt and a private key that allows only the owner to decrypt.

It is named after Rivest, Shamir and Adleman. Shamir is at WIS, the other two are in MIT, where the algorithm was developed in 1977. I think it made them rich.

Clifford Cocks, an English mathematician, working for the British intelligence, had developed an equivalent system earlier, in 1973, but it was not declassified until 1997.

The story is nicely told in Simon Singh book *Codes*. Wikipedia will tell you what RSA has to do with drinking too much Manischwitz wine on Passover.

You can use RSA also to sign: Using your private key you can sign a random message while anybody can read and verify your signature using the public key.

RSA is a practical method in the sense that it does not guarantee security forever, it only guarantees security for any finite time you care about, say one year.

13.1.1 RSA challenge

The security of RSA rests on the presumed difficulty of factoring semi-primes—numbers of the form $N = pq, \quad p, q \in \mathbb{N}$. You only need N to encrypt, but to decipher you need to know the factors.

RSA assumes that if you made public a semi-prime with few hundred digits, the factors q and p are still secret for all practical purposes, ie. it will take, say, a year to find the factors.

RSA is not more difficult than factoring. It is not known if it is as difficult, i.e it is not known if one can break RSA without factoring.

Security depends on what resources you grant the adversary. Of course if the adversary can access your computer you are in trouble. Shamir showed that

RSA is compromised if the adversary is are allowed to listen to the noises the hard drive of the decrypting computer.

13.1.2 Quantum threat

In 1994 Shor gave an efficient algorithm for factoring integers that has a quantum subroutine. To run this subroutine we need a quantum computer. If one could be built that could handle few hundred qubits then RSA would be insecure. Quantum computers, if they exist, compromise RSA. I do not know if it the algorithm made Shor rich. It definitely made him famous.

It is not known if there is a classical algorithm that factors semi-primes quickly. None is known.

13.2 Number theory

The fundamental theorem of arithmetic, going back to Euclid (300 BC[1]), states that any integer $N \geq 2$ has unique factorization into prime factors.

$$N = p_1^{n_1} \dots p_n^{n_n}, \quad p_j \in Primes, \quad n_j \in \mathbb{Z}$$

13.2.1 GCD

Definition 13.1. *r and s are relatively prime if $GCD(r, s) = 1$*

Euclid algorithm computes GCD efficiently, more precisely, if r, s are two n digit numbers, then GCD costs $O(n^2)$.

A basic fact is that $GCD(r, s) = 1$ implies that r has an inverse $\mod s$. In particular, if p is a prime then $GCD(r, p) = 1$ for all $r \neq 0 \mod p$ and so every such r has n inverse.

For example, since 7 is prime

$$5 \times 3 = 15 = 1 + 14 = 1 \; Mod \; 7 \Longrightarrow 5^{-1} \; Mod \; 7 = 3$$

To understand how RSA works we shall need:

Theorem 13.2 (Fermat little theorem)**.** *If $x \in \mathbb{Z}$ and p prime*

$$x^p = x \mod p \tag{13.1}$$

If $x \neq 0 \mod p$ we also have

$$x^{p-1} = 1 \mod p \tag{13.2}$$

[1] Roughly the times of Alexander the great. Between the Persian and Greeks in Jewish history

Fermat, as usual, gave no proof. A proof, by induction, taken from Wikipedia, goes like this: Observe first that if p is a prime then for $1 \leq n < p$

$$\underbrace{\binom{p}{n}}_{integer} = \frac{p!}{(p-n)!n!} = \underbrace{p\frac{(p-1)!}{(p-n)!n!}}_{Unique\ factorization} = 0 \mod p, \tag{13.3}$$

Since the denominator can not cancel p the result follows from unique factorization.

We shall now prove by induction that $x^p = x \mod p$. The induction assumption holds for $x = 0$:

$$x^p = x \mod p, \quad x = 0$$

By Eq. 13.3 for any integer k

$$(k+1)^p = k^p + 1 \mod p$$

By the induction hypothesis

$$k^p + 1 = k + 1 \mod p$$

This proves the first half of the theorem. The second follows from the fact that $j \neq 0 \mod p$ has an inverse.

In fact, for RSA makes use of a special case of a stronger result known as of Fermat-Euler theorem. The special case we shall need is:

Suppose x is not divisible by p or q then

$$x^{(p-1)(q-1)} = 1 \, mod \, pq \tag{13.4}$$

Exercise 13.3. *Show this using Fermat little theorem and the identity* $\left(s^{p-1}\right)^{q-1} = \left(s^{q-1}\right)^{p-1}$)

13.3 RSA

13.3.1 Encryption

Encryption algorithm[2]

- You broadcast (N, e). $N = PQ$ is large semi-prime and e can be a prime, say $e = 3$.
- The factors P, Q are kept secret and you chose e so that it has a inverse mod $(P-1)(Q-1)$. This is automatic if e is prime.
- The message M, an integer, is encrypted as

$$E(M) = M^e \; Mod \; N$$

- $E(M)$ is broadcasted publicly.

Example 13.4. *Suppose I put on my web site the public key* $(N, e) = (187, 3)$ *and keep secret the fact that*

$$187 = \underbrace{11 \times 17}_{secret},$$

You want to transmit to me your secret $KABALA = 137$ *(also the inverse of the fine structure constant). You broadcast the encrypted message*

$$E(137) = 137^3 = 103 \mod 187$$

13.3.2 Decryption

The message M is deciphered with the private key (N, d) which is kept secret. The pair of keys (N, e) and (N, d) are chosen so that (e, d) are modular inverses

$$ed = 1 \; Mod \; (P-1)(Q-1)$$

Note that (e, d) are inverses $Mod \; (P-1)(Q-1)$ not $Mod \; N$.

Using Euclid extended GCD you can efficiently find d even for large N. Since only you know the number $(P-1)(Q-1)$, no one else can figure out the private key (N, d).

The deciphering is done just like the encryption, but with the private key [3]

$$\begin{aligned} D(E(M)) &= \big(E(M)\big)^d \; Mod \; N \\ &= (M^e)^d \; Mod \; N \\ &= M^{ed} \; Mod \; N \\ &= M^{1+k(P-1)(Q-1)} \; Mod \; N, \quad k \in \mathbb{Z} \\ &= M \; Mod \; N \end{aligned}$$

[2] I will use notation where capital letter denote large integers.

[3] The trivial case $M = 0 \mod N$ needs to be dealt with separately.

and Euler-Fermat, Eq. 13.4, was used in the last step [4].

Example 13.5 (Continued). *Continuing the example:*

$$\underbrace{d = 107}_{private\ key}\ ,\ \underbrace{E^d(m) = 103^{107} = 137 \mod 187}_{deciphering}$$

Remark 13.6 (Private-Public keys). *The private and public keys can be interchanged, of course.*

Exercise 13.7 (Toy RSA). *Write a computer program for making RSA with p and q primes with say 30-40 digits.*

- *Mathematica picks random primes p with 30 decimal digits with:* `RandomPrime[{`$10^{30}, 10^{31}$`}]`
- *Choose randomly an odd (prime) e and accept this choice as a public key if $GCD(e, \varphi(N)) = 1$. This will succeed with finite probability.*
- *Compute*[5] *the private key $d = e^{-1}\ Mod\ (P-1)(Q-1)$. d is your secret key for decoding the messages. Keep it safe. d may turn out to be huge.*
- *When you encrypt and decipher use* `PowerMod[m,e,N]`*. If you use instead* `Mod[`s^d`,N]` *with d large you will get an overflow.*

Remark 13.8 (Primality test). *You may worry that the difficulties in making public and private keys are comparable to the difficulties in breaking them. This is a legitimate worry, and the step to worry about is how do you (randomly) choose really large primes p, q? Of course, it is easy to choose a large random integer N, and the probability that by chance you picked a prime is $1/\log N$ is not too small, even for very large N. But, to verify that the number you picked is a prime you need to find its factors, which is the same difficulty as breaking the key. What saves RSA is an efficient test (due to Miller and Rabin) that guarantees primality. If Riemann hypothesis holds the test is sure. If not, it is probabilistic.*

[4] Recall that m is assumed not be divisible by p and q.
[5] Mathematica command: `PowerMod[e,-1,M]`

Chapter 14

Factoring

14.1 Breaking RSA

You break RSA once you gain access to the factors of $N = PQ$. In principle, since N is public, by the unique factorization, P and Q are determined by N. So, if one can factor large semi-primes efficiently, RSA is compromised. This appears to be hard.

14.2 Complexity for pedestrians

14.2.1 Resources

Complexity theory classifies problems according to the growth of resources you need to solve a problem with the number input bits. The resources can be time, space, hardware etc. For example, how many operations you are allowed to make, how much memory you are allocated, etc.

14.2.2 $Poly(n)$

There is a common belief among Computer Scientists (something they call Thesis) that if a problem is solved by one Turing machines in time $Poly(n)$, it will be solved by in time $poly(n)$ by any other reasonable Turing machine.

Let us look at some standard algorithms and how the number of operations scale with the input . Given two numbers of with n digits adding or subtracting them has complexity that is linear in n. Schoolbook long multiplication has complexity of n^2. Amusingly, these are sophisticated multiplication algorithms that scale almost like n.

Operation	Complexity
add two n digits numbers	n
schoolbook multiplication of two n digits numbers	n^2
Schonhagen-Strassen multiplication	$n \log n \log \log n$
GCD	n^2

14.2.3 $Exp(Poly)$

Tasks that diverge exponentially with the number of digits are hard. Here are some examples.

Operation	Complexity
list all integers with n digits	2^n
list all primes with n digits	$(2^n)/n$

14.3 $Poly(n)$ versus $Exp(Poly)$

Complexity deals with the limit $n \to \infty$. A problem in $poly(n)$ that runs in time $2^{100}n$ is more difficult, than one with running time 2^n for $n \leq 100$. However, eventually, when $n \geq 100$ the second becomes harder.

Example 14.1. *Suppose your computer has a clock of GHz and you are willing to complete a task that takes a year. The number of operations of your computer is of order*

$$\underbrace{3 \times 10^7}_{sec\ in\ year} \times 10^9 = 3 \times 10^{16} \approx 2^{55}$$

This is your available resources.

Now consider tasks that scale with the size of the input n like

$$2^{n^\alpha} = \begin{cases} 2^n & \alpha = 1, \quad exponential \\ 2^{\sqrt{n}} & sub - exponential \\ 2^{n^2} & super - exponential \\ n^3 & \alpha = 0, \quad polynomial \end{cases} \tag{14.1}$$

How long an input can you feed in? Assuming that all factors are $O(1)$ is

$$n = \begin{cases} 55 & \alpha = 1, \quad exponential \\ 3025 & sub - exponential \\ 7 & super - exponential \\ 2^{27} & \alpha = 0, \quad polynomial \end{cases} \tag{14.2}$$

For polynomial complexity, you can feed essentially infinitely long inputs. But in the exponential case the input is modest.

14.4 Factoring

Silly factorization of large $N = O(2^n)$ has complexity of

$$\sqrt{N} \times (\log N)^2$$

You test all integers less than $\sqrt{N}$ if they divide N. The cost of division is $(\log N)^2$ and there are $\sqrt{N}$ such tests.

You can do a little better if you use the fact that the number of primes less than N is asymptotically

$$\pi(N) \approx \frac{N}{\times} \log N$$

and it is enough to test primes. This improves the complexity to

$$\sqrt{N} \log N$$

Mathematicians put a lot of effort into factoring. Sophisticated method for factoring large integers have sub exponential complexity, e.g.

$$e^{(\log N)^{1/3}}$$

and are collectively known as number sieve. These can be effectively applied to numbers with about hundred digits. This sets the size of RSA keys.

The basic idea behind several methods of factorization goes back to Fermat. Write

$$N = PQ = \left(\frac{P+Q}{2}\right)^2 - \left(\frac{P-Q}{2}\right)^2 = R^2 - S^2$$

If N is an odd semi-prime, p and q are odd, r and s are integers. Factoring is now reduced to finding integers

$$R^2 = S^2 \mod N, \quad R \pm S \neq 1 \mod N$$

One needs a method to effectively search for R and S.

14.5 Functions that are hard to compute

If you could figure out directly $(P-1)(Q-1)$ from $N = PQ$, then you could efficiently compute the private key d from the public key e:

$$ed = 1 \mod (P-1)(Q-1)$$

This function has name: Euler totient function $\varphi(N)$. It counts the number of divisors of N and has the desired property

$$\varphi(PQ) = (P-1)(Q-1)$$

Unfortunately, Euler totient function is one of those functions that are hard to compute.

Example 14.2. *Certain functions that are easy to invert over the reals are difficult to invert in modular arithmetic. For example the square root and the discrete logarithm are hard to compute.*

For example, with $P = 5$, the square and power are easy to compute but the root and log are hard to compute:

n	0	1	2	3	4	missing in range
n^2	0	1	4	4	1	2,3
2^n	1	2	4	3	1	0
$\sqrt{n}$	0	1, 4	–	–	2, 3	
$\log_2 n$	–	0,4	1	3	2	

14.6 Order

As we shall see, the key to Shor algorithm is that there are functions that are hard to compute on a classical computer, but easy to compute on a quantum computer. One such function that helps in factoring is order finding.

Definition 14.3 (Order)**.** *Given an integer x and a positive integer N with $GCD(x, N) = 1$, the multiplicative order of x modulo N is the smallest positive integer R with*

$$x^R = 1 \mod N$$

Example 14.4. $P = 2 \times 5$

x	1	2	3	4	5	...	order
1	1	1	1	1	...	...	1
2	2	4	8	6	2	...	not defined
3	3	9	7	1	...	...	4
5	5	5	...	...	...	...	not defined
7	7	9	3	1	...	...	4
9	9	1	...	...		...	2

The sequence

$$x, x^2, x^3 \ldots, x^N, \qquad \mod N$$

takes at most N (different) values. If x is invertible, then 0 is not in the list and therefore for some $1 < m \leq N$ one must encounter a (non-zero) value encountered before

$$x^j = x^m \mod N, \quad N \geq \quad j < m \geq 1$$

Since x has a multiplicative (modular) inverse x^{-1}

$$x^{m-j} = 1 \mod N$$

$R = m - j$ is the smallest solution to $x^R = 1$ *Mod* N.

From Euler- Fermat, Eq. 13.4, we have for $N = PQ$ semi-prime

$$x^{(P-1)(Q-1)} = 1 \mod PQ, \quad GCD(x, PQ) = 1$$

It follows that R is a divisor of $(P-1)(Q-1)$.

In the example above where $PQ = (2-1)(5-1) = 4$ the periods are 4 (for $x = 3$) and 2 (for $x = 9$).

Order finding is an arithmetic operation that has no efficient algorithm: to find the order of j you need to make the table j^x with $x = 1, \ldots, N$ and search for the first time 1 shows up in the table.

14.7 Factorization with order finding oracle

If one had on oracle that would determine the order, one could factorize integers. The method goes back to Fermat. Write

$$1 = x^R \; mod N \Longrightarrow (x^R - 1) = 0 \mod N$$

Suppose the period R of x is even then, for $N = PQ$ semi-prime

$$(x^R - 1) = (x^{R/2} - 1)(x^{R/2} + 1) = 0 \mod N \Longrightarrow (x^{R/2} - 1)(x^{R/2} + 1) = k_1 k_2 PQ$$

for some integers $k_{1,2}$. Comparing sides we get the following possibilities

$$(x^{R/2} - 1) = \begin{cases} k_1 & GCD(x^{R/2} - 1, N) = 1 \\ k_1 P & GCD(x^{R/2} - 1, N) = P \\ k_1 Q & GCD(x^{R/2} - 1, N) = Q \\ k_1 PQ & \end{cases}$$

The last line does not happen, since it implies $x^{R/2} = 1 \mod PQ$, but, by assumption R was the shortest period. The first line does not teach us anything, but the second and third lines give us a factor. It turns out that the good case occurs with profitability of $O(1)$.

Remark 14.5 (What can fail)**.** *In the bad cases you did not succeed factoring. (This can all happen as the example below show). The method works because there is a finite success probability. Showing this (you may argue that this is the interesting part of the algorithm) is the business of people who do probabilistic number theory. I will not address this.*

Example 14.6. *In the example with $PQ = 2 \times 5$, $x = 3$ has period $R = 4$ so $x^{R/2} - 1 = 3^2 - 1 = 8$ and $GCD(8, 10) = 2$ is the factor.*

We conclude that one could solve factoring efficiently if one had an effective tool to find the order. But finding an order is finding the period of a periodic function. Finding periods is something that Fourier transform does. So, we conclude that if we had an effective way to compute Fourier transforms, we might be able to crack the factoring problem.

Chapter 15

Fourier

15.1 Fourier transform

This section is a review of Fourier transforms. I use QM notation partly because this is a physics class and you are familiar with it and partly to emphasize that functions may be viewed as vectors in Hilbert space.

To a Physicists the Fourier transform[1]

$$\underbrace{\langle p|\psi\rangle}_{p-rep} = \int dx \langle p|x\rangle\langle x|\psi\rangle = \int \frac{dx}{\sqrt{2\pi}}\, e^{-ipx}\, \underbrace{\langle x|\psi\rangle}_{x-rep}$$

is a map from coordinate space to its dual, the momentum space. But, if you forget about the fact that x and p have different dimensions, or, more precisely change units so that x and p are dimensionless, i.e. $x \mapsto x/\sqrt{\hbar}$ and $p \mapsto p/\sqrt{\hbar}$ you can then identify the two spaces and the Fourier transfrom may be viewed as a unitary change of bases in one Hilbert space. Mathematicians think of course of Hilbert space, $L^2(\mathbb{R})$, as dimensionless.

Fourier is intimately related to shift T_ξ and boosts S_η. They act in coordinate space by

$$\langle x|T_\xi|\psi\rangle = \langle x-\xi|\psi\rangle, \quad \langle x|S_\eta|\psi\rangle = e^{-ix\eta}\langle x|\psi\rangle$$

and in momentum space by

$$\langle p|S_\eta|\psi\rangle = \langle p-\eta|\psi\rangle, \quad \langle p|T_\xi|\psi\rangle = e^{ip\xi}\langle p|\psi\rangle$$

15.2 Discrete FT

A matrix analog of the Fourier transform on an N dimensional vector space is:

[1] In units where $\hbar = 1$.

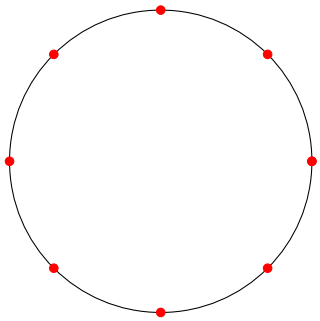


Figure 15.1: The 8-th root of unity

Definition 15.1. *The Fourier transform is defined by*

$$F|j\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega^{jk} |k\rangle, \qquad \omega = e^{2\pi i/N}$$

ω the $N-th$ root of unity and j, k are counted modulo N.

ω^{jk} is the discrete analog of the continuous e^{ipx}.

The simplest example is $N = 2$ then the Fourier transform is simply Hadamard:

$$F = H$$

Exercise 15.2. *Show that with ω the N-th root of unity $\omega^N = 1$,*

$$\sum_{j=0}^{N-1} \omega^{jm} = N\delta_{m,0} \tag{15.1}$$

F_ω is unitary. This follows from writing F in matrix form and observing that its rows are columns form an orthonormal basis.

This follows by inspection as the columns of F are mutually orthogonal and normalized.

Exercise 15.3. *Show that*

$$(F_\omega)^\dagger = (F_\omega)^* = F_{\omega^*}$$

15.3 Discrete translations and boosts

Definition 15.4 (Modular translation)**.** *Define the (modular) translation T by*

$$T|k\rangle = |k+1\rangle,$$

with k counted modulo N.

The shift is diagonalized by the Fourier transforms

$$FT = SF \tag{15.2}$$

where S is the diagonal matrix

$$S|j\rangle = \omega^j |j\rangle$$

Indeed

$$\begin{aligned} FT|j\rangle &= F|j+1\rangle \\ &= \frac{1}{\sqrt{N}} \sum_k \omega^{(j+1)k} |k\rangle \\ &= \frac{1}{\sqrt{N}} \sum_k \omega^{jk} S|k\rangle = \\ &= S\left(\frac{1}{\sqrt{N}} \sum_k \omega^{jk} |k\rangle \right) = \\ &= SF|j\rangle \end{aligned}$$

We shall make much use of this fact.

An amusing property of the Fourier transform is that it square to the inversion, i.e

$$F^2|j\rangle = |-j\rangle$$

Exercise 15.5. *Show this.*

Since inverting twice is the identity we get:

Theorem 15.6 (Spectrum)**.** *The spectrum of the Fourier transform is the set* $\{\pm 1, \pm i\}$.

Each eigenvalue is, of course, highly degenerate when N is large.

15.4 Cost

The computational cost in evaluating the Fourier transform from the definition is large: Each component requires N multiplications and N additions of complex numbers. This involves at least

$$O(N^2) = O\left(2^{2n}\right)$$

operations. A lot.

You can not do better than N because this is the price to write the N amplitudes.

It seems that the computation of Fourier transform is a hard problem, as it scales exponentially with the number of qubit n. How could then Shor ever hope to factor efficiently using Fourier?

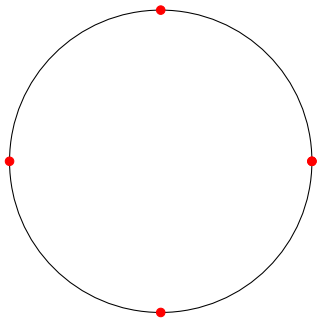


Figure 15.2: The spectrum of Fourier

15.5 Fourier as Spectral analyzer

Suppose we have a vector that has a period R, namely[2]

$$|\psi\rangle = \sum_{j=0}^{N-1} \psi_j |j\rangle, \quad \psi_j = \psi_{j+R},$$

The index j is counted $\mod N$.

Assume that R is a divisor of N. This is a bad assumption to make in general, but we shall make it because the general case is too messy for me.

The shift T operates on the basis vectors by

$$T|j\rangle = |j+1\rangle, \quad j \in \mathbb{Z}_N$$

For a periodic vector

$$|\psi\rangle = T^R|\psi\rangle \Rightarrow \psi_j = \langle j|\psi\rangle = \langle j|T^R|\psi\rangle = \langle j-R|\psi\rangle = \psi_{j-R}$$

Eq. 15.2 implies that

$$FT^R = S^R F$$

Hence for a periodic vector

$$\left|\tilde{\psi}\right\rangle = F|\psi\rangle = S^R F|\psi\rangle = S^R\left|\tilde{\psi}\right\rangle$$

The amplitudes of the Fourier transform then satisfy

$$\tilde{\psi}_j = \langle j|F|\psi\rangle = \langle j|S^R F|\psi\rangle = \omega^{-jR}\langle j|F|\psi\rangle = \omega^{-jR}\tilde{\psi}_j$$

which can be written as

$$(1-\omega^{-jR})\tilde{\psi}_j = 0$$

It follows that all non-zero amplitudes $\tilde{\psi}_j \neq 0$ are for j that satisfy

$$\omega^{jR} = 1 \Longrightarrow jR = 0 \mod N$$

[2]Here N need not be 2^n.

Example 15.7. *Suppose $N = 2^n$ with n large, say 50, and $R = 2^m$ with m large too, say $m = 20$. The non-zero Fourier components are necessarily of the form*

$$j2^m = k2^n \Longrightarrow j = k2^{n-m}$$

with k an integer. This is a very small fraction of all the integers $0, \ldots, N-1$.

If you feed into a Fourier circuit a periodic vector

$$|\psi\rangle = \sum \psi_j |j\rangle \quad \boxed{\mathcal{F}} \quad F|\psi\rangle = \sum \tilde{\psi}_j |j\rangle$$

then measuring the output qubits (always in the computational basis) will select for you only those basis vector $|j\rangle$ for which $\tilde{\psi}_j \neq 0$. This will form a comb of integers:

$$j = k \underbrace{\left(\frac{N}{R}\right)}_{integer}$$

Since, by assumption, R divides N, k must divide j. In particular, $k \leq j$. Write this as

$$R = k \underbrace{\frac{N}{j}}_{known}, \quad k \in 1, \ldots, j-1$$

The period is a multiple of a known integer. If $j \geq 2$, you have learned something useful about the period.

The quantum punch is that you you do not try to compute the $O(N)$ number of amplitudes $\tilde{\psi}_j$. Instead, you collapses the superposition to $|j\rangle$ associated to the large amplitudes. To find this j you only make n on the n qubits. This avoids the classical cost of computing all the components of the Fourier transform which is necessarily larger than N.

15.5.1 What if R does not divide N

In general you do not expect N, a machine property, to be a multiple of R, a problem property. This means that $|\psi\rangle$ is not strictly periodic and N/j need not be an integer, but will be close to an integer if N is large enough.

If you think about this like a physicist then the question is the same as what is the effect of a crystal boundaries on Bragg peaks: They will acquire width. One can make the analysis quantitative, but I would not.

15.6 Quantum Period algorithm

You are given the gate G that computes a periodic function with period R

$$g(x) = g(x+R), \quad x \in \mathbb{Z}_N$$

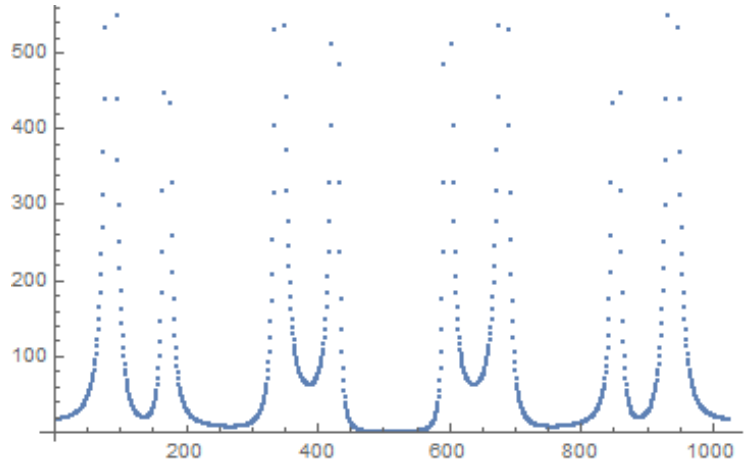


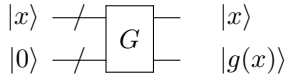
Figure 15.3: This shows the (power spectrum of the) Fourier transform of a^x mod pq for $p = 19$, $q = 13$ and $a = 84$. The period is $R = 12$. The Fourier transform is taken with $n = 10$ bits where $N = 1024$ is not a multiple of $R = 12$. The largest peak (for $j > 1$) is for $j = 86$ and the ratio $N/j = 11.907$ gives a good approximation to the period.

We want to find the period R (a divisor of N).

The standard construction of a function gate is

$$G|x\rangle \otimes |0\rangle = |x\rangle \otimes |g(x)\rangle$$

Note that the first factor takes N values while the second factor takes only N/R values. Pictorially



Now suppose you feed the gate with the democratic superposition in the first factor

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

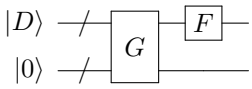
The gate outputs

$$G|D\rangle \otimes |0\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \otimes |g(x)\rangle$$

Now, insert the output of the *first* factor into Fourier gate. This gives

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} F|x\rangle \otimes |g(x)\rangle = \frac{1}{N} \sum_{x,y=0}^{N-1} |y\rangle \otimes |g(x)\rangle \omega^{yx}$$

In picture:



Decompose x into a unit cell and cell number, i.e.

$$x = z + kR, \quad z \in 0, \ldots, R-1, \quad k = 0, \ldots, \lfloor N/R \rfloor - 1$$

We can rearrange the summation accordingly

$$\begin{aligned}\sum_{x,y=0}^{N-1} |y\rangle \otimes |g(x)\rangle \omega^{yx} &= \sum_{y=0}^{N-1} \sum_{z=0}^{R-1} |y\rangle \otimes |g(z)\rangle \left(\sum_{k=0}^{N/R-1} \omega^{-(z+kR)} \right) \\ &= \sum_{y=0}^{N-1} \sum_{z=0}^{R-1} \omega^{-y} |y\rangle \otimes |g(z)\rangle \left(\sum_{k=0}^{N/R-1} \omega^{kyR} \right)\end{aligned}$$

It follows that if you measure the output qubits, the probability to find $|y\rangle \otimes |g(z)\rangle$ is

$$Prob(y) = \frac{1}{N^2} \left| \sum_{k=0}^{N/R-1} \omega^{kyR} \right|^2$$

(The right hand side is aka power spectrum). The geometric series can be summed explicitly

$$\sum_{k=0}^{\lfloor N/R \rfloor - 1} \omega^{kyR} = \frac{1 - \omega^{yR\lfloor N/R \rfloor}}{1 - \omega^{yR}} = \underbrace{\frac{\sin(\pi y R/N \lfloor N/R \rfloor)}{\sin(\pi y R/N)}}_{Interference}$$

The situation is particularly simple if R is a divisor of N. Then the numerator is $\sin \pi y = 0$ for all integer y . A non-zero amplitude for y occurs when also the denominator vanishes, i.e.

$$\omega^{yR} = 1 \Longrightarrow yR = 0 \mod N$$

It follows that N/y is either the period or a factor of the period:

$$R = k\frac{N}{y}$$

Applying this algorithm to the function $a^x \mod PQ$, we have an oracle that allows to find the period R_a and we can therefore break RSA.

Chapter 16

The Quantum Fourier circuit

We have seen that with a quantum circuit that performs Fourier transform we can effectively find periods. The question now arises can we make the Fourier circuit efficiently? Does the number of gates we need to construct Fourier circuit scale polynomially with n? As we shall see we need n^2 gates. The Fouri is efficient indeed.

16.1 The quantum Fourier circuit

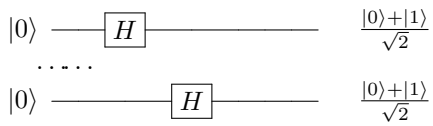
To build the quantum Fourier circuit let start by first building a circuit that will take the input $|0\rangle$ and output $F_\omega|0\rangle$:

$$|0\rangle \mapsto F_\omega|0\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} |j\rangle$$

$F|0\rangle$ is the democratic superposition

$$\begin{aligned} F_\omega|0\rangle &= |+\rangle \otimes |+\rangle \cdots \otimes |+\rangle \\ &= \frac{1}{\sqrt{N}} \Big(|0\rangle + |1\rangle\Big) \otimes \Big(|0\rangle + |1\rangle\Big) \cdots \otimes \Big(|0\rangle + |1\rangle\Big) \end{aligned}$$

This is implement with n single qubits Hadamards



To figure out how F_ω acts on any basis vectors use the relation between shifts and Fourier transform:

$$F_\omega|j\rangle = F_\omega T^j|0\rangle = S^j F_\omega|0\rangle \tag{16.1}$$

S is a diagonal matrix in the computational basis

$$S|j\rangle = \omega^j |j\rangle$$

S factors to action on individual qubits. To see this write the binary representation of j

$$j = j_{n-1}2^{n-1} + \dots j_0 = \underbrace{j_{n-1}\dots j_0}_{binary\ rep}, \quad j_k \in 0, 1$$

and

$$|j\rangle = |j_{n-1}\rangle \otimes \cdots \otimes |j_0\rangle,$$

The action of S factors to actions on qubits:

$$S|j\rangle = \omega^{j_{n-1}2^{n-1}} |j_{n-1}\rangle \otimes \dots \omega^{j_0} |j_0\rangle$$

We may summarize this by

Theorem 16.1 (Quantum Fourier factorization formula)**.**

$$F_\omega|j\rangle = \frac{1}{\sqrt{N}} \left(|0\rangle + \left(\underbrace{\omega^{2^{n-1}}}_{-1} \right)^j |1\rangle \right) \otimes \cdots \otimes \left(|0\rangle + \omega^j |1\rangle \right) \tag{16.2}$$

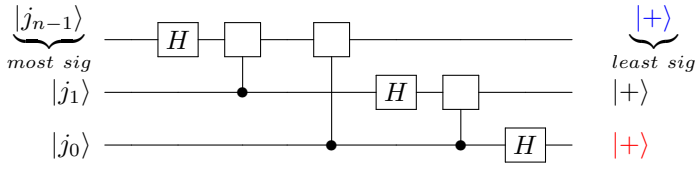
and more explicitly:

$$F_\omega|j_{n-1}\dots j_0\rangle = \frac{1}{\sqrt{N}} \underbrace{\left(|0\rangle + (-1)^{j_0}|1\rangle \right)}_{affected\ only\ by\ j_0} \otimes \cdots \otimes \underbrace{\left(|0\rangle + \omega^{j_{n-1}\dots j_0}|1\rangle \right)}_{affected\ by\ all\ j_k} \tag{16.3}$$

We have succeeded in factoring the Fourier transform so that it acts on the qubits. This formula give exponential gain in complexity since we reduced Fourier to bitwise action.

In some way factorization appears to be too good to be true: We know that just to write down the column vector with the N Fourier coefficient is $O(N)$. The superposition in Eq. 16.3 has only $O(\log N)$ terms. Where have we cheated? The point is that most of the information is lost when you measure: The superposition collapses to one of the computations basis vectors (with the probability determined by the amplitude).

From Eq. 16.3 we see that if we add to the Hadamard control gates that float up, the circuit below will still work for $|0\rangle$.



For the input $|0\rangle$ it does not matter what you put in the (blank) controlled gates since none is operational.

It is clear from Eq. 16.3 that the unitaries will be phase gates related to powers of ω. In the next section we shall see how.

The top line is affected by all qubits, while the bottom is only affected by itself. This reflects the structure of Eq. 16.3.

Note the reorder of the input bits relative to the output bits.

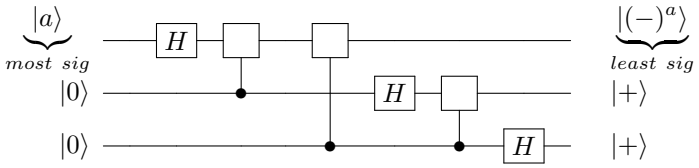
16.2 Filling the blank controls

For $j = 0, N/2$ we have $\omega^j = \pm 1$. From the factorization formula Eq. 16.2 we see that

$$F_\omega \begin{cases} |0\rangle \\ |N/2\rangle \end{cases} = \frac{1}{\sqrt{N}} \Big(|0\rangle + |1\rangle \Big) \otimes \cdots \otimes \underbrace{\Big(|0\rangle \pm |1\rangle \Big)}_{least\ sig\ bit}$$

The action differs just on the last bit.

The circuit below reproduces the equation above:

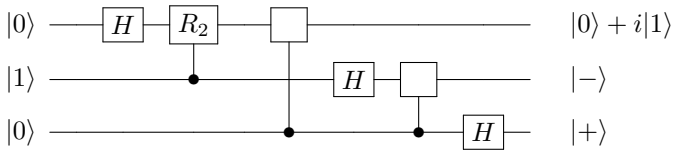


Now that we have the skeleton of the circuit, lets fill in the gates. We can determine the blank conditional gates one by one by feeding in inputs that would switch them one at a time. Let us look at $n = 3$ where $\omega = \sqrt{i}$ and the circuit below.

To determine the first gate, marked R_2, on the left, feed in the state $|010\rangle = |2\rangle$ which makes only this gate operate. From the factorization formula

$$\begin{aligned} \sqrt{8} F|010\rangle &= \big(|0\rangle + i^4|1\rangle\big) \otimes \big(|0\rangle + i^2|1\rangle\big) \otimes \big(|0\rangle + i|1\rangle\big) \\ &= \big(|0\rangle + |1\rangle\big) \otimes \big(|0\rangle - |1\rangle\big) \otimes \big(|0\rangle + i|1\rangle\big) \end{aligned}$$

The quibit in the middle line is not affected by the gate on the top line and indeed transforms correctly (to $|-\rangle$). The top-right qubit should get a relative phase i and this fixes R_2

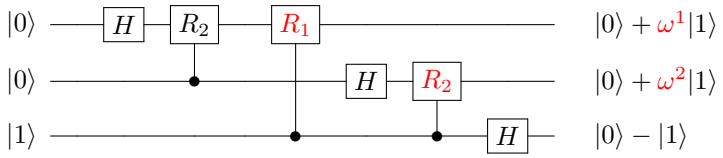


where we introduce the notation

$$-\boxed{R_j}- \quad = \quad \begin{pmatrix} 1 & 0 \\ 0 & \omega^j \end{pmatrix}, \quad \omega = e^{2\pi i/N}$$

For this part it does not matter what the blank controlled gates are.

The remaining two blank gates are determined by considering an input states that activates them, namely $|001\rangle$. The bottom line is automatically satisfied. The top two lines give two equations for two unknowns gates (marekd red).



Remark 16.2 (Ordering and anti-ordering)**.** *Note that the order of bits by significance is opposite on the two sides.*

16.3 Computational cost

When you generalize the construction above to n qubits you see that there are $n-1$ different gates R_j on the top line, $n-2$ on the second etc. In total there are $n(n-1)/2$ gates R_j and n Hadamard. The resources are therefore $O(n^2) = O(\log^2 N)$ gates. The quantum Fourier transform is efficient. It is an exponential improvement on the fast Fourier transform whose cost is $N \log N$ which is itself much better than the simple minded Fourier transform whose cost is N^2. However, as we discussed the QFT is not really useful to determine amplitudes but rather as a a spectral analyzer.

16.4 Phase estimation

QM comes with simple rules how to measure eigenvalues of Hermitian operators. Measuring eigenvalues of unitary operators is measuring phases. It is more complicated and we need to measure interference. Let us see how to do that.

For simplicity, let me assume I have a two qubit, $n = 2$ so I can count integers from $0, \ldots, 3$, and fractions that are multiples of $1/4$. With the fraction $\varphi = 0.\varphi_1\varphi_2$ we associate the integer

$$4\varphi = \underbrace{\varphi_1\varphi_2}_{\text{binary rep}}$$

The phases I can count are then

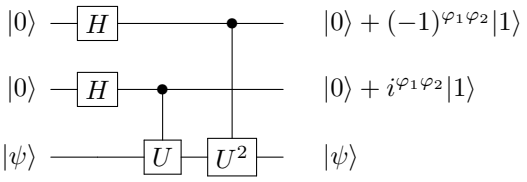
$$e^{i2\pi\varphi} = (e^{i\pi/2})^{4\varphi} = (i)^{\varphi_1\varphi_2}$$

Let $|\psi\rangle$ be an eigenvector of U with eigenvalue φ

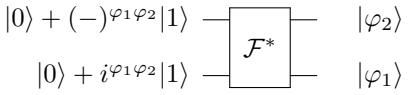
$$U|\psi\rangle = e^{i2\pi\varphi}|\psi\rangle$$

Task: Given the function gates U, U^2, the Fourier gate, and $|\psi\rangle$ construct a gate whose outcome gives φ.

The circuit



Now, if you feed the output to the (inverse) Fourier you will get



A single query of the output gives the binary digits of the phase φ.

16.5 Order finding

You have a gate with n qubits, so you can count up to $N = 2^n$. I also give you a function gate U_a for a modular multiplication

$$U_a|j\rangle = |aj \ ModN\rangle$$

We are interested in finding the period r be the period, i.e. $a^r = 1 \ Mod \ N$. Of course, we need $gcd(a, N) = 1$ for the period to exist. Assume this is the case. We also assume that r is a divisor of N, so Fourier works cleanly. Clearly

$$Spect(U_a) \subset \{\text{r-th roots of unity}\}$$

Now, with n registers, we can hope to identify periods that are shorter than $N/2$. This means that each eigenvalue will be at least two-fold degenerate. The shorter the period, the larger the degeneracy.

16.5.1 Bloch states

Let U_a be a modular multiplication with $a \in 1, \ldots, N-1$ with $gcd(a, N) = 1$

$$U_a|j\rangle = |aj \ ModN\rangle$$

and let r be the period, i.e. $a^r = 1 \ Mod \ N$, with r a divisor of N, then for $k \in 0, \ldots, r-1$

$$|\psi_k\rangle = \frac{1}{\sqrt{r}} \sum_{j=0}^{N-1} e^{i2\pi kj/r} U_a^j |1\rangle$$

are eigenstate of U_a with eigenvalues $e^{-2\pi ik/r}$. Conversely

$$\frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} |\psi_k\rangle = |1\rangle$$

Proof: The first part follows from

$$\begin{aligned}\sqrt{r}U_a|\psi_k\rangle &= \sum_{j=0}^{N-1} e^{i2\pi kj/r}U_a^{j+1}|1\rangle \\ &= \sum_{j=1}^{N} e^{i2\pi k(j-1)/r}U_a^{j}|1\rangle \\ &= e^{-2\pi ik/r}\sum_{j=1}^{N} e^{i2\pi kj/r}U_a^{j}|1\rangle \\ &= \sqrt{r}e^{-2\pi ik/r}|\psi_k\rangle\end{aligned}$$

The converse part follows from the fact about sums of roots of the identity.

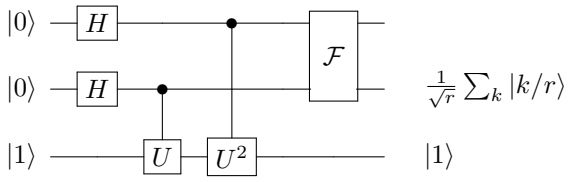
16.5.2 Circuit

It follows that if I give you U_a and its powers $(U_a)^2$, $(U_a)^4$ etc and I also give you $|\psi_k\rangle$ and a Fourier gate, phase estimation gives the for a measurement in the computational basis the fraction k/r.

Now. because of the converse piece, I do not even need to give you $|\psi_k\rangle$. You just construct for yourself $|1\rangle$. This gives you a superposition of $|\psi_k\rangle$ and the outcome of the circuit will be a superposition of r basis vectors in the computational basis so that each vector points at the fraction k/r:

$$\frac{1}{\sqrt{r}}\sum_k |k/r\rangle$$

and by $|k/r\rangle$ we mean a basis vector in the computational basis corresponding to the phase k/r. A measurement will collapse the superposition on the computational basis. It is a collapse on one of the terms k. The result of a query of the circuit will be one of the phases k/r. This gives information on the period r.



Chapter 17

Entropy and information

17.1 Shannon

In 1948 Claude Shannon launched the information age in a ground breaking paper "A Mathematical Theory of Communication". He formulated precise notions that allow to describe information quantitatively. This allowed him to solve two big issues in information theory:

- Recovery from errors
- Data compression.

To do that we need to quantifying the notion of information content of strings of n bits.

17.2 Kolmogorov Complexity

Kolmogorv gave a conceptually satisfactory definition of the information content of a list x, now known as Kolmogorov Complexity, $K(x)$

$$K(x) = \texttt{Length of the shortest algorithm that generates } x$$

For example, the complexity of an integer N is

$$K(N) = \begin{cases} O(\log n) & N = 2^n \\ O(\log N) & otherwise \end{cases}$$

The algorithm gives the binary digits of N (or n in the first case).

The problem with Kolmogorov complexity is that there is no algorithm to compute the complexity for a given list.

17.3 Shannon entropy

Suppose we get a list x from a bank of possible lists $\{x\}$. How much information do we gain when we get x? The information we gain reflects our ignorance before receiving x. The larger the bank, the larger our ignorance the more information the message transmitted. The ignorance is a property of the bank of lists, and the probability distribution P that assigns the probability $p(x)$ to the event x, and Shannon called it entropy:

Definition 17.1 (Shanon entropy)**.** *The entropy*

$$H(P) = -\sum_x p(x) \log p(x) \geq 0, \quad 1 \geq (x) \geq 0, \quad 0 \log 0 = 0 \tag{17.1}$$

We follow the tradition in communication theory and take log in base 2. ln stands for the natural log.

There is a story that when Shannon was developing these ideas he was in a search for a good name for the information content and consulted with von Neumann who suggested the entropy for two reasons, first because of the similarity to entropy in statistical mechanics and second because, so said von Neumann, nobody really understands what entropy really means, so Shannon would have the upper hand in an argument.

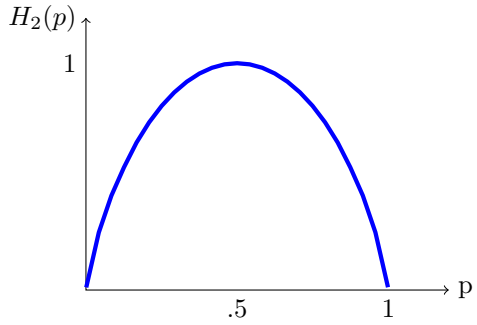


Figure 17.1: $H_2(p)$

Example 17.2. *The pool has N lists.*

- *Suppose the probability of receiving the list x is the same $p(x) = \frac{1}{N}$. The entropy is*

$$H = -N \times \frac{1}{N} \log \frac{1}{N} = \log N$$

The information you got with a list is the number of bits needed to encode all the lists.

- *Suppose*

$$p(x) = \begin{cases} 1 & x = x_0 \\ 0 & otherwise \end{cases}$$

Now, the entropy vanishes

$$H = 1 \log 1 + 0 \log 0 = 0$$

You have learned nothing when you received the message.

Entropy captures your initial ignorance and the gained knowledge when the event happened.

Definition 17.3. *For a binary random variable with $P_X = \{p, 1-p\}$ we define $H_2(p)$ for $0 \leq p \leq 1$ by*

$$H_2(p) = -p \log p - (1-p) \log(1-p)$$

Proposition 17.4. *$H_2(p)$ is a concave function of p, symmetric about $p = 1/2$ and $H_2(1/2) = 1$.*

Example 17.5. *Prof. X flunks every student and Prof. Y passes every one. In either case, if you know which prof gave the grade, there is no information in the grade itself. At least not about the student. But, if you do not know which Professor gave the grade then there is information in the grade, if not about the student, but about the professor.*

The entropy of two isolated glasses of water is the sum of their entropies. The corresponding fact in Shannon is: If x and y are independent random variables, i.e.

$$p(x, y) = p(x)q(y)$$

then Shannon entropy is additive

$$H\big(P_{x,y}\big) = H(P) + H(Q)$$

17.3.1 Typical sequences

There are 2^n different lists obtained by n coin tosses. Now consider the list made from biased coin tosses, with probability p for head and $q = 1 - p$ for tail. There are still 2^n different lists, however, many of these are extremely rare. When n is large, a typical long list will have m heads near the expected number, namely

$$m \approx np$$

The ingenious idea of Shannon was to focus on typical lists. The number of typical is

$$\begin{aligned}\binom{n}{m} &= \frac{n!}{m!(n-m)!} \underbrace{\approx}_{Stirling} \frac{n^n}{m^m(n-m)^{n-m}} \\ &= \left(\frac{n}{m}\right)^m \left(\frac{n}{n-m}\right)^{n-m} \\ &\approx \frac{1}{\sqrt{2\pi n\, p\, q}}\, p^{-np} q^{-nq}\end{aligned}$$

The set of typical lists is exponentially large with n and the rate of growth is $H_2(p)$. Since $H_2(p) \leq H_2(1/2) = 1$ an unbiased coin gives the set with exponentially (in n) more members than any other set.

This says that there is more information in a list of n bits that comes from an unbiased coin than from a list that comes from a biased coin.

The Shannon entropy counts the size of the pool of lists. To see this let us compute H. The probability $p(x)$ for list x with m heads is

$$p(x) = p^m q^{n-m}. \quad m = \#\text{heads} \in x$$

and there are $\binom{n}{m}$ such lists.

The Shannon entropy is

$$\begin{aligned}H &= -\sum_x p(x) \log p(x) \\ &= -\sum_x p^m q^{n-m} (m \log p + (n-m) \log q) \\ &= -\log p \left(\sum_x m p^m q^{n-m}\right) - \log q \left(\sum_\ell (n-m) p^m q^{n-m}\right) \\ &= -\log p \underbrace{\left(\sum_m m p^m q^{n-m} \binom{n}{m}\right)}_{\text{Expectation of m}=np} - \log q \underbrace{\left(\sum_m (n-m) p^m q^{n-m} \binom{n}{m}\right)}_{\text{expectation of } n-m=nq}\end{aligned}$$

The sums in the brackets can be evaluate explicitly

$$\begin{aligned}H &= -\langle m \rangle \log p - \langle n-m \rangle \log q \\ &= -n\big(p \log p + q \log q\big) \\ &= n H_2(p)\end{aligned}$$

We see that Shanon entropy measures the logarithm of the set of typical sequences.

The rate of information gain with any additional bit depends on p and is given by $H_2(p) \leq 1$. Since the information per bit is $H_2(p) \leq 1$, a message of length n bits can, at least in principle, be compressed to a message whose length is shorter $n \times H_2(p)$. The compressed list always looks like a list from a fair coin. This is basically Shanon compression theorem.

17.4 Relative entropy

Relative entropy is a notion of distance between probability distributions. It is always non-negative, and vanishes when two distributions coincide. Unfortunately, it is not a bona-fide distance wince it is a-symmetric:

Definition 17.6. *$H(P||Q)$, the relative entropy of two probability distributions, P and Q, defined on the same set of events labeled by x, is defined by*

$$H\big(P||Q\big) = \sum_x p(x) \log \left(\frac{p(x)}{q(x)} \right) \tag{17.2}$$

The basic property of the relative entropy is

$$H\big(P||Q\big) \geq 0$$

To prove this we need the elementary inequality

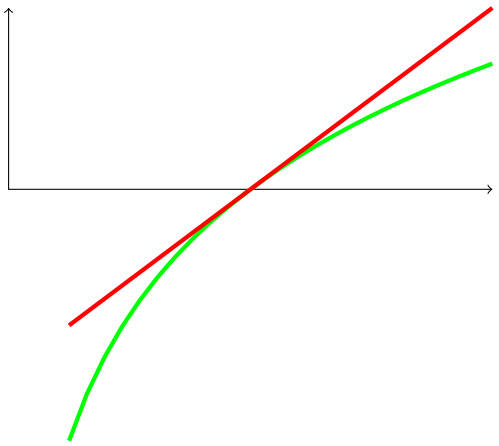


Figure 17.2: $\ln x$ and $x - 1$

$$\ln x \leq x - 1 \Longrightarrow \log_2 x \leq (\log_2 e)(x - 1) \tag{17.3}$$

The positivity of the relative entropy now follows easily:

$$\begin{aligned} -H\big(P||Q\big) &= -\sum_x p(x)\log\left(\frac{p(x)}{q(x)}\right) \\ &= \sum p(x)\log\left(\frac{q(x)}{p(x)}\right) \\ &\le \log e \sum p(x)\left(-1+\frac{q(x)}{p(x)}\right) \\ &= \log e \sum \big(p(x)-q(x)\big) \\ &= 0 \end{aligned}$$

Example 17.7. *What is the probability that a long sequence of length $n \gg 1$ will appear as if it was taken as if the unbiased coin if the coin is actually biased with $\{p, 1-p\}$?*

The probability of finding $m = n/2$ in a sequence taken from an biased coin is

$$\begin{aligned} \binom{n}{n/2} p^{n/2} q^{n/2} &\approx 2^n 2^{(n/2)(\log p + \log q)} \\ &= 2^{n\left(1+\frac{1}{2}\log p + \frac{1}{2}\log q\right)} \\ &= 2^{-nH(1/2||p)} \end{aligned}$$

where we have used

$$H(1/2||p) = -\frac{1}{2}\log\frac{1}{2p} - \frac{1}{2}\log\frac{1}{2(1-p)} = 1 + \frac{1}{2}\log p + \frac{1}{2}\log(1-p)$$

You see that making such an error has an exponentially small probability, and the rate is the relative entropy.

17.5 Convexity

Perhaps the most important property of the entropy is that it is a concave function.

A function $h(x)$ is concave if for $0 \le \lambda \le 1$

$$h\big(\lambda x + (1-\lambda)y\big) \ge \lambda h(x) + (1-\lambda)h(y)$$

This is also called Jensen inequality. In particular, $-x\log x$ is a concave functions (see figure).

Theorem 17.8. *The Shannon entropy $H\big(P\big)$ is a* concave *function of the probability distribution $p(x)$. This means that the entropy of a mixture of two probabilities distributions P and Q (on the same set of events) is larger than the mixture of entropies:*

$$H\big(\lambda P + (1-\lambda)Q\big) \ge \lambda H(P) + (1-\lambda)H(Q)$$

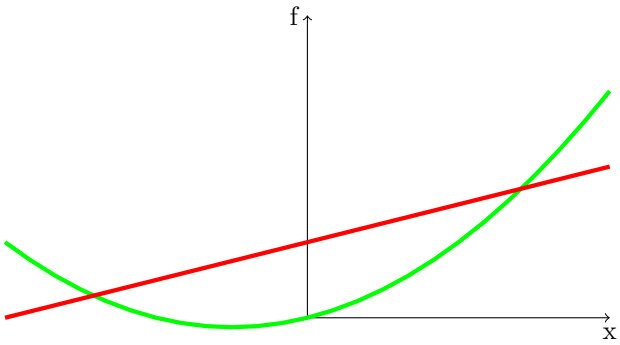


Figure 17.3: A convex function $f(x)$: The weighted sum of the values (at the intersections) lie above the value of the weighted sum.

This is a consequence of the concavity of $-x \log x$ and the fact that the sum of concave functions is a concave function.

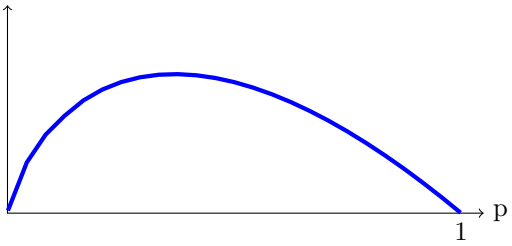


Figure 17.4: The graph of $-x \log_2 x$ for $0 \leq x \leq 1$. The graph is not symmetric and its maximum is at $1/e$.

Remark 17.9. *In thermodynamics the concavity of the entropy is related to irreversibility. When you mix 1/2 a glass of cold water with 1/2 glass of hot water, you get a glass of tepid water, whose entropy is larger than the entropy of the two half glasses. The second law says that the entropy of isolated systems can only increase. This makes mixing irreversible.*

Theorem 17.10. *Suppose the sample space has N events then*

$$0 \leq H \leq \log N \tag{17.4}$$

The lhs is obvious. The rhs can be seen as follows. Let P_π be the probability distribution obtained from P by permuting the the points in the sample space, i.e.

$$p(x) = p\left(\pi(x)\right)$$

Clearly

$$H(P) = H(P_\pi)$$

There are $N!$ such permutations. Consider mixing them all. This gives the uniform distribution $p_j = 1/N$ with entropy $\log N$. By Jensen

$$\log N = H\,(i.i.d) \geq \sum_\pi \frac{1}{N!} H(P_\pi) = H(P)$$

17.5.1 Monotonicity and Sub-additivity

We have seen that the Shannon entropy of independent random variables is additive. Since correlations decrease entropy we expect Shannon entropy to be sub-additive

$$H(P_{x,y}) \leq H(P_x) + H(P_y) \tag{17.5}$$

To see this use 17.3

$$\begin{aligned} H(P_{x,y}) - H(P_x) - H(P_y) &= \sum p(x,y) \log \frac{p_x(x) p_y(y)}{p(x,y)} \\ &\leq \log_2 e \sum p(x,y) \left(\frac{p_x(x) p_y(y}{p(x,y)} - 1 \right) \\ &= 0 \end{aligned} \tag{17.6}$$

Entropy is monotonic in the sense that

$$H(P_{x,y}) \geq H(P_x) \tag{17.7}$$

This follows from

$$\begin{aligned} H(P_{x,y}) - H(P_x) &= -\sum_{xy} p(x,y) \log p(x,y) + \sum_x p(x) \log p(x) \\ &= -\sum_{xy} p(x,y) \log p(x,y) - p(x,y) \log p(x) \\ &= -\sum_{xy} p(x,y) \log \left(\frac{p(x,y)}{p(x)} \right) \\ &= -\sum_{xy} p(x,y) \log \left(p(y|x) \right) \geq 0 \end{aligned}$$

17.6 Mutual information

Definition 17.11. *Mutual information is defined by*

$$H(X:Y) = H(X) + H(Y) - H(X,Y) \tag{17.8}$$

By sub-additivity

$$H(X:Y) \geq 0$$

To see what this means consider

- Encryption with a fixed key: X is an n bit message, taken from an ensemble which is uniformly distributed over all n-bit messages. The encryption of X is

  $$Y = E(X) = X \oplus k$$

  with a fixed n bit key k. Then, since a message x and the fixed key k fix a unique $y = x \oplus k$ we have

  $$H(X) = H(Y) = n, \quad H(X,Y) = n, \quad H(X:Y) = n$$

  The encryption with a fixed key has as much information as the unencrypted message.

- Encryption with a random uniformly distributed key: Now

  $$H(X) = H(Y) = n, \quad H(X,Y) = 2n, \quad H(X:Y) = 0$$

  There is no mutual information between the encryption and the original message.

Chapter 18

von Neuman entropy

A quantum state ρ can be viewed as the non-commutative analog of a probability distribution

$$\rho \Longleftrightarrow P$$

In particular, since ρ is positive with unit trace its eigenvalues ρ_j may be interpreted as probabilities.

Definition 18.1 (von Neumann entropy)**.** *The von Neumann entropy of a quantum state ρ is the Shannon entropy of its eigenvalues:*

$$S(\rho) = -Tr(\rho \log \rho) = -\sum \rho_j \log \rho_j = H(\rho_j)$$

where the ρ_j are the eigenvalues of ρ.

A pure state is a one dimensional projection $\rho_j \in (0, 1)$ and hence has zero entropy

$$S\big(|\psi\rangle\langle\psi|\big) = 0$$

The zero entropy expresses the fact that we have complete knowledge of the quantum state.

The entropy of a single qubit in the state $\rho = \frac{\mathbb{1}+\mathbf{n}\cdot\sigma}{2}$ is

$$S(\rho) = H_2\left(\frac{1+|\mathbf{n}|}{2}\right)$$

is a monotonically decreasing, concave function of $|\mathbf{n}|$. The fully mixed state has maximal entropy, 1.

Exercise 18.2. *Show that mixing two qubits always increases the entropy.*

The upper and lower bound on the Shannon entropy transfer to von Neumann

$$\underbrace{0}_{pure} \leq S(\rho) \leq \underbrace{\log \dim \mathcal{H}}_{max\ \ mixed} \tag{18.1}$$

The fully mixed state maximizes the entropy and saturates the upper bound.

We have talked about the fact that Hilbert space is big. You do not see this in the entropy. n qubits do not have more entropy as n bits:

The von Neumann entropy of a state of n qubits is bounded by n.

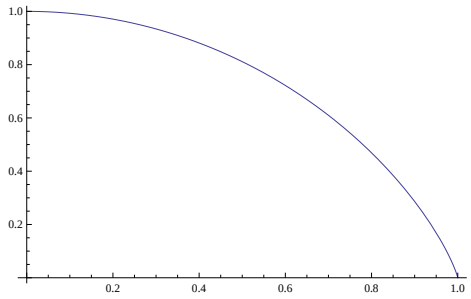


Figure 18.1: The von Neuman entropy of a qubit $S(\rho)$, drawn as function of $|n|$, the distance from the center of the Bloch ball.

The state $\rho_A \otimes \rho_B$ is the quantum analog of independent random variables. In particular

$$Spec(\rho_A \otimes \rho_B) = \{(\rho_A)_j(\rho_B)_j\}$$

The additivity of Shannon implies that the entropy of a tensor product is additive:

$$S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$$

18.1 Convexity

The notions of relative entropy and mutual information and the properties such as convexity and sub-additivity have natural generalization to von Neumann, but the proofs are more difficult.

Theorem 18.3 (Löwener Heinze)**.** *The von Neuman entropy is a concave function of* ρ

$$S\big(p\rho_1 + (1-p)\rho_2\big) \geq pS(\rho_1) + (1-p)S(\rho_2), \quad 0 \leq p \leq 1$$

18.1.1 Klein inequality

The relative entropy is positive

$$S(\rho\|\sigma) = Tr\rho\log\rho - Tr\rho\log\sigma \geq 0$$

To prove this we need something more fancy than Eq. 17.3. It is still true that

$$\ln\rho \leq \rho - 1$$

but the inquality that hold in the comutative case

$$\underbrace{\rho(\ln\sigma - \ln\rho)}_{non-hermitian} \geq \rho - \sigma$$

does not even make sense in the non-commutative case: the left hand side is not even hermitian.

Theorem 18.4 (Klein)**.** *Suppose $f(x)$ is convex and A, B Hermitian matrices. Then*

$$Tr\big(f(B) - f(A)\big) \geq Tr\big((B-A)f'(A)\big)$$

A short proof is in Wikipedia.

Let us use Klein to prove positivity of relative entropy. Take $f(x) = x\ln x$, which is convex.

$$f'(x) = 1 + \ln x$$

and then

$$Tr\rho\ln\rho - Tr\sigma\ln\sigma \geq Tr(\rho-\sigma)(1+\ln\sigma) = tr(\rho-\sigma)\ln\sigma$$

which is the result we wanted.

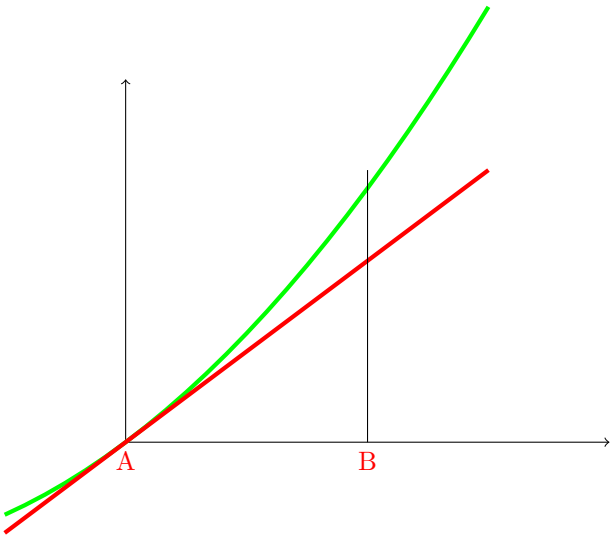


Figure 18.2: Klein

18.2 Quantum statistical mechanics

von Neumann definition of entropy coincides with the entropy of Gibbs states in equilibrium statistical mechanics. The quantum version of Gibbs principle

fixes the density matrix at thermal equilibrium

$$\rho(\beta) = \frac{e^{-\beta H}}{Z(\beta)}, \qquad \underbrace{Z(\beta) = Tr\, e^{-\beta H}}_{quantum\ partition\ function} \quad , \quad \beta = 1/k_B T, \quad k_B = 1$$

H is the quantum Hamiltonian[1] A cool name for β is coolness.

From von-Neumann

$$\begin{aligned} S(\rho) &= -Tr\left(\frac{e^{-\beta H}}{Z}\log\left(\frac{e^{-\beta H}}{Z}\right)\right) \\ &= \beta Tr\left(H\frac{e^{-\beta H}}{Z}\right) + Tr\left(\frac{e^{-\beta H}}{Z}\right)\log Z \\ &= \beta\left\langle E\right\rangle + \log Z \end{aligned}$$

This can be rearranged as

$$\underbrace{F = -T\log Z}_{\text{free energy}} = \left\langle E\right\rangle - TS$$

which is one of the possible definitions of entropy in thermodynamics.

Exercise 18.5. *Show the thermodynamic identity*

$$S = -\frac{\partial F}{\partial T}$$

F is a concave function of T.

Exercise 18.6. *Show that*

$$\frac{d^2F}{dT^2} = -\beta^3\left(\left\langle E^2\right\rangle - \left\langle E\right\rangle^2\right) \leq 0, \quad \left\langle E^k\right\rangle = \frac{Tr\, H^k e^{-\beta H}}{Z(\beta)}$$

Example 18.7. *The density matrix for the Harmonic oscillator in thermal equilibrium with $\gamma = \beta\omega$ is*

$$\rho = p_n|n\rangle\langle n|, \quad p_n = (1-e^{-\gamma})e^{-\gamma n} \quad n = 0, \ldots, \infty$$

The von Neumann entropy can be computed explicitly by summing geometric series and one finds

$$S(\rho) = -\sum p_n \log p_n = (q+1)\log(q+1) - q\log q, \quad q = \frac{1}{e^{\gamma}-1}$$

[1] H needed to be bounded below for the trace to make sense.

18.3 The growth of entropy

The second law of thermodynamics says that the entropy of an isolated system will evolve towards a maximum. It grows. This is what happens when you let cold and hot water mix.

In quantum mechanics the evolution of an isolated system is unitary. The von Neumann entropy is unitary invariant:

$$S(\rho) = S(U\rho U^\dagger)$$

The entropy is a constant of motion.

Moreover, from the definition it is clear that the relative entropy is unitary invariant

$$S(\rho\|\sigma) = S(U\rho U^\dagger\|U\sigma U^\dagger) \tag{18.2}$$

In particular, if $\rho = \rho_{T_h} \otimes \rho_{T_c}$ represents two isolated teacups at different temperatures at σ the thermal equilibrium of the two teacups after they are allowed to interact, then

$$U\sigma U^\dagger\sigma$$

and we expect approach to equilibrium, i.e.

$$S(\rho\|\sigma) \geq S(U\rho U^\dagger\|\sigma)$$

But, this is not what Eq. 18.2 says.

It seems as though we can prove results that are in conflict with common experience. Or alternatively that the notion of entropy in thermodynamics and von Neuman entropy are not quite the same.

This problem is not unique to QM. The same problem arises in the SM description of entropy. In SM entropy measures the volume in phase space of thin energy shell. If the system evolves with a classical Hamiltonian, then Liouville theorem says that the volume is conserved. So, in classical SM entropy does not seem to grow either. In fact, if we wait long enough any finite system will come back close to its initial state.

Entropy grows if we allow for course graining. For example, the state $\rho_t = U_t\rho U_t^\dagger$ can be rapidly change in time. If we do not resolve time accurately, a candidate for a course grained state is a time average, say

$$\bar{\rho} = \frac{1}{N}\sum_{j=1}^{N} U_j\rho U_j^\dagger$$

By concavity of the entropy

$$S(\bar{\rho}) \geq \sum \frac{1}{N} S(U_j\rho U_j^\dagger) = S(\rho)$$

Similar ideas can be used to show approach to equilibrium. Forgetting a subsystems decreases the relative entropy:

$$S(\rho_{A\otimes B}\|\sigma_{A\otimes B}) \geq S(\rho_A\|\sigma_A)$$

For a proof see e.g. Nielsen and Chuang. It is a difficult result as it depends on strong sub-additivity.

18.4 Entanglement entropy

Suppose that the system and its bath are in a pure state $|\psi\rangle$. By Schmidt

$$|\psi\rangle_{SB} = \sum \sqrt{p_j}|j\rangle_S \otimes |j\rangle_B$$

and

$$\rho_S = \sum p_j |j\rangle_S\langle j|, \quad \rho_B = \sum p_j |j\rangle_B\langle j|$$

The entropy of the system and the bath are equal and are given by the Shannon entropy of the Schmidt coefficients:

$$S(\rho_{SB}) = 0, \quad S(\rho_A) = S(\rho_B) = -\sum p_j \log p_j$$

The entropy is non-zero, if and only if, the system and the bath are entangled. This is the entanglement entropy. In a quantum system the whole may have less entropy than its parts.

Exercise 18.8. *Suppose the system is made of n qubits and the bath with $m > n$ qubits. Show that the entanglement entropy is bounded above by n.*

Quantum information is peculiar in that we may know everything about the entire system, say the state of a Bell pair, so the entropy vanishes. But, yet we may know very little about its subsystems: The entropy of Alice qubit is maximal. A theorem I shll not prove says that

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

18.5 Area law

For a system of n qubits we have

$$S(\rho) \leq n$$

It can be shown that for treasonable Hamiltonians the entropy of thermal syayes is extensive:

$$S(\rho^(\beta)) = ns(T, V)$$

In contrast, in the theory of Black holes a fundamental discovery of Bekenstein is that

$$S_{bh} = \frac{A}{4}$$

where A is the area of the horizon.

This raises the intriguing possibility that Black hole entropy may be identified as some kind of entanglement entropy.

18.6 Entanglement entropy of a one dimensional chain

Consider a chain where of $n+2$ qubits labeled by $j = 0, \ldots, n+1$. Let

$$H_j = Z_{j-1} \otimes X_j \otimes Z_{j+1}, \quad j \in 1, \ldots, n$$

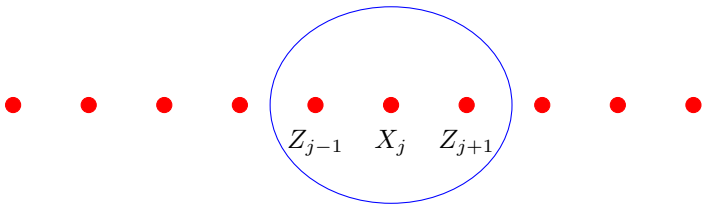
It is easy to see that all the H_+j commute. Clearly

$$[H_j, H_k] = 0$$

if $|j-k| \geq 2$ and if $j = k$. In the case $j - k = \pm 1$

$$[H_j, H_{j+1}] = [Z_{j-1}X_jZ_{j+1}, Z_jX_{j+1}Z_{j+2}] = Z_{j-1}\underbrace{[X_jZ_{j+1}, Z_jX_{j+1}]}_{=0}Z_{j+2}$$

For the sake of concreteness let us impose boundary conditions so the edges 0



and $n+1$ are at $|0\rangle$ and consider the Hamiltonian

$$H = \sum_{j=1}^{n} H_j$$

Since

$$Spect(H_j) = \pm 1$$

the ground state has energy $-n$ and the top state n.

Let us find the ground state. Denote

$$|j\rangle = |a_0 \ldots a_{n+1}\rangle, \quad a_j \in 0, 1$$

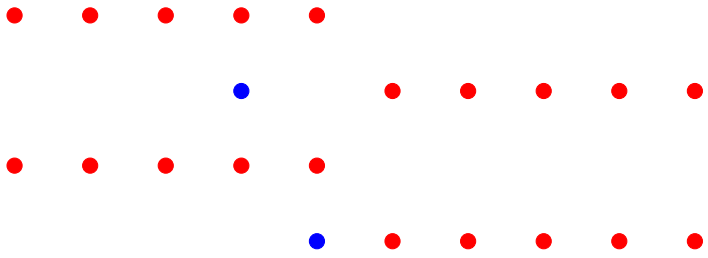
a basis vector in the computational basis. The

$$H_j|a_0 \ldots a_j \ldots a_{n+1}\rangle = (-)^{a_{j-1}+a_{j+1}}|a_0 \ldots a_j \oplus 1 \ldots a_{n+1}\rangle$$

Let us look at a picture. There are two cases shown in the two pictures below. Here the number of jumps in the configuration increased by two, and the overall sign remains. In the second case shown below the number of jumps stays the same but the overall phase changes This means that an eigenvector of H_j with eigenvalue -1 is

$$|\Psi\rangle = \frac{1}{2^{n/2}} \sum (-)^{k(j)}|j\rangle$$

where $k(j)$ is the number of up jumps in the configuration j.

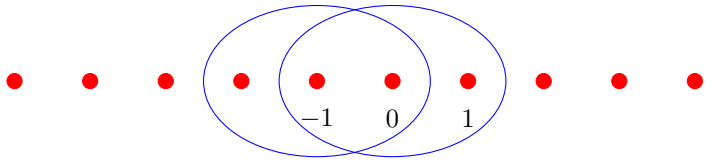


18.6.1 Schmidt decomposition

We can now make the Schmidt decomposition by hand. The bi-bipartite decompositions of the ground state has 16 terms

$$|\Psi\rangle = \frac{1}{2^2} \sum_{a,b,c,d \in 01} \pm \underbrace{|L\rangle \otimes |ab\rangle}_{bi-partite} \otimes \underbrace{|cd\rangle \otimes |R\rangle}_{bi-partite}$$

The Schmidt decomposition minimizes the number of terms so that there are only 4 terms.



For example, tracking the signs due to jumps we see that the four terms

$$|0\rangle|0\rangle - |0\rangle|1\rangle - |1\rangle|0\rangle + |1\rangle|1\rangle = 2|+\rangle|-\rangle$$

There are four such terms. It follows that the Schmidt coefficients are all equal, and there are 4 of them. The entanglement entropy is 2 independent of the size of the chain. This is an example of the area law.

Chapter 19

Introduction to error correction

19.1 Shannon and error correction

Error correction works on redundancy. Suppose a classical communication channel introduces errors at rate p small so the input bit $a \in 0, 1$ exits as mixture

$$a \mapsto \begin{cases} a & \text{with probability } q \\ 1 \oplus a & \text{with probability } p = 1 - p \end{cases}$$

The probability for getting the wrong answer is proportional to p.

You can try and improve the odds be representing the logical $a \in 0, 1$ by, say, 3 copies. Then

$$a^3 \mapsto \begin{cases} a^3 & \text{with probability } q^3 \\ \big((1 \oplus a)a^2\big) \ \vee \ \big(a(1 \oplus a)a\big) \ \vee \ \big(a^2(1 \oplus a)\big) & \text{with probability } 3pq^2 \\ \big((1 \oplus a)^2 a\big) \ \vee \ \big(a(1 \oplus a)^2\big) \ \vee \ \big((1 \oplus a)a(1 \oplus a)\big) & \text{with probability } 3p^2 q \\ (1 \oplus a)^3 & \text{with probability } p^3 \end{cases}$$

If you use the majority rule to correct the output you see that

$$a^3 \mapsto \begin{cases} a^3 & \text{with probability } q^2(1 + 2p) \\ (1 \oplus a)^3 & \text{with probability } p^2(1 + 2q) \end{cases}$$

The probability of getting the wrong answer is proportional to p^2. If $p < 1/2$ you are doing better.

This method is not good if you want to correct a long string. If you send m such triplets, the probability that the majority rule fixes all the errors is:

$$Prob(\text{majority fixes all errors}) = \big(q^2(1 + 2p)\big)^m$$

The term in the bracket, being a probability, is bounded by one, and takes the value 1 for $p = 0$. So for $p > 0$ he probability to fix all errors decays exponentially with the length of the chain. This method is, therefore not good enough for communication.

It is one of the great discoveries of Shannon that provided the mutual information between the source and the output of the channel is positive, then, then one can find an encoding (which will appropriately inflate the message from n bits to $const \times n$ bits) so that with probability close to 1, for chains that are arbitrarily long, all errors can be recovered.

19.1.1 Shannon noisy channel coding theorem

I will not do justice to Shannon theory but rather sketch the basic idea.

A message of n bits can be identified with a corner of the n-dimensional unit cube. The origin

$$(0 \dots 0)$$

has n neighbors all at (Hamming) distance 1:

$$(10 \dots 0), (01 \dots 0), \dots, (0 \dots 01)$$

It has $\binom{n}{2}$ neighbors at distance 2 and $\binom{n}{m}$ neighbors at distance m. Most of the neighbors are at a distance $n/2$.

Consider the neighbors that are at distances $m \leq n/3$. There are about $2^{nH_2(1/3)}$ of these. Remove all these points from the cube. This removes $2^{nH_2(1/3)}$ out of 2^n, a small fraction.

Now pick any of the remaining vertexes and repeat the procedure by removing all vertexes that are at a distance less than $n/3$. This will removes at most $2^{nH_2(1/3)}$ additional vertexes.

You can proceed in this manner $2^{n(1-H_2(1/3)}$ times pruning the 2^n vertexes of the unit cube until you are left with

$$2^{n(1-H_2(1/3))}$$

vertexes that are all at a distance at least $n/3$ from each other. This is you code space.

The different messages in the code space can be encoded in $m < n$ bits where

$$m = n(1 - H_2(1/3))$$

The encoding of the m bits in $n > m$ bits gives protection. Indeed all the points in the coding space messages are separated one from the other by a distance of at least $n/3$. If the error rate per bit is $p < 1/3$ then a long message m will, on the average, acquire pn errors. With $pn < n/3$ the error can not mode from one code word to another. You can then identify what is the original message uniquely.

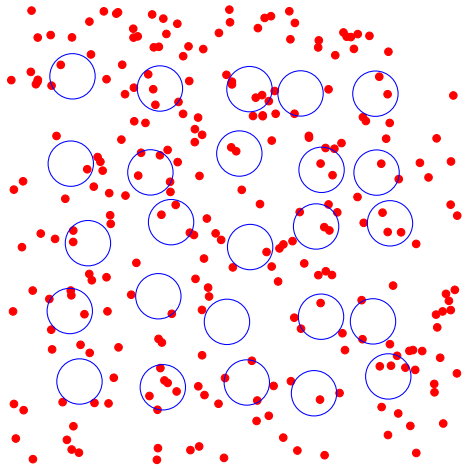


Figure 19.1: The red dots are all 2^n messages. The blue circles represent the the np neighborhoods of the code space. Since the blue circles do not overlap, you can encode an $n(1-H_2(1/3))$ bit sequence in the codes space in such a way that you can correct np errors.

19.2 Quantum error correction

Suppose we have a quantum channel that introduces errors. A bit flip error is

$$|a\rangle \mapsto \begin{cases} |a\rangle & \text{with probability } q \\ X|a\rangle & \text{with probability } p = 1-p \end{cases}$$

A phase flip error is

$$|a\rangle \mapsto \begin{cases} |a\rangle & \text{with probability } q \\ Z|a\rangle & \text{with probability } p = 1-p \end{cases}$$

and a general error may be

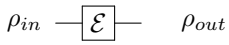
$$|a\rangle \mapsto \begin{cases} |a\rangle & \text{with probability } q \\ U|a\rangle & \text{with probability } p = 1-p \end{cases}$$

with an arbitrary unitary. Recall that (up to an overall phase) any unitary acting on a qubit can be written as

$$U = a_0\mathbb{1} + ia_1X + ia_2\underbrace{Y}_{iXZ} + ia_3Z, \quad \sum a_\mu^2 = 1, \quad a_\mu \in \mathbb{R}$$

This suggests that if we can get rid of X and Z errors we are done.

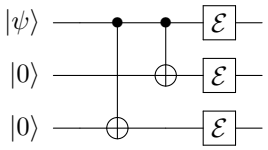
In a picture:



and in a formula

$$\rho_{in} = |\psi\rangle\langle\psi| \xrightarrow[\text{corrupt}]{\mathcal{E}} q\rho_{in} + pU\rho_{in}U^\dagger$$

If the error can occur on any qubit the picture is:



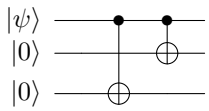
The task is to identify and correct the error without destroying the interference.

19.3 Bit flip

QM does not allow cloning but allows for cloning in the computational basis. We can encode the logical bits 0_L and 1_L by multiple identical qubits

$$0_L \mapsto |0\rangle^{\otimes 3}, \quad 1_L \mapsto |1\rangle^{\otimes 3}$$

The encoding can be accomplished by



The circuit encoded $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ as

$$|\psi\rangle \otimes |0\rangle \otimes |0\rangle \mapsto \alpha \underbrace{|000\rangle}_{|\mathbf{0}_L\rangle} + \beta \underbrace{|111\rangle}_{|\mathbf{1}_L\rangle}$$

The uncorrupted Hilbert space where the qubit is encoded is

$$\mathcal{H}_0 = Span\{|000\rangle, |111\rangle\} \tag{19.1}$$

We need $\mathcal{H}_0$ to be a Hilbert space because we need to protect superpositions.

Suppose for example that a bit flip-error occurred in the j-th qubit. This is represented by a unitary map acting on the uncorrupted space

$$\mathcal{H}_0 \mapsto X_j\mathcal{H}_0$$

Each one of these 4 two dimensional spaces is a linear space. The choice of $\mathcal{H}_0$ implies that all these spaces are mutually orthogonal:

$$\mathcal{H}_0 \perp X_j\mathcal{H}_0 \perp X_k\mathcal{H}_0, \quad j \neq k$$

Together, they span the 8 dimensional Hilbert space of 3 qubits.

Now, if we know that an error X_j occurred we can easily correct for it

$$\mathcal{H}_0 \xrightarrow[\text{corrupt}]{X_j} \mathcal{H}_0 \xrightarrow[\text{correct}]{X_j} \mathcal{H}_0$$

without corrupting the superposition. For example:

$$\alpha|000\rangle + \beta|111\rangle \xrightarrow[\text{corrupt}]{X_1} \alpha|100\rangle + \beta|011\rangle \xrightarrow[\text{correct}]{X_1} \alpha|000\rangle + \beta|111\rangle$$

We shall assume that at most one error occurred.

More generally, if the error U can occur in any one of the qubits

$$\rho_{in} = |\psi\rangle\langle\psi| \xrightarrow[\text{corrupt}]{\mathcal{E}} (1-p)\rho_{in} + p\sum_{j=1}^{3} U_j \rho_{in} U_j^\dagger$$

19.4 Non demolition and error syndromes

We need a measurement that would allow us to determine if and what error occurred in such a way that it does not cause a collapse in $\mathcal{H}_0$.

Now

$$Z_1Z_2, \quad Z_2Z_3, \quad Z_1Z_3$$

is a stabilizer for $\mathcal{H}_0$: Every vector in $\mathcal{H}_0$ is an eigenvector of Z_jZ_k with eigenvalue 1:

$$Z_jZ_k|\psi\rangle = |\psi\rangle, \quad |\psi\rangle = \alpha|000\rangle + \beta|111\rangle$$

This means that if we measure Z_jZ_k the measurement is non-demolition in $\mathcal{H}_0$.

The three measurements are mutually commuting and so can be performed in any order.

Since

$$Z_jZ_kX_\ell = \begin{cases} X_\ell Z_j Z_k & \text{if } \ell \neq j,k \\ -X_\ell Z_j Z_k & \text{otherwise} \end{cases}$$

A measurement of Z_jZ_k of any vector in the corrupted spaces $X_\ell\mathcal{H}_0$ is also a non-demolition measurement, only that now some of the eigenvalues are -1.

For example with $|\psi\rangle = \alpha|000\rangle + \beta|111\rangle$

$$Z_1Z_2X_1|\psi\rangle = Z_3Z_1X_1|\psi\rangle = -Z_2Z_3X_1|\psi\rangle = -X_1|\psi\rangle$$

Z_jZ_k are called the error syndrome.

19.5 Recovery from continuous errors

One reason for the success of digital computers is error corrections. No one knows how to error correct analog (classical) computers: You can not use the

majority rule because every computer drift independently and the N computers will give you N different results. The best you can do is average.

Since qubits live in the Broch sphere, you may worry that they may drift independently and you will not be able to error correct. But at the same time we know that a von Neumann measurement of a qubit yields only yes and no. In this respect a qubit behaves like a discrete bit. As we shall now see it is indeed the collapse of the wave function that allows to correct for a continuum of errors.

Suppose any one of the 3 qubits can be affected by an error of a continuous rotation about the X axis:

$$|a\rangle \mapsto \begin{cases} |a\rangle & \text{with probability } q \\ (\mathbb{1}\cos\theta + iX\sin\theta)|a\rangle & \text{with probability } p = 1-q \end{cases}$$

For the sake of concreteness suppose the mistake occurred on the first qubit.

This means that $|\psi\rangle = \alpha|000\rangle + \beta|111\rangle$ is mapped to

$$|\psi\rangle \mapsto \begin{cases} |\psi\rangle & \text{with probability } q \\ (\mathbb{1}_1\cos\theta + iX_1\sin\theta)|\psi\rangle & \text{with probability } p = 1-q \end{cases}$$

I could write this also as a map of density matrices:

$$|\psi\rangle\langle\psi| \xrightarrow[\text{corrupt}]{U_1} q|\psi\rangle\langle\psi| + p(\mathbb{1}_1\cos\theta + iX_1\sin\theta)|\psi\rangle\langle\psi|(\mathbb{1}_1\cos\theta + iX_1\sin\theta)$$

Now, suppose you measure the syndrome Z_1Z_2. This projects the state on the spectral subspaces of Z_1Z_2. In particular,

$$(\mathbb{1}_1\cos\theta + iX_1\sin\theta)|\psi\rangle \xrightarrow[\text{measure}]{Z_1Z_2} \begin{cases} |\psi\rangle & \text{if } Z_1Z_2 = 1 \\ X_1|\psi\rangle & \text{if } Z_1Z_2 = -1 \end{cases}$$

In the first case, $Z_1Z_2 = 1$, we recovered from the error. In the second case, $Z_1Z_2 = -1$, we identified the error exactly and can recover by the unitary gate X_1.

19.6 Phase flip error

Suppose now that we know that the channel does not bit flip, but it may phase flip, with small probability. For example

$$\alpha|000\rangle + \beta|111\rangle \xrightarrow[\text{corrupt}]{Z_1} \alpha|000\rangle - \beta|111\rangle \xrightarrow[\text{recovery}]{Z_1} \alpha|000\rangle + \beta|111\rangle$$

Let us first address the question how to identify which qubit suffered a flip.

Recall that H unitarily interchanges X and Z

$$HZ = XH$$

It follows that the bif-flip error syndromes are

$$H^{\otimes 3}(Z_1Z_2)H^{\otimes 3} = X_1X_2, \quad X_2X_3, \quad X_1X_3$$

The syndromes act as the identity on $\mathcal{H}_0$, and therefore do not demolish the state.

We can translate all the results about bit-flips to phase flips with the dictionary $X \leftrightarrow Z$.

19.7 5 qubits suffice

We are now faced with the bigger task: We want to identify and correct in a non-demolition fashion, bit-flips or phase flips one any qubit.

To do that we need more qubits than 3. Let us see how many. Let $\mathcal{H}_0$ be the 2-dimensional Hilbert space of the logical qubits. Suppose this is encoded with n qubits.

There are n bit flip and n phase flip errors

$$X_j, \quad Z_j \quad j \in 1, \ldots, n$$

If we want to have a syndrome that would uniquely identify the error we need the spaces

$$X_j\mathcal{H}_0, \quad Z_j\mathcal{H}_0$$

to be mutually orthogonal. They also need to fit in the big Hilbert space, i.e

$$2(1+2n) \leq 2^n$$

The smallest n that satisfies this is $n = 5$.

In fact, with 5 qubits you can correct three errors, X_j, Y_j, Z_j, since

$$2(1+3n) = 2^n\Big|_{n=5} = 32$$

This is, indeed, the minimal number of qubits needed to give a full protection to a single logical qubit.

19.8 The Shor code

The Shor code uses 9 qubits to encode a single logical qubit. The space $\mathcal{H}_0$ is spanned by

$$|0_L\rangle = \big(|000\rangle + |111\rangle\big) \otimes \big(|000\rangle + |111\rangle\big) \otimes \big(|000\rangle + |111\rangle\big)$$

and

$$|1_L\rangle = \big(|000\rangle - |111\rangle\big) \otimes \big(|000\rangle - |111\rangle\big) \otimes \big(|000\rangle - |111\rangle\big)$$

The syndrome for bit flips are organized by triplets. There are nine of them

$$\begin{aligned} S_3 &= Z_1 Z_2, \quad S_1 = Z_2 Z_3, \quad S_2 = Z_3 Z_1, \\ S_6 &= Z_4 Z_5, \quad S_4 = Z_5 Z_6, \quad S_5 = Z_6 Z_4, \\ S_9 &= Z_7 Z_8, \quad S_7 = Z_8 Z_9, \quad S_8 = Z_7 Z_9, \end{aligned}$$

Clearly they commute and leave the logical qubits invariant.

We need additonal 3 syndromes to detect a phase flip in the first bracket, or the second bracket, or the third. This has the syndromes[1]

$$\begin{aligned} S'_3 &= (X_1 X_2 X_3)(X_4 X_5 X_6) \\ S'_1 &= (X_4 X_5 X_6)(X_7 X_8 X_9), \\ S'_2 &= (X_1 X_2 X_3)(X_7 X_8 X_9) \end{aligned}$$

The two sets of syndromes commute as you can see by looking at e.g.

$$[Z_1 Z_2, X_1 X_2 X_3] = 0$$

The price for error correction is that we are storing the two logical bits in a 2^9 dimensional Hilbert space. The eigenspaces of the syndromes allow us to decompose the space into orthogonal pieces where the possible errors lie.

The logical bits sit in the "error free" subspace, which is the range of

$$\mathcal{P}_0 = \prod_{j=1}^{9} \frac{\mathbb{1} + S_j}{2} \prod_{k=1}^{3} \frac{\mathbb{1} + S'_k}{2}$$

We can now repeat the arguments we had in the case of single bit-flip error, to show that we can recover from any error in the qubit.

[1] This does not identify which of the qubits in the triplet flipped. But it is also no important. Pick one.