Lecture notes for the course:
Quantum Information

given by Dr. Benni Reznik
Spring 2003, Tel-Aviv University
written by Amir Seginer

last modified: March 3rd 2005

Caveat: These lecture notes were written while I was studying for the final exam in the quantum information course. I have gone over some of the parts again, but not over all. As a result, the notes are both missing in some places and plain wrong in others (I just don't know where). If you find a mistake, or have any comments, please let us know.[1]

Note: These notes are based mostly on notes I took during the lectures given, but were also supplemented, in some places, by other material. Most of this extra material was taken either from the book by A. Peres [1], or from the online lecture notes of J. Preskill [2] (see references at the end).

[1] Send emails to reznik@post.tau.ac.il and to aseginer@post.tau.ac.il.

Contents

Part 1

Introduction

CHAPTER 1

Physics and Information

1.1. Maxwell's demon (classical)

Assume two adjoining rooms A and B, each filled with a gas at equal temperature T. Now, imagine also a small demon[1] who control's a shutter between the rooms. The shutter is assumed to be ideal, so that no work is produced by opening and closing it. The demon opens and closes the shutter so as to allow fast particle to pass from room B to room A, and allow slow particles to pass in the opposite direction, from room A to room B. As a result, at the end of the process, we have faster gas particles in room A and slower ones in B, and therefore

$$T_A > T_B$$

without any work being done.

Now, from thermodynamics we know that for quasi-static processes the entropy S obeys[2]

$$dS = \frac{đQ}{T}.$$

Since the system of the two rooms is a closed one, then the heat $đQ_A$ going into room A must come from room B so that

$$đQ_A = -đQ_B.$$

Thus, the change in entropy ΔS of the whole system (both rooms) is

$$\Delta S = \int \left(\frac{đQ_A}{T_A} + \frac{đQ_B}{T_B} \right) = \int \left(\frac{đQ_A}{T_A} - \frac{đQ_A}{T_B} \right) < 0,$$

which is negative since $đQ_A > 0$ (the fast energetic particles are going into room A) and $T_A \geq T_B$. We have therefore managed to reduce the entropy of a closed system without doing any work (the opening and closing of the shutter requires no work), thus contradicting the second law of thermodynamics.

1.1.1. The Szilard model. One way out of the paradox is to say that in order to decide when to open and close the shutter, one must know which particle is approaching the shutter, and thus we have a connection between entropy and information (see also the end of this section). As an example of this let us study the *Szilard model* (1929).

In the Szilard model one assumes a small cell with a single particle in it. If the particle is in the left-hand side of the cell, we say that the cell is in state "0", and if the particle is in the right-hand side, we say the cell is in state "1". We can thus code information in a sequence of such cells, each cell in one of the binary states 0 or 1.

Now, in order to keep the particle on the left-hand side of the cell (or on the right-hand), we must put a barrier in the middle. However, before we put the barrier the particle may be anywhere in the cell. Thus, in order to force it to be on one side, we have to push a piston from one side of the cell (until we reach the middle). Pushing such a piston *isothermally* requires work to be done, and so we connect work and information.

[1] This paradox was suggested by Maxwell and is therefore called "Maxwell's demon paradox".

[2] Recall that dS is an exact differential, while $đQ$ is an inexact one. Inexact differential means that the change in the heat Q depends not only on the initial and final physical (macroscopical) states of the system (defined by pressure, volume, temperature, ...), but also on the path taken from one state to the other.

To find the work done by the piston when moving (isothermally) to the middle of the cell, we assume for the moment that the cell is filled with an ideal gas. The force acting on the piston, by the gas, equals the piston's area A times the gas pressure P. The change in volume of the cell, when the piston moves a distance δx is $|\delta V| = |A\delta x|$. Thus the work δW done (on the system, by the piston) when the volume of the cell is changed by δV (due to the piston's movement) is[3]

$$\delta W = F\delta x = -P\delta V.$$

We have assumed that the gas we use is an ideal gas, so it obeys

$$PV = Nk_BT,$$

where N is the number of particles in the gas and k_B is Boltzmann's constant. Thus the work done by the piston may be written as

$$W = -\int PdV = -\int_{V_1}^{V_2} \frac{Nk_BT}{V} dV = Nk_BT \ln\frac{V_1}{V_2}.$$

Now, since the temperature was unchanged during the process, the velocity distribution of the gas particles has not changed either (the internal energy of the system has not changed). The question is therefore where has the work-energy gone to? It has gone to heating the heat bath surrounding our system.

As we just saw, the internal energy U of our gas has not changed. We know from thermodynamics that we may write the internal energy U as

$$U = F + TS,$$

or in differential form

$$\delta U = \delta F + T\delta S,$$

but also

$$\delta U = W + \delta Q = W + T\delta S,$$

where F is the free energy of the system, W is the work done on the system, and δQ is the heat which entered the system. Since in our case the temperature T is constant, and we have $\Delta U = 0$, then we must have[4]

$$\Delta F = W = -T\Delta S.$$

Thus, from the result we had for the work done, we may write

$$\Delta F = Nk_BT \ln\frac{V_1}{V_2}.$$

$$\Rightarrow \Delta S = -Nk_B \ln\frac{V_1}{V_2}$$

If we now return to the case of our cell, having a single particle ($N = 1$) confined to half of the cell ($V_2 = \frac{1}{2}V_1$), we find that

$$\Delta F = k_BT \ln 2,$$

and

$$\Delta S = -k_B \ln 2.$$

This last result should not be surprising, since entropy may also be defined as

$$S = -k_B \ln \Omega,$$

[3] Note the minus sign. When the piston is pushed (doing work on the system), the volume of the cell is reduced, and so δV is negative.

[4] Since for $\delta U = 0$

$$0 = \delta F + T\delta S = W + T\delta S.$$

where Ω is the total number of possible states of the system. In our case the system can be in one of two states ("0" left-hand side, or "1" - right-hand side), which gives the above result.

1.1.2. The Landauer principle.

The Szilard model has shown us that there is connection between information and energy/work. Landauer used this connection to give a lower bound on the energy expenditure needed for performing a computation.

The *Landauer principle* says that in order to *erase* information we must expend energy which then goes into heating the environment. We shall show that this leads to a lower bound on the energy expenditure for performing a computation.

We shall first examine why it requires energy to erase information. For this we start with a Szilard cell. Assume that we are given a cell which has a particle either on the left or on the right (we don't know where). We shall say that the cell is erased, if the particle is (for certain) on the left-hand side.[5] A method of achieving this, is to take the barrier out of the cell and then push our piston half way from the right, thus confining our particle to the left half. As we have already seen, the work done in pushing the piston, when done isothermally, goes to heating the environment by $\Delta Q = k_B T \ln 2$. Thus, we see that the process of erasing information causes the heating of the surroundings.

Now, if we look at logical gates in a computer, they are schematically described as irreversible process in which two bits of information go in, while only *one* comes out. Thus, in the *irreversible* process of a logic gate we have necessarily erased one bit, which requires an energy of at least $k_B T \ln 2$.[6] We have therefore found a lower bound for the energy expenditure for doing a calculation. Note, that in today's computers the energy expenditure ($\sim 10^8 k_B T$ per bit) is much higher than Landauer's lower bound.

1.1.3. Bennett's reversible computer.

It has been emphasized that the lower bound given by Landauer is only good for irreversible gates. Bennett (1973) has shown that if one uses reversible gates, one may construct a computer which requires no energy expenditure at all. In Bennett's computer a gate still accepts two bits as input, however (unlike before), the output is also two bits: One bit, is the logical result we wanted (from the gate) while the second bit (together with the first) allows us to find the initial input bits.[7]

Although, such a gate gives us superfluous information for the calculation, it does allow us to reverse the process. Now, during the computation, using reversible gates, we shall not erase any cells and therefore no energy will be wasted. However, in order to make a different calculation (after the first) we must reuse our cells which means erasing them, and thus seemingly returning to Landauer's principle. But, as you recall we used reversible gates, therefore we can write down the result at the end of the first computation and then reverse the process of computation. This reverse computation will bring us back to the initial conditions with no net energy expenditure. The cells in their initial condition can then be used for our next calculation, and so we have built a computer which requires no energy.[8]

Having found a connection between storage of information and entropy/energy, we can now return to the Maxwell's demon paradox. Bennett (1982) suggested a resolution between the demon paradox and the second law of thermodynamics. The resolution is that

[5] We use this definition of erasure since we are assuming that the computer has a limited amount of memory. It therefore has to recycle its bits, which means erasing them as we defined here. To manipulate a cell we must first know in which state it is, and we know this for the erased cells.

[6] After passing through the gate, we no longer know the state of one of the cells. In order to reuse this cell (our computer has a finite amount of memory cells), we must erase it and thus waste energy.

[7] Since we have two bits of input, and two bits of output, then we can code the input in the output (for logic operators such as "and" and "or").

[8] Note that the cells in the initial conditions are all in known states, either "0" or "1", but known to us. With this information we can construct any other initial conditions by flipping the necessary cells. The process of flipping requires no work; we don't push a piston we simply flip the whole cell.

every time the demon opens and shuts the shutter he is actually performing a computation (he is performing an “if” statement which can be broken into logical gates). a computation means that he needs memory bits. Assuming that the number of bits is finite, the demon will have to erase them and thus the demon will give rise to work and entropy (although the shutter itself requires no work to operate it). This entropy will ensure that the second law of thermodynamics is upheld.

1.2. Quantum information

In the previous section we studied information using classical objects. We now wish to introduce information theory using quantum objects. The following table compares the classical and quantum manifestations of the main points of importance in information theory:

	Classical	Quantum
basic information unit	bit: $\{0,1\}$	qubit: $\alpha|0\rangle+\beta|1\rangle$ (superposition principle)
dynamics	deterministic (causal)	deterministic (unitary evolution)
measurements	do not influence system	effect the system (uncertainty principle + collapse)

We shall see that the superposition principle and the different effects of measurements will cause the quantum theory of information to display very different traits from those of the classical theory.

1.2.1. the qubit. In the classical case, the basic unit of information we used was the bit, which could accept either the value “0” or the value “1”. In the quantum case, the basic unit we use is a two state system.[9] We shall generally denote the two states as $|0\rangle$ and $|1\rangle$,[10] however, due to the superposition principle, the general state of such a system is

$$|\psi\rangle=\alpha|0\rangle+\beta|1\rangle \quad (\langle\psi|\psi\rangle=1\Rightarrow|\alpha|^2+|\beta|^2=1).$$

Since α and β are complex numbers, they are each described by two parameters (real and imaginary parts) which gives us four parameters describing the state $|\psi\rangle$. However, we also have the requirement $|\alpha|^2+|\beta|^2=1$ (due to the normalization $\langle\psi|\psi\rangle=1$), which reduces us to just three continuous parameters. Of these parameters, one is the global phase of the system which has no physical importance. Thus we are left with just two (physical) continuous parameters for describing $|\psi\rangle$.

One method of writing $|\psi\rangle$ with two parameters is

$$|\psi\rangle=\cos\frac{\theta}{2}e^{-i\frac{\phi}{2}}|0\rangle+\sin\frac{\theta}{2}e^{+i\frac{\phi}{2}}|1\rangle.$$

Since we have two continuous parameters, one might think that we can use a single qubit to store an infinite amount of information (unlike the classical bit which can store only 0 or 1). This is indeed true, we can store in a qubit an infinite amount of information, however Holevo (1961) has shown that we can extract from a qubit (with 100% certainty) a maximum of only one bit of information. Thus, for all practical reasons we can store in a qubit only a single bit of information.

[9] The simplest non-trivial Hilbert space is a two dimensional one.

[10] The two state system can be any kind of system with two orthonormal states. For example, it can be a spin $\frac{1}{2}$ system with the two states $|\uparrow\rangle$ and $|\downarrow\rangle$, or a system with two energy states $|E_0\rangle$ and $|E_1\rangle$.

1.2.2. no-cloning theorem. As we noted above, one cannot extract more then one bit of information from a qubit. In spite of this let us now try. Assume two qubits: the first we shall denote as $|\nearrow\rangle_\theta$

$$|\nearrow\rangle_\theta = \cos\frac{\theta}{2}e^{-i\frac{\phi}{2}}|0\rangle + \sin\frac{\theta}{2}e^{+i\frac{\theta}{2}}|1\rangle$$

and the second will simply be the spin up qubit

$$|\uparrow\rangle = |0\rangle.$$

These two states together give

$$\langle\uparrow|\nearrow\rangle_\theta = e^{-i\frac{\phi}{2}}\cos\frac{\theta}{2},$$

so that the probability of measuring spin-up for a state $|\nearrow\rangle_\theta$ is $\cos^2\frac{\theta}{2}$. If we could now make many such measurements, then according to the statistics of our measurement we could deduce θ up to any accuracy. Thus, apparently we can encode in a qubit a continuous parameter and then extract it (to any desired precision).

The problem with the previous scheme, is that in order to perform a multiple number of measurements, we must first replicate, or clone, our initial state $|\nearrow\rangle_\theta$ while we do not know what it is. Only then (after cloning) can we do the measurements and determine θ. The problem is that in quantum mechanics we cannot clone (unknown states). This is called the no-cloning theorem.

PROOF. The proof of the no-cloning theorem rests on the fact that the evolution of a quantum state must be described by a unitary operator.[11] In order to clone our particle N times we must start with N particles in a known state, which we shall denote as $|0\rangle$. Thus, our initial state before cloning starts, is

$$|\Psi_i\rangle = |0\rangle|0\rangle\cdots|0\rangle|\psi\rangle.$$

At the end of the process we want to have a state

$$|\Psi_f\rangle = U|0\rangle|0\rangle\cdots|0\rangle|\psi\rangle = |\psi\rangle|\psi\rangle\cdots|\psi\rangle.$$

Now, assume that we have found such an operator U, which we use on two states $|\psi^{(1)}\rangle$ and $|\psi^{(2)}\rangle$:

$$|\Psi_f^{(1)}\rangle = U|\Psi_i^{(1)}\rangle = U|0\rangle|0\rangle\cdots|0\rangle|\psi^{(1)}\rangle = |\psi^{(1)}\rangle|\psi^{(1)}\rangle\cdots|\psi^{(1)}\rangle,$$

$$|\Psi_f^{(2)}\rangle = U|\Psi_i^{(2)}\rangle = U|0\rangle|0\rangle\cdots|0\rangle|\psi^{(2)}\rangle = |\psi^{(2)}\rangle|\psi^{(2)}\rangle\cdots|\psi^{(2)}\rangle.$$

Since the operator U it is unitary ($U^\dagger = U^{-1}$) then necessarily

$$\langle\Psi_f^{(1)}|\Psi_f^{(2)}\rangle = \langle\Psi_i^{(1)}|U^\dagger U|\Psi_i^{(2)}\rangle = \langle\Psi_i^{(1)}|\Psi_i^{(2)}\rangle.$$

However, by definition

$$\langle\Psi_i^{(1)}|\Psi_i^{(2)}\rangle = \Big(\langle\psi^{(1)}|\langle 0|\cdots\langle 0|\Big)\Big(|0\rangle\cdots|0\rangle|\psi^{(2)}\rangle\Big) = (\langle 0|0\rangle)^N\langle\psi^{(1)}|\psi^{(2)}\rangle = \langle\psi^{(1)}|\psi^{(2)}\rangle,$$

while

$$\langle\Psi_f^{(1)}|\Psi_f^{(2)}\rangle = (\langle\psi^{(1)}|\cdots\langle\psi^{(1)}|\langle\psi^{(1)}|)(|\psi^{(2)}\rangle\cdots|\psi^{(2)}\rangle|\psi^{(2)}\rangle) = (\langle\psi^{(1)}|\psi^{(2)}\rangle)^{N+1}.$$

[11] Recall that the Hamiltonian in quantum mechanics must be Hermitian ($H^\dagger = H$). The (time) evolution operator is then $U(t) = e^{-\frac{i}{\hbar}Ht}$:

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}Ht}|\psi(t=0)\rangle = U(t)|\psi(t=0)\rangle,$$

which is necessarily unitarian ($U^\dagger = U^{-1}$). Note, that this is all true, assuming that the Hamiltonian is time independent . If the Hamiltonian is time dependent, then the evolution operator is $e^{-\frac{i}{\hbar}\int Hdt}$, which is also Unitary.

Thus, if there exists a unitary cloning operator U, then we must have (since we need $\langle\Psi_f^{(1)}|\Psi_f^{(2)}\rangle = \langle\Psi_i^{(1)}|\Psi_i^{(2)}\rangle$) that for any two states

$$\left(\langle\psi^{(1)}|\psi^{(2)}\rangle\right)^{N+1} = \langle\psi^{(1)}|\psi^{(2)}\rangle.$$

This is certainly not true for *any* two states, and therefore there cannot exist a cloning operator. □

Please note, however, that if we choose an orthonormal basis, we can create a unitary operator which clones the elements of the basis, but not their linear combinations.[12]

1.2.3. Bit vs. qubit. Although we can extract from a qubit only one bit of information, the qubit is not equivalent to a classical bit. For example, assume that we are given the integral

$$\int_0^1 f(t)dt = n\alpha,$$

where we know $f(t)$ and α, and we know that n (an integer) is either even or odd. Now, in order to find whether n is even or odd, classically we require an infinite number of bits, since t is continuous, and we need an infinite number of bits to describe a continuum (to calculate the integral numerically). However, if we use qubits, it suffices to use just a single qubit to find whether n is even or not.

To solve the problem quantum mechanically we take a spin "up" in the x direction $|\uparrow\rangle_x$, and construct a Hamiltonian

$$H(t) = \lambda f(t)S_z = \lambda f(t)\frac{1}{2}\hbar\sigma_z,$$

where σ_z is one of the Pauli matrices

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and[13]

$$\sigma_z|\uparrow\rangle_x = |\downarrow\rangle_x.$$

[12] Such an operator for two particles could be

$$U = \sum_i |i\rangle|i\rangle\langle i|\langle 0| + \sum_{\substack{i,j \\ j\neq i,0}} |j\rangle|i\rangle\langle i|\langle j| + \sum_i |0\rangle|i\rangle\langle i|\langle i|,$$

which gives

$$U|0\rangle|i\rangle = |i\rangle|i\rangle$$

and

$$UU^\dagger = \sum_i |j\rangle|i\rangle\langle i|\langle j| = \mathbb{1}.$$

[13] Recall that in the z basis

$$|\uparrow\rangle_x = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_z + |\downarrow\rangle_z\right),$$

$$|\downarrow\rangle_x = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_z - |\downarrow\rangle_z\right).$$

The evolution of the spin $|\uparrow\rangle_x$ is then given by[14]

$$\begin{aligned} U(t)|\uparrow\rangle_x &= e^{-\frac{i\lambda}{2}\int_{t'=0}^{t} Hdt\sigma_z}|\uparrow\rangle_x = e^{-\frac{i\lambda}{2}n\alpha\sigma_z}|\uparrow\rangle_x \\ &= (\cos\frac{\lambda n\alpha}{2} - i\sigma_z\sin\frac{\lambda n\alpha}{2})|\uparrow\rangle_x \\ &= \cos\frac{\lambda n\alpha}{2}|\uparrow\rangle_x - i\sin\frac{\lambda n\alpha}{2}|\downarrow\rangle_x. \end{aligned}$$

Now if we choose λ so that

$$\lambda\alpha = \pi,$$

then we have

$$U(t)|\uparrow\rangle_x = \cos\frac{\pi n}{2}|\uparrow\rangle_x - i\sin\frac{\pi n}{2}|\downarrow\rangle_x,$$

and thus, if n is odd, we get $|\downarrow\rangle_x$ (up to a multiplicative factor), and if n is even, we get $|\uparrow\rangle_x$ (again, up to a multiplicative factor). Therefore, by measuring the spin in the x direction at the end, we can determine whether n is even or odd.

We have thus been able, with just one qubit, to find something that we couldn't do classically at all. Note however, that the information we got was just a single bit ("even" or "odd").

1.2.4. simulating a quantum computer with a classical one. As we saw above, we can use qubits to get results which are much harder, or even impossible to reach using just simple classical bits. However, when we consider a computer, it is simply some black box which accepts some vectors as input, operates on them, and returns a new vector as an output. All the operations which we do quantum mechanically we can also simulate classically (manipulate vectors, take their projections, ...). The question that should be asked is how much resources does this require?

Assume N qubits. The state describing them is

$$|\psi\rangle = \prod_i(\alpha_i|0\rangle_i + \beta_i|1\rangle_i) = \sum_{j=1}^{2^N} c_j|\varphi\rangle_j,$$

where $|\varphi\rangle_j$ are N-particle states, which give all 2^N possible combinations of N particles being in either state $|0\rangle$ or state $|1\rangle$. For example for the case of $N=3$ we have

$$\begin{aligned} |\psi\rangle &= \prod_{i=1}^{3}(\alpha_i|0\rangle_i + \beta_i|1\rangle_i) \\ &= c_1|0\rangle|0\rangle|0\rangle + c_2|0\rangle|0\rangle|1\rangle + c_3|0\rangle|1\rangle|0\rangle + c_4|0\rangle|1\rangle|1\rangle \\ &\quad + c_5|1\rangle|0\rangle|0\rangle + c_6|1\rangle|0\rangle|1\rangle + c_7|1\rangle|1\rangle|0\rangle + c_8|1\rangle|1\rangle|1\rangle. \end{aligned}$$

The number of parameters describing such a state is $2\cdot 2^N - 2$: We have 2^N coefficients c_i, each one of those is actually two number since these are complex numbers, however if we require that ψ be normalized (one constraint) and don't mind if it is multiplied by a global phase $e^{i\theta}$, then two parameters may be dropped giving us $2\cdot 2^N - 2$. If we assume that we need at least one bit for every such parameter,[15] this means that for an N qubit system we need at least $2\cdot 2^N - 2$ bits for the classical simulation. Such a fast increase makes simulations impossible very quickly.

1.2.5. examples.

[14] since $\sigma_z^2 = \mathbb{1}$, then $\sigma_z^{2m} = \mathbb{1}$ and $\sigma_z^{2m+1} = \sigma_z$. Therefore the Taylor series for $e^{i\theta\sigma_z}$ can be written as

$$e^{i\theta\sigma_z} = \sum_n \frac{1}{n!}(i\theta\sigma_z)^n = \sigma_z \sum_{n\text{ odd}} \frac{1}{n!}(i\theta)^n + \sum_{n\text{ even}} \frac{1}{n!}(i\theta)^n = i\sigma_z\sin\theta + \cos\theta.$$

[15] Since the c_i are continuous parameters we need an infinite number of bits to describe each parameter. However, if we settle for a finite precision for the c_i's, then a finite number of bits will suffice to describe each one of them.

1.2.5.1. *Deutsch's problem.* Assume a black box which accepts a single bit as input and gives a single bit as output. We shall denote the effect of the box as $f(x)$ [if the input bit is x then we get $f(x)$ as output]. There are of course 4 different possible functions $f(x)$ which may describe the black box (each of the two possible inputs has two possible outcomes). We would like to know whether $f(x)$ is a constant function, i.e. $f(0) = f(1)$, or whether it is a balanced function, i.e. $f(0) \neq f(1)$.[16]

Classically, to determine the type of function, we must make *two* runs of the system. First we enter a "0" input and see the result, and then we enter "1" as input and see what the outcome is. Such a test would give $f(x)$ exactly and will therefore also tell us if $f(x)$ is constant or balanced. However, as we shall see, using quantum mechanics and the superposition principle we can find the type of function (constant or balanced) with just a single run.

Now, in order to use quantum mechanics, the effect of our black box must be describable by a unitary operator. If $f(x)$ is "balanced" there is no problem, however if $f(x)$ is constant, then we do have a problem: A unitary operator cannot transform two orthogonal states into the same state (a unitary transformation, sends a basis to a new basis, and a constant $f(x)$ lowers the dimension of the basis). We therefore need a slightly different box.

Instead of $f(x)$ we shall use a unitary operator U_D. This operator will both accept and give as output two qubits of information according to the rule

$$|x\rangle_1|y\rangle_2 \xrightarrow{U_D} |x\rangle_1|y \oplus f(x)\rangle_2,$$

where $\oplus$ means adding and then taking the modulo 2 of the result:

$$|1 \oplus 0\rangle = |1\rangle,$$

$$|1 \oplus 1\rangle = |0 \oplus 0\rangle = |0\rangle.$$

Before using this new operator let us first check that it is indeed unitary. Clearly by the definition of U_D we have

$$U_D|x\rangle_1|0\rangle_2 \neq U_D|x\rangle_1|1\rangle_2$$

and

$$U_D|0\rangle_1|y\rangle_2 \neq U_D|1\rangle_1|y'\rangle_2 \quad \text{(any } y, y'\text{)},$$

where in the second relation y and y' may be the same or different. Therefore (if $x = 0$ or 1, and $y = 0$ or 1) we must have (since $\langle 0|1\rangle = 0$)[17]

$$U_D|x\rangle_1|0\rangle_2 \perp U_D|x\rangle_1|1\rangle_2,$$

and

$$U_D|0\rangle_1|y\rangle_2 \perp U_D|1\rangle_1|y'\rangle_2,$$

or simply

$$U_D|x\rangle_1|y\rangle_2 \perp U_D|x'\rangle_1|y'\rangle_2 \quad \left(\begin{array}{c} x \neq x' \text{ and/or } y \neq y' \\ x, x', y, y' = 0, 1 \end{array} \right).$$

By this last result we see that applying U_D to the orthogonal basis

$$\{|0\rangle_1|0\rangle_2, |0\rangle_1|1\rangle_2, |1\rangle_1|0\rangle_2, |1\rangle_1|1\rangle_2\}$$

[16] Note that we don't care what $f(x)$ is exactly. If $f(0) = f(1) = 0$ or $f(0) = f(1) = 1$ doesn't matter to us. In both cases the function is constant.

[17] If $x_i = 0$ or 1 and $y_i = 0$ or 1, then applying U_D on $|x_i\rangle_1|y_i\rangle_2$ will give $|x_i'\rangle_1|y_i'\rangle_2$ with $x_i' = 0$ or 1 and $y_i = 0$ or 1. Thus if we know that two states ($|\psi_1\rangle = U_D|x_1\rangle_1|y_1\rangle_2$ and $|\psi_2\rangle = U_D|x_2\rangle_1|y_2\rangle_2$) are different, it necessarily means that their inner product ($\langle\psi_1|\psi_2\rangle$) must include $\langle 0|1\rangle$ (or $\langle 1|0\rangle$), and since $\langle 0|1\rangle = 0$, then they must be orthogonal.

gives us a new set of four mutually orthogonal states.[18] Since the four new states are mutually orthogonal, they must constitute a basis. Thus, the transformation U_D took us from one orthonormal basis to another, which means that U_D must be unitary, as claimed.[19]

To find, using our new quantum black box, whether $f(x)$ is a constant function or a balanced one, we can of course run it twice (once putting $x = 0$ and once $x = 1$) and see . However, we can also use the superposition principle to determine this with just a single run. To see this let us first try as input $|x=0\rangle_1$ and $\frac{1}{\sqrt{2}}(|y=0\rangle_2 - |y=1\rangle_2)$. By applying U_D we have

$$U_D\left[\frac{1}{\sqrt{2}}|0\rangle_1(|0\rangle_2 - |1\rangle_2)\right] = \frac{1}{\sqrt{2}}|0\rangle_1(|0\oplus f(0)\rangle - |1\oplus f(0)\rangle) = \frac{(-1)^{f(0)}}{\sqrt{2}}|0\rangle_1(|0\rangle - |1\rangle),$$

where the last equality is due to the fact that

$$|0\oplus f(0)\rangle - |1\oplus f(0)\rangle = |0\rangle - |1\rangle \quad \text{for } f(0) = 0,$$

and

$$|0\oplus f(0)\rangle - |1\oplus f(0)\rangle = |1\rangle - |0\rangle \quad \text{for } f(0) = 1.$$

By the same logic, if we input $|x=1\rangle_1$ and $\frac{1}{\sqrt{2}}(|y=0\rangle_2 - |y=1\rangle_2)$ we get

$$U_D\left[\frac{1}{\sqrt{2}}|1\rangle_1(|0\rangle_2 - |1\rangle_2)\right] = \frac{(-1)^{f(1)}}{\sqrt{2}}|1\rangle_1(|0\rangle_2 - |1\rangle_2).$$

Taking a super position $\frac{1}{2}(|0\rangle_1 + |1\rangle_1)(|0\rangle_2 - |1\rangle_2)$ of the two inputs will therefore give us

$$\begin{aligned} U_D\left[\frac{1}{2}(|0\rangle_1 + |1\rangle_1)(|0\rangle_2 - |1\rangle_2)\right] &= \frac{1}{2}\left((-1)^{f(0)}|0\rangle_1 + (-1)^{f(1)}|1\rangle_1\right)(|0\rangle_2 - |1\rangle_2) \\ &= \frac{(-1)^{f(0)}}{2}\left(|0\rangle_1 + (-1)^{f(1)-f(0)}|1\rangle_1\right)(|0\rangle_2 - |1\rangle_2). \end{aligned}$$

If we now examine particle 1 after applying U_D, we see that we get (up to a global multiplicative factor)

$$\begin{cases} |0\rangle_1 + |1\rangle_1 & \text{if } f(0) = f(1) \\ |0\rangle_1 - |1\rangle_1 & \text{if } f(0) \neq f(1) \end{cases}.$$

These two new states are orthogonal to one another, and so may be distinguished by a single measurement [simply measure particle 1 in the basis $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$]. Thus, by applying U_D to a single state

$$\frac{1}{2}(|0\rangle_1 + |1\rangle_1)((|0\rangle_2 - |1\rangle_2)$$

and performing a single *quantum* measurement we can distinguish whether $f(x)$ is constant or balanced, a feat we could not accomplish classically (we had to apply $f(x)$ twice, to two different inputs).

Note, that once again we manged to extract by our measurement just a single bit of information (f is constant or not). The power of quantum mechanics entered in the fact that we can use superposition which cannot be used classically.

1.2.5.2. *Beam splitters and the Mach-Zender interferometer*.

[18] The new states are all different, since two identical states would not be mutually orthogonal to each other (unless they are identically zero, which U_D cannot produce).

[19] If $|e_i\rangle$ is an orthonormal basis and $|h_i\rangle$ is a second orthonormal basis then the transformation from e to h is

$$U = \sum_i |h_i\rangle\langle e_i|,$$

which is clearly unitary.

1.2.5.3. *dense coding.* As we saw earlier one can store a lot of information in a qubit, however only 1 bit may be extracted with certainty. We shall now see how, with the use of entanglement, we can communicate *two* bits of information by transferring a *single* qubit.

Our system will include two qubits (A and B), which we will describe using a special basis known as *Bell states*

$$\begin{aligned}
\psi^- &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B\right)\\
\psi^+ &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right)\\
\phi^- &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B - |1\rangle_A|1\rangle_B\right)\\
\phi^+ &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right),
\end{aligned}$$

where the subscripts A, B tell us to which qubit/particle the ket belongs (in many cases we shall drop the subscripts and keep the order of the kets constant).

This special basis has the convenient traits that it is orthonormal and that all four states are entangled (see section 2.4). The basis is also the set of mutual eigenvectors of the complete set of commuting operators[20][21]

$$\sigma_{x_A}\sigma_{x_B} \quad \text{and} \quad \sigma_{z_A}\sigma_{z_B}$$
$$[\sigma_{x_A}\sigma_{x_B}, \sigma_{z_A}\sigma_{z_B}] = 0.$$

We now define (using the Pauli matrices) a new set of unitary operators $U_{ij}^{(A)}$ such that

$$\begin{aligned}
U_{00}^{(A)} &= \mathbb{1}_A,\\
U_{01}^{(A)} &= \sigma_{x_A},\\
U_{10}^{(A)} &= \sigma_{y_A},\\
U_{11}^{(A)} &= \sigma_{z_A}.
\end{aligned}$$

From the traits of the Pauli matrices,[22] it is easy to see that applying these operators on the Bell state $|\psi^-\rangle$ gives[23]

$$\begin{aligned}
U_{00}^{(A)}|\psi^-\rangle &= |\psi^-\rangle,\\
U_{01}^{(A)}|\psi^-\rangle &= -|\phi^-\rangle,\\
U_{01}^{(A)}|\psi^-\rangle &= i|\phi^+\rangle,
\end{aligned}$$

[20] Note that the operators commute since there are two particles. For just one particle we have

$$\sigma_x\sigma_z = -\sigma_z\sigma_x = -i\sigma_y \Rightarrow [\sigma_x, \sigma_z] = -2i\sigma_y.$$

However, when we have two particles the minuses cancel and we get

$$\sigma_{x_A}\sigma_{z_A}\sigma_{x_B}\sigma_{z_B} = \sigma_{z_A}\sigma_{x_A}\sigma_{z_B}\sigma_{x_B} = -\sigma_{y_A}\sigma_{y_B} \Rightarrow [\sigma_{x_A}\sigma_{x_B}, \sigma_{z_A}\sigma_{z_B}] = 0$$

[21] Note that $\sigma_{y_A}\sigma_{y_B}$ also commutes with $\sigma_{x_A}\sigma_{x_B}$ and $\sigma_{z_A}\sigma_{z_B}$. However, it suffices to look for eigenvectors common to $\sigma_{x_A}\sigma_{x_B}$ and $\sigma_{z_A}\sigma_{z_B}$ in order to uniquely define the four Bell states (see section 5.1). Thus $\sigma_{x_A}\sigma_{x_B}$ and $\sigma_{z_A}\sigma_{z_B}$ (or any other pair of operators from $\sigma_{x_A}\sigma_{x_B}$, $\sigma_{z_A}\sigma_{z_B}$and $\sigma_{y_A}\sigma_{y_B}$) constitute a complete set of commuting operators.

[22] Recall that

$$\begin{aligned}
\sigma_z|\uparrow\rangle = |\uparrow\rangle \quad &; \quad \sigma_z|\downarrow\rangle = -|\downarrow\rangle,\\
\sigma_x|\uparrow\rangle = |\downarrow\rangle \quad &; \quad \sigma_x|\downarrow\rangle = |\uparrow\rangle,\\
\sigma_y|\uparrow\rangle = i|\downarrow\rangle \quad &; \quad \sigma_z|\downarrow\rangle = -i|\uparrow\rangle.
\end{aligned}$$

[23] Note that the operators $U^{(A)}$ operate only on the single particle A. Thus to be rigorous, the operator operating on $|\psi^-\rangle$ is actually $U^{(A)}\mathbb{1}_B$. That is, it's an operator which applies $U^{(A)}$ on particle A and does nothing to particle B.

$$U_{11}^{(A)}|\psi^-\rangle = |\psi^+\rangle.$$

Having constructed our tools, we can now turn to our original problem: communicating two bits of information using only a single qubit. To see how this may be done let Alice and Bob be two distant persons, Alice holding particle A and Bob holding particle B. The particles, as Alice and Bob both know, are given to be in the Bell state $|\psi^-\rangle$. Now, assume that Alice wishes to communicate to Bob *two* bits of information: i and j ($i = 0,1$ and $j = 0,1$). To do this Alice operates locally on her particle with the operator $U_{ij}^{(A)}$ we defined. As a result (according to the effect of $U_{ij}^{(A)}$ on $|\psi^-\rangle$ given above) the two particles A and B together, are now in one of the orthonormal Bell states (up to a global phase). Next, Alice sends her particle, which is a single qubit, to Bob. Having both particles, Bob can now make a measurement (locally) on the state and determine in which of the orthogonal states the two particles are.[24] since Bob knows, that the particles were originally in state $|\psi^-\rangle$, he can therefore infer which operator Alice applied on her particle and thus find i, j.

We have thus seen that by merely passing a single qubit from Alice to Bob, Alice could communicate (to Bob) two bits of information. The extra (second) bit communicated was hidden in the entanglement of the two particles.

Note, that a benefit of this method is encryption. If a third person tries to intercept the message, all he gets is a single qubit, which gives *him* no information at all. Unlike Bob, any other person who gets the transmitted particle has no extra information and therefore cannot infer from it anything.

[24] since the possible states are orthogonal, Bob can make a measurement which distinguishes between all four. For example he can measure the operator

$$O = 1\cdot|\psi^-\rangle\langle\psi^-| + 2\cdot|\psi^+\rangle\langle\psi^+| + 3\cdot|\phi^-\rangle\langle\phi^-| + 4\cdot|\phi^+\rangle\langle\phi^+|.$$

If the result we measure is 1, we know the particles were in state $|\psi^-\rangle$ if we measure 2 we know the particles were in state $|\psi^+\rangle$, and so on.

CHAPTER 2

Basics of quantum information

2.1. Basics of quantum mechanics

Every physical theory is defined by the following:

- The method of describing a system.
- The dynamics of a system.
- The method of measuring a system.

In quantum mechanics observable quantities are described by Hermitian operators ($O^\dagger = O$). Such operators have the following traits:

- All eigenvalues are real:
$$\lambda_a \in \mathbb{R} \quad (O|a\rangle = \lambda_a|a\rangle).$$
- Eigenvectors of *different* eigenvalues are orthogonal:
$$\lambda_a \neq \lambda_{a'} \Rightarrow \langle a|a'\rangle = 0.$$
- Every Hermitian operator may be written in a *spectral decomposition* form
$$O = \sum_a \lambda_a \Pi_a,$$
where Π_a is the projection onto the subspace of eigenvectors with eigenvalues λ_a
$$\begin{aligned}\Pi_a^2 &= \Pi_a,\\ \Pi_a \Pi_{a'} &= \delta_{aa'}\Pi_a,\\ \sum_a \Pi_a &= \mathbb{1}.\end{aligned}$$

In general, projections are Hermitian operators (i.e. they are observables) such that if Π is a projection then
$$\Pi^2 = \Pi \quad (\Pi^\dagger = \Pi).$$
From the spectral decomposition trait, every state may be written as
$$|\psi\rangle = \sum_a \Pi_a|\psi\rangle.$$

Using this decomposition we can define the effect of measurement in quantum mechanics as follows: A measurement of the quantity A for a state $|\psi\rangle$ results in a collapse of the state into one of the eigenvalue subspaces of A, i.e.
$$|\psi\rangle \xrightarrow{\text{measure } A} \frac{\Pi_a|\psi\rangle}{\sqrt{\langle\psi|A|\psi\rangle}}.$$
The probability of the collapse to the subspace a is given by
$$\text{prob}(\Pi_a = 1) = \text{prob}(A = a) = \langle\psi|\Pi_a|\psi\rangle.$$

CONCLUSION. If two states are not orthogonal ($\langle\psi_1|\psi_2\rangle \neq 0$), then one cannot distinguish between them with certainty. In other words, there exists no projection Π such that
$$\text{prob}(\Pi = 1) = \langle\psi_1|\Pi|\psi_1\rangle = 1$$

and

$$\text{prob}(\Pi = 0) = \langle\psi_2|\Pi|\psi_2\rangle = 0.$$

PROOF. Let us assume that there exists such a projection (for $\langle\psi_1|\psi_2\rangle \neq 0$). We define by a Grahm-Schmidt process a new state $|\varphi_2\rangle$ orthonormal to $|\psi_1\rangle$:

$$|\tilde{\varphi}_2\rangle = |\psi_2\rangle - \langle\psi_1|\psi_2\rangle|\psi_1\rangle,$$

$$|\varphi_2\rangle = \frac{|\tilde{\varphi}_1\rangle}{\sqrt{\langle\tilde{\varphi}_2|\tilde{\varphi}_2\rangle}}.$$

Since $|\psi_1\rangle$ and $|\psi_2\rangle$ are not orthogonal while $|\psi_1\rangle$ and $|\varphi_2\rangle$ are,[1] then there exist $\alpha, \beta \neq 0$ such that

$$|\psi_2\rangle = \alpha|\psi_1\rangle + \beta|\varphi_2\rangle.$$

If we now substitute this new form of $|\psi_2\rangle$ into the assumption $\langle\psi_2|\Pi|\psi_2\rangle = 0$, we get

$$\begin{aligned} 0 &= (\alpha^*\langle\psi_1| + \beta^*\langle\varphi_2|)\,\Pi\,(\alpha|\psi_1\rangle + \beta|\varphi_2\rangle) \\ &= |\alpha|^2\langle\psi_1|\Pi|\psi_1\rangle + |\beta|^2\langle\varphi_2|\Pi|\varphi_2\rangle + \alpha^*\beta\langle\psi_1|\Pi|\varphi_2\rangle + \alpha\beta^*\langle\varphi_2|\Pi|\psi_1\rangle. \end{aligned}$$

Now, since $\langle\psi_1|\Pi|\psi_1\rangle = 1$ and also $\langle\psi_1|\psi_1\rangle = 1$, then we must have $\Pi|\psi_1\rangle = |\psi_1\rangle$ (and $\langle\psi_1|\Pi = \langle\psi_1|$).[2] Thus if $|\psi_1\rangle$ and $|\varphi_2\rangle$ are orthogonal, then

$$\langle\psi_1|\Pi|\varphi_2\rangle = \langle\varphi_2|\Pi|\psi_1\rangle = 0 \quad \left(\Leftarrow \begin{array}{c} \langle\psi_1|\varphi_2\rangle = 0, \\ \Pi|\psi_1\rangle = |\psi_1\rangle \end{array}\right),$$

and since Π is hermitian, then

$$\langle\psi_1|\Pi|\psi_1\rangle \geq 0 \quad \text{and} \quad \langle\varphi_2|\Pi|\varphi_2\rangle \geq 0.$$

Therefore, together with $|\alpha|^2, |\beta|^2 > 0$ and $\langle\psi_1|\Pi|\psi_1\rangle = 1$, the expression we just found for $\langle\psi_2|\Pi|\psi_2\rangle = 0$ becomes

$$0 = |\alpha|^2 + |\beta|^2\langle\varphi_2|\Pi|\varphi_2\rangle > 0,$$

which is a contradiction. Hence, there does *not* exist a projection Π such that

$$\text{prob}(\Pi = 1) = \langle\psi_1|\Pi|\psi_1\rangle = 1$$

and

$$\text{prob}(\Pi = 0) = \langle\psi_2|\Pi|\psi_2\rangle = 0,$$

if $\langle\psi_1|\psi_2\rangle \neq 0$ □

Before going on it should be mentioned that in quantum mechanics one can distinguish between two types of systems. The first is that of a closed system: a system for which all elements are known as well as their interactions with one another. The second type, is that of an open system, where in addition to the elements we are interested in, there is also an environment. This environment interacts with our system, however, in a way that we do *not* know exactly.

[1] And we assume $|\psi_1\rangle \neq |\psi_2\rangle$.

[2] We may always write

$$\Pi|\psi_1\rangle = \alpha|\psi_1\rangle + \beta|\psi_\perp\rangle,$$

such that $|\alpha|^2 + |\beta|^2 = 1$, and $\langle\psi_1|\psi_\perp\rangle = 0$. Thus

$$\langle\psi_1|\Pi|\psi_1\rangle = \alpha\langle\psi_1|\psi_1\rangle + \beta\langle\psi_1|\psi_\perp\rangle = \alpha\langle\psi_1|\psi_1\rangle.$$

Since we also have $\langle\psi_1|\Pi|\psi_1\rangle$, then we must have $\alpha = 1$ and $\beta = 0$, which proves that $\Pi|\psi_1\rangle = |\psi_1\rangle$.

2.2. Spin $\frac{1}{2}$ and the Pauli matrices

Since spin $\frac{1}{2}$ particles are used very often in quantum information it is worth while to make a short review of (some of) their traits. The observables for measuring spin $\frac{1}{2}$ in the x, y, and z directions are the Pauli operators σ_x, σ_y, and σ_z respectively. These are also denoted as σ_1, σ_2, and σ_3 respectively. The operators obey the commutation relation of angular momentum

$$[\sigma_i, \sigma_j] = i\varepsilon_{ijk}\sigma_k,$$

where ε_{ijk} is the antisymmetric tensor:

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{for } \varepsilon_{123} \text{ and all cyclic permutations of } i,j,k \\ -1 & \text{for } \varepsilon_{321} \text{ and all cyclic permutations of } i,j,k \\ 0 & \text{otherwise} \end{cases}.$$

In other words

$$[\sigma_x, \sigma_y] = i\sigma_z,$$
$$[\sigma_z, \sigma_x] = i\sigma_y,$$
$$[\sigma_y, \sigma_z] = i\sigma_x,$$

with all other cases obvious from these.

The Pauli operators also have the following traits:

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}\mathbb{1},$$
$$\sigma_i\sigma_j = \delta_{ij}\mathbb{1} + i\varepsilon_{ijk}\sigma_k,$$
$$\text{Tr}[\sigma_i] = 0,$$
$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbb{1},$$
$$\sigma_i^\dagger = \sigma_i \quad (\text{Hermitian}),$$

and

$$\sigma_i^\dagger\sigma_i = \mathbb{1} \quad (\text{unitarty}).$$

Note, that the Pauli operators are both unitary and Hermitian.

The Pauli operators are usually represented by the standard Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

and

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

The eigenstates of the of the Pauli matrices are the "up" (eigenvalue $+1$) and "down" (eigenvalue -1) states in the appropriate direction. The relations between these eigenstates are as follows:

$$|\uparrow_x\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle + |\downarrow_z\rangle\right)$$
$$|\downarrow_x\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle - |\downarrow_z\rangle\right)$$
$$|\uparrow_y\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle + i|\downarrow_z\rangle\right)$$
$$|\downarrow_y\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle - i|\downarrow_z\rangle\right).$$

The effect of the Pauli operators on the "up" and "down" states in the z direction is as follows:

$$\sigma_x|\uparrow_z\rangle = |\downarrow_z\rangle,$$

$$\sigma_x|\downarrow_z\rangle = |\uparrow_z\rangle,$$
$$\sigma_y|\uparrow_z\rangle = i|\downarrow_z\rangle,$$
$$\sigma_y|\downarrow_z\rangle = -i|\uparrow_z\rangle.$$

Thus we say that σ_x flips in the z direction, while σ_y not only flips in the z directions, but also adds a phase (dependent on the original state).

2.3. Open systems, mixtures and the density matrix

So far we have only dealt with closed quantum systems, we shall now turn to treat open systems. There are two general cases in which we encounter open systems:

(1) Lack of knowledge of the full system: We might not know some of the initial conditions (the initial state of the system), or some of the parameters of the system, or not know exactly the dynamics of the system.
(2) We are dealing with a system of two (or more) subsystems, which we fully know how to describe, however, we are interested in making measurements only on part of the full system.

In both these cases the treatment is different than that of closed systems. We shall see that we have to use mixtures instead of regular states, where these mixtures will be described by density matrices. Further more, probabilities will behave slightly different: Instead of one state evolving in time, we shall have several, each with a different probability to occur. This is different from a linear of combination of states (superposition), since here each state is treated separately and there is no interference effect.[3]

To see the difference between open and closed systems let us study an example. Assume two states; state $|\psi_A\rangle$

$$|\psi_A\rangle = a_0|0\rangle + a_1|1\rangle \quad (|a_0|^2 + |a_1|^2 = 1),$$

with probability p_a to occur; and state $|\psi_B\rangle$

$$|\psi_B\rangle = b_0|0\rangle + b_1|1\rangle \quad (|b_0|^2 + |b_1|^2 = 1),$$

with probability $p_b = 1 - p_a$ to occur. What is the probability to measure $|0\rangle$ in this case? i.e. what is

$$\text{prob}(\Pi_0 = 1) = ? \quad (\Pi_0 \equiv |0\rangle\langle 0|).$$

If we make many measurements, in p_a of them the measurement will be of state $|\psi_A\rangle$ and in $p_b = 1 - p_a$ they will be of state $|\psi_b\rangle$. Therefore, the probability to measure $|0\rangle$ will be p_a times the probability to measure $|0\rangle$ in case $|\psi_A\rangle$ plus p_b times the probability to measure $|0\rangle$ in case $|\psi_B\rangle$:

$$\begin{aligned}\text{prob}(\Pi_0 &= 1) = p_a\langle\psi_A|\Pi_0|\psi_A\rangle + p_b\langle\psi_B|\Pi_0|\psi_B\rangle \\ &= p_a|a_0|^2 + p_b|b_0|^2 = p_a|a_0|^2 + (1-p_a)|b_0|^2.\end{aligned}$$

If on the other hand, instead of having a probability for each state ($|\psi_A\rangle$ and $|\psi_B\rangle$), we make a superposition

$$|\psi_{AB}\rangle = \alpha|\psi_A\rangle + \beta|\psi_B\rangle = (\alpha a_0 + \beta b_0)|0\rangle + (\alpha a_1 + \beta b_1)|1\rangle \quad (|\alpha|^2 + |\beta|^2 = 1),$$

then, in this case, we shall find

$$\text{prob}(\Pi_0 = 1) = |\alpha a_0 + \beta b_0|^2.$$

If we now compare the two results, we see that they are markedly different. In the mixture (assuming $a_0, b_0 \neq 0$), no matter the value of p_a there will always be a finite probability to measure $|0\rangle$. However, in the superposition case, we may choose α and β such that the probability to measure $|0\rangle$ will be zero. The difference, as mentioned above, is that in the latter case we have interference: all the coefficients appear within one absolute value

[3] Recall, that in a linear combination of states, the coefficients appearing are not the probabilities of each state, but their amplitude. You must take the absolute value squared to find the probability.

(squared). However in the mixture case, there is no interference and we have a sum with two absolute values (squared).

2.3.1. the density matrix.

The mathematical tool we use to describe mixtures is the density matrix. Assume a set $\{p_i, |\psi_i\rangle\}$ of possible states $|\psi_i\rangle$ (not necessarily orthogonal, but $\langle\psi_i|\psi_i\rangle = 1$), each with probability p_i to occur. We define the density matrix/operator ρ as

$$\rho \equiv \sum_i p_i|\psi_i\rangle\langle\psi_i| \quad (0 \le p_i \le 1, \sum_i p_i = 1).$$

If we write it in an orthonormal basis $|n\rangle$ ($\langle n|m\rangle = \delta_{nm}$), then

$$\rho_{nm} = \langle n|\rho|m\rangle$$

and

$$\mathrm{Tr}\,\rho = \sum_n \langle n|\rho|n\rangle = \sum \rho_{nn}.$$

The density matrix has the following traits ($|n\rangle$ is an orthonormal basis):

(1) Its trace is 1:

$$\mathrm{Tr}\,\rho = \sum_n \langle n|\rho|n\rangle = \sum \rho_{nn} = 1.$$

(2) The density matrix is Hermitian and the sum of its eigenvalues is 1

$$\rho = \rho^\dagger \Rightarrow \sum_k \lambda_k = 1 \quad (\rho|\varphi_k\rangle = \lambda_k|\varphi_k\rangle).$$

(3) The density matrix is a *positive operator*, i.e. for every state $|\psi\rangle$ in the Hilbert space, we have

$$\langle\psi|\rho|\psi\rangle \ge 0 \quad (\forall|\psi\rangle),$$

which is equivalent to having all its eigenvalues nonnegative

$$\langle\psi|\rho|\psi\rangle \ge 0 \Leftrightarrow \lambda_k \ge 0.$$

Note, that together with the previous trait ($\sum_k \lambda_k = 1$), we must have

$$0 \le \lambda_k \le 1.$$

PROOF. (1) To prove that $\mathrm{Tr}\,\rho = 1$ we shall use the definition of the density matrix. The trace of an operator is independent of the (orthonormal) basis we work in. If $|n\rangle$ is some orthonormal basis, then

$$\begin{aligned}
\mathrm{Tr}\,\rho &= \sum_n \rho_{nn} \equiv \sum_n \langle n| \left(\sum_i p_i|\psi_i\rangle\langle\psi_i|\right)|n\rangle \\
&= \sum_{n,i} p_i\langle n|\psi_i\rangle\langle\psi_i|n\rangle = \sum_{n,i} p_i\langle\psi_i|n\rangle\langle n|\psi_i\rangle \\
&= \sum_i p_i\langle\psi_i|\left(\sum_n |n\rangle\langle n|\right)|\psi_i\rangle = \sum_i p_i\langle\psi_i|\psi_i\rangle \\
&= \sum p_i = 1,
\end{aligned}$$

where we have used the trait of orthonormal bases

$$\sum_n |n\rangle\langle n| = \mathbb{1}.$$

(2) Proving that ρ is Hermitian is very simple from its definition. Since the p_i are real ($0 \le p_i \le 1$), then

$$\rho^\dagger = \left(\sum_i p_i|\psi_i\rangle\langle\psi_i|\right)^\dagger = \sum_i p_i|\psi_i\rangle\langle\psi_i| = \rho.$$

Since ρ is Hermitian, then it may be diagonalized. The sum of its eigenvalues, is its trace, and thus from the previous trait we must have

$$\sum_k \lambda_k = \operatorname{Tr}\rho = 1.$$

(3) To show that the density matrix ρ is a positive operator ($\langle\psi|\rho|\psi\rangle \geq 0$, $\forall\psi$) we shall calculate $\langle\psi|\rho|\psi\rangle$ using the definition of the density matrix. For any state $|\psi\rangle$:

$$\begin{aligned}\langle\psi|\rho|\psi\rangle &= \langle\psi|\left(\sum_i p_i|\psi_i\rangle\langle\psi_i|\right)|\psi\rangle \\ &= \sum_i p_i\langle\psi|\psi_i\rangle\langle\psi_i|\psi\rangle = \sum_i p_i|\langle\psi|\psi_i\rangle|^2.\end{aligned}$$

Since $p_i \geq 0$ and $|\langle\psi|\psi_i\rangle|^2 \geq 0$, then we necessarily have

$$\langle\psi|\rho|\psi\rangle \geq 0,$$

as required.

With this result we can now show that all eigenvalues of ρ are non-negative. To show this, we choose $|\psi\rangle = |\varphi_k\rangle$, where $|\varphi_k\rangle$ is an eigenvector of ρ with eigenvalue λ_k ($\rho|\varphi_k\rangle = \lambda_k|\varphi_k\rangle$). Using the last result we find

$$\begin{gathered}0 \leq \langle\varphi_k|\rho|\varphi_k\rangle = \langle\varphi_k|\lambda_k|\varphi_k\rangle = \lambda_k \\ \Rightarrow \lambda_k \geq 0,\end{gathered}$$

which is just what we wished to prove.[4]

□

As we saw, the density matrix describes a mixture of states, however, it may also describe a regular state. This latter case occurs when the mixture includes only a single state with probability $p = 1$. We say that such a mixture is a *pure state* (otherwise it is called a *mixed state*). In other words, a system is in a pure state if there exists a state $|\psi\rangle$ such that

$$\rho = |\psi\rangle\langle\psi| \quad \text{(pure state)}.$$

The density matrix of a pure state has the special trait that $\rho^2 = \rho$. Note, that this trait holds only for pure states, i.e.

$$\rho^2 = \rho \Leftrightarrow \text{pure state}.$$

PROOF. Clearly, if we have a pure-state density matrix ρ ($\rho = |\psi\rangle\langle\psi|$, $\langle\psi|\psi\rangle = 1$), then

$$\rho^2 = (|\psi\rangle\langle\psi|)(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \rho,$$

which proves one direction (pure state $\Rightarrow \rho^2 = \rho$).

As for the opposite direction, it is simple to show (see the end of this proof) that a density matrix ρ obeys $\rho^2 = \rho$ if and only if ρ has a single eigenvector $|\varphi_1\rangle$ with eigenvalue $\lambda = 1$, while all other eigenvectors have an eigenvalue $\lambda = 0$. If this is indeed the case, then by the spectral decomposition we may write

$$\rho = \sum \lambda_i \Pi_{\lambda_i} = 1 \cdot \Pi_1 + 0 \cdot \Pi_0 = |\varphi_1\rangle\langle\varphi_1|,$$

[4] We showed above that if $\langle\psi|\rho|\psi\rangle \geq 0$ for any state $|\psi\rangle$, then necessarily all eigenvalues obey $\lambda_k \geq 0$. To show the opposite direction (assuming the operator can be diagonalized), simply write $|\psi\rangle$ in the basis of eigenvectors $|\varphi_k\rangle$

$$|\psi\rangle = \sum_k \alpha_k|\varphi_k\rangle.$$

Now $\langle\psi|\rho|\psi\rangle$ will give

$$\langle\psi|\rho|\psi\rangle = \sum_{k,k'} \alpha_k\alpha_k^*\langle\varphi_{k'}|\rho|\varphi_k\rangle = \sum_{k,k'} \lambda_k\alpha_k\alpha_k^*\langle\varphi_{k'}|\varphi_k\rangle = \sum_{k,k'} \lambda_k|\alpha_k|^2\delta_{kk'} = \sum_k |\alpha_k|^2\lambda_k \geq 0,$$

which is the desired result $\langle\psi|\rho|\psi\rangle \geq 0$.

or simply

$$\rho = |\varphi_1\rangle\langle\varphi_1|.$$

We have thus proven the second direction, i.e. that $\rho^2 = \rho$ implies that ρ has the form of a pure state.

To complete the proof, we still have to fill in one gap. We must show, as claimed above, that $\rho^2 = \rho$ corresponds to ρ having a single eigenvalue 1 with all others 0. To show this, simply diagonalize ρ to give ρ_D. The diagonal elements of ρ_D are the eigenvalues of ρ. Since ρ is a density matrix, then these eigenvalues obey $0 \leq \lambda_k \leq 1$ and $\sum \lambda_k = 1$. Clearly, in such a case the diagonal matrix ρ_D obeys $\rho_D^2 = \rho_D$ if and only if a single element on the diagonal is 1 and all others are 0 (i.e. if ρ has a single eigenvalue of 1 and all the rest 0).[5] Further more, since the diagonalization of ρ to ρ_D is just a base change, then

$$\rho^2 = \rho \Leftrightarrow \rho_D^2 = \rho_D.$$

Consequently, $\rho^2 = \rho$ if and only if ρ has a single eigenvalue 1, while all others are 0. □

The density matrix (pure or not) has one more important trait: For any projection operator Π, the probability of it measuring true (i.e. of the mixture collapsing, due to the measurement, to the subspace of Π) is

$$\mathrm{prob}(\Pi = 1) = \mathrm{Tr}(\rho\Pi) = \mathrm{Tr}(\Pi\rho).$$

This trait may be further generalized as follows: the average value $\langle O\rangle$ of an observable O, when measured, is

$$\langle O\rangle = \mathrm{Tr}(\rho O) = \mathrm{Tr}(O\rho).$$

PROOF. We shall start by proving the simple form of the trait. By definition (recall $\rho \equiv \sum_i p_i|\psi_i\rangle\langle\psi_i|$)

$$\mathrm{prob}(\Pi = 1) = \sum_i p_i\langle\psi_i|\Pi|\psi_i\rangle.$$

If we now use an orthonormal basis $|n\rangle$, we know that $\sum_n |n\rangle\langle n| = \mathbb{1}$, and we can therefore write the last relation as

$$\begin{aligned}
\mathrm{prob}(\Pi = 1) &= \sum_i p_i\langle\psi_i|\Pi\left(\sum_n |n\rangle\langle n|\right)|\psi_i\rangle \\
&= \sum_{i,n} p_i\langle\psi_i|\Pi|n\rangle\langle n|\psi_i\rangle = \sum_{i,n} p_i\langle n|\psi_i\rangle\langle\psi_i|\Pi|n\rangle \\
&= \sum_n \langle n|\left(\sum_i p_i|\psi_i\rangle\langle\psi_i|\right)\Pi|n\rangle = \mathrm{Tr}\left[\left(\sum_i p_i|\psi_i\rangle\langle\psi_i|\right)\Pi\right] \\
&= \mathrm{Tr}(\rho\Pi),
\end{aligned}$$

which proves the simpler trait.

We can now use the last result to prove the more general trait. By definition

$$\langle O\rangle = \sum_i p_i\langle\psi_i|O|\psi_i\rangle.$$

[5] If ρ_D has more than one non-zero elements on its diagonal, then these elements must be different from 1 (due to $\sum \lambda_i = 1$, $\lambda_i \geq 0$). As a result ρ_D^2will not give ρ_D. For example

$$\begin{pmatrix} \frac{1}{4} & & & & \\ & \frac{3}{4} & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}^2 = \begin{pmatrix} \frac{1}{16} & & & & \\ & \frac{9}{16} & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} \neq \begin{pmatrix} \frac{1}{4} & & & & \\ & \frac{3}{4} & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}.$$

Since O is an observable we can always write it in a spectral decomposition

$$O=\sum_k \lambda_k|\varphi_k\rangle\langle\varphi_k| \equiv \sum_k \lambda_k \Pi_k.$$

Inserting this into the previous relation and using the trait $\text{prob}(\Pi=1)=\text{Tr}(\rho\Pi)$ we get

$$\langle O\rangle=\sum_k \lambda_k \sum_i p_i \langle\psi_i|\Pi_k|\psi_i\rangle=\sum_k \lambda_k \,\text{Tr}(\rho\Pi_k).$$

Now, since Tr is a linear operator then

$$\langle O\rangle=\sum_k \lambda_k \,\text{Tr}(\rho\Pi_k)=\text{Tr}\left(\rho\sum_k \lambda_k\Pi_k\right)=\text{Tr}(\rho O).$$

To prove that $\text{Tr}(\rho\Pi)=\text{Tr}(\Pi\rho)$ and $\text{Tr}(\rho O)=\text{Tr}(O\rho)$, we can simply use the trait of the trace that

$$\text{Tr}(AB)=\text{Tr}(BA).$$

On the other hand, in proving the simpler form, we could have started with

$$\text{prob}(\Pi=1)=\sum_i p_i\langle\psi_i|\left(\sum_n |n\rangle\langle n|\right)\Pi|\psi_i\rangle$$

instead of

$$\text{prob}(\Pi=1)=\sum_i p_i\langle\psi_i|\Pi\left(\sum_n |n\rangle\langle n|\right)|\psi_i\rangle,$$

which would have led us to

$$\text{prob}(\Pi=1)=\text{Tr}(O\rho).$$

□

The traits we have found for the density matrix put constraints on its elements ρ_{nm}. We might therefore ask how many independent (real) parameters describe an $N\times N$ density matrix. If we had no constraints, then there would be N^2 complex elements in the matrix, which would therefore give $2N^2$ independent *real* parameters. However, we have three constraints

$$\rho^\dagger=\rho,$$
$$\text{Tr}\,\rho=1,$$

and

$$\lambda_i\geq 0 \quad \forall i,$$

where λ_i are the eigenvalues of the density matrix. The first constraint ($\rho^\dagger=\rho$) is actually N^2 equations since on the diagonal $\rho^\dagger=\rho$ gives

$$\rho_{nn}=\rho^*_{nn} \quad (N \text{ equations}),$$

and off the diagonal ($n\neq m$) we have

$$\rho_{nm}=\rho^*_{mn} \quad (N^2-N \text{ equations}).$$

Note, that in the off-diagonal case, the number of equations takes into account that we should have both doubled and halved the number of equations (relative to the N^2-N off-diagonal elements). The number of equations should have been doubled since each equality gives two equations: one for the real part and a second for the imaginary part. On the other hand, the number of equations should have been halved since it suffices to count only the pairs n,m above the diagonal [$\frac{1}{2}(N^2-N)$ pairs], as those below will give us the same equations again. We did not double the equations for the elements on the diagonal, since these equations only tell us that the imaginary part is zero, but do not tell us anything about the real part (it equals itself, which is trivial).

To the above constraints we must also add the one on the trace:

$$\text{Tr}\,\rho=1 \quad (1 \text{ equation}).$$

This constraint is just a single equation, since we already know that the trace has no imaginary component (the diagonal elements are all real, due to $\rho^\dagger = \rho$). Subtracting the number of equations from the total number of parameters (in the case of no constraints) we finally get[6]

$$\#\text{of independent parameters} = N^2 - 1.$$

Now that we know that density matrices have $N^2 - 1$ parameters, the question might be, how do we parametrize these matrices, i.e. how do we write the matrices as a function of $N^2 - 1$ parameters. To do this we note that not only density matrices have $N^2 - 1$ parameters, but that the $SU(N)$ group of matrices[7] also has $N^2 - 1$ parameters . Since both sets have the same number of parameters, we might try and relate the two somehow. In general, it is impossible (at least in a simple way) to construct density matrices using a linear combination of unitary matrices from $SU(N)$. However, we may use their generators.[8] We shall next see how this is done for the case of $N = 2$.

For $N = 2$, one possible set of generators, of $SU(2)$, is the Pauli matrices σ_i.[9] For convenience, we define a vector of matrices

$$\vec{\sigma} \equiv (\sigma_x, \sigma_y, \sigma_z) \equiv (\sigma_1, \sigma_2, \sigma_3),$$

and an inner product of matrices[10]

$$\langle A, B\rangle \equiv \mathrm{Tr}(A^\dagger B).$$

Using this last definition we find that the Pauli matrices are orthogonal to one another

$$\langle \sigma_i, \sigma_j\rangle = 2\delta_{ij}.$$

If we also add the unit matrix to the Pauli matrices, we now have $N^2 = 4$ matrices, and these four (using complex coefficients) span the space of 2×2 matrices. To see this, note that if we define

$$\sigma_0 \equiv \mathbb{1},$$

[6] Note, that we have not used the constraint that all eigenvalues must be non-negative. This constraint does not change the number of parameters, it just reduces the range of the parameters. It reduces the region of allowed parameters in the $N^2 - 1$ dimensional space.

[7] The $SU(N)$ group (**S**pecial **U**nitary group) is the group of all $N \times N$ unitary matrices with determinant $+1$:

$$U \in SU(N) \Leftrightarrow U^\dagger U = \mathbb{1}, \det(U) = +1.$$

(In general, unitary matrices have a determinant of $e^{i\theta}$, with $\theta \in \mathbb{R}$).

It is easy to see that $SU(N)$ has $N^2 - 1$ independent (real) parameters, since $U^\dagger U = \mathbb{1}$ gives N^2 equations, and $\det(U) = +1$ is another equation. Thus we have $N^2 + 1$ equations for the $2N^2$ real parameters of U (i.e. $N^2 - 1$ real independent parameters). The $2N^2$ original real parameters are due to the fact that an $N \times N$ matrix has N^2 elements, but each has two components: a real part and imaginary part. Note, that this logic does not work for the number of equations in $U^\dagger U = \mathbb{1}$ and $\det(U) = +1$. This is because $U^\dagger U$ mixes real and imaginary parts, and thus we cannot split the N^2 complex equation into $2N^2$ equations (real and imaginary).

[8] The generators of a group, in this case, are a set of matrices g_i such that any element in the group may be written as

$$\prod e^{i\theta_j g_j} \quad (\text{or } e^{i\Sigma\theta'_j g_j}),$$

where the θ_j's are real. For the $SU(N)$ group there are $N^2 - 1$ generators g_i.

[9] Reminder: The Pauli matrices are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

[10] It can easily be checked that $\mathrm{Tr}(A^\dagger B)$ obeys all the requirements of an inner product (α a scalar):

$$\langle A+B, C\rangle = \langle A, C\rangle + \langle B, C\rangle,$$
$$\langle \alpha A, B\rangle = \alpha^* \langle A, B\rangle,$$
$$\langle B, A\rangle = \langle A, B\rangle^*,$$
$$\langle A, A\rangle \geq 0 \quad \text{where } \langle A, A\rangle = 0 \Leftrightarrow A = 0.$$

then the above inner product $\langle\sigma_i,\sigma_j\rangle = 2\delta_{ij}$ still holds, even when i, j run from 0 to 3. Since the four matrices are orthogonal to each other, then they necessarily constitute a basis of all the 2×2 matrices (a four dimensional space).

Since the Pauli matrices, together with the unit matrix, constitute a basis, then any density matrix may be written in the form

$$\rho = a_0\mathbb{1} + a_1\sigma_x + a_2\sigma_2 + a_3\sigma_3 \equiv a_0\mathbb{1} + \vec{a}\cdot\vec{\sigma}.$$

To find the coefficients a_i we apply the constraints we had on the density matrix. First we apply the constraint on the trace: $\mathrm{Tr}\,\rho = 1$. Since $\mathrm{Tr}\,\sigma_i = 0$ for all three Pauli matrices ($i = 1,2,3$), then the condition $\mathrm{Tr}\,\rho = 1$ becomes (recall that here $\mathbb{1}$ is a 2×2 matrix)

$$1 = \mathrm{Tr}\,\rho = a_0\,\mathrm{Tr}\,\mathbb{1} + 0 = 2a_0$$

$$\Rightarrow a_0 = \frac{1}{2}.$$

Now, the second requirement we had is that ρ be Hermitian ($\rho^\dagger = \rho$). Since the Pauli matrices themselves (and $\mathbb{1}$) are Hermitian, the requirement becomes

$$\rho^\dagger = (\frac{1}{2}\mathbb{1} + a_1^*\sigma_x + a_2^*\sigma_y + a_3^*\sigma_z) = (\frac{1}{2}\mathbb{1} + a_1\sigma_x + a_2\sigma_y + a_3\sigma_z) = \rho,$$

which means that[11]

$$a_i = a_i^* \Rightarrow a_i \in \mathbb{R}.$$

For convenience we define

$$\vec{p} \equiv 2\vec{a},$$

which allows us to write the density matrix as

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{p}\cdot\vec{\sigma}) \quad (\vec{p}\in\mathbb{R}^3),$$

or in matrix form

$$\begin{aligned}\rho &= \frac{1}{2}\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + p_1\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + p_2\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} + p_3\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right]\\ &= \frac{1}{2}\begin{pmatrix}1+p_3 & p_1 - ip_2\\ p_1 + ip_2 & 1 - p_3\end{pmatrix}.\end{aligned}$$

The final requirement of the density matrix, is that it be a positive operator. Since we are dealing with a 2×2 matrix with a positive trace, then a necessary and sufficient condition is that the determinant be non-negative[12]

$$\det(\rho) \geq 0.$$

From the matrix form we found for ρ, this means that

$$\det(\rho) = \begin{vmatrix}1+p_3 & p_1 - ip_2\\ p_1 + ip_2 & 1 - p_3\end{vmatrix} = 1 - (p_1^2 + p_2^2 + p_3^2) = 1 - \vec{p}^2 \geq 0.$$

Therefore, we finally have the general form of the density matrix

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{p}\cdot\vec{\sigma}) \quad (\vec{p}\in\mathbb{R}^3, |\vec{p}| \leq 1).$$

[11] Since the Pauli matrices together with the unit matrix constitute a basis, then there is only a single choice of coefficients a_i which gives a certain matrix (if there were more, the matrices would not be linearly independent, and therefore not a basis). The above condition gives two sets of coefficients $\{a_i\}$ and $\{a_i^*\}$. For these sets to be the same we must have $a_i = a_i^*$.

[12] Recall, that the trace of a matrix equals the sum of its eigenvalues, and its determinant is the product of the eigenvalues. Since we are dealing with a 2×2 matrix, it has two eigenvalues. The trace is 1, which is positive, and therefore the product of the eigenvalues, i.e. the determinant, must be non-negative for both eigenvalues to be non-negative.

The vector $\vec{p}$ is called the *polarization vector*. The name is used since $\vec{p}$ gives the average direction (polarization) of the spin:

$$\begin{aligned}\langle\sigma_x\rangle &= \mathrm{Tr}(\rho\sigma_x) = p_x,\\ \langle\sigma_y\rangle &= \mathrm{Tr}(\rho\sigma_y) = p_y,\\ \langle\sigma_z\rangle &= \mathrm{Tr}(\rho\sigma_z) = p_z.\end{aligned}$$

Note, that as promised, we have three parameters ($N^2-1=3$, for $N=2$) which describe the density matrix. These parameters are the three real components of the vector $\vec{p}$.

To conclude, we see that we can represent all possible density matrices (of a two dimensional Hilbert space) by the possible vectors $\vec{p}$. The possible vectors $\vec{p}$ ($\vec{p}\leq 1$) form a ball of radius 1. This ball is known as the *Bloch sphere*.[13] We shall see that the case $|\vec{p}|=1$ (points on the surface of the Bloch sphere) coincides with pure states.

CLAIM. A 2×2 density matrix describes a *pure* state if and only if the polarization vector $\vec{p}$ is a unit vector ($\vec{p}=\hat{n}$), i.e.

$$\rho = |\psi\rangle\langle\psi| \Leftrightarrow \rho = \frac{1}{2}(\mathbb{1}+\hat{n}\cdot\vec{\sigma}).$$

PROOF. Let us start with the reverse direction, i.e. that $\vec{p}=\hat{n}$ gives a pure state. We know that if we have $\rho^2=\rho$ then ρ describes a pure state, thus it will suffice to show that $\vec{p}=\hat{n}$ implies $\rho^2=\rho$. Let us therefore start by calculating ρ^2 for $\vec{p}=\hat{n}$. By definition (when $\vec{p}=\hat{n}$)

$$\rho^2 = \left[\frac{1}{2}(\mathbb{1}+\hat{n}\cdot\vec{\sigma})\right]^2 = \frac{1}{4}[\mathbb{1}+2\hat{n}\cdot\vec{\sigma}+(\hat{n}\cdot\vec{\sigma})^2],$$

and

$$\begin{aligned}(\hat{n}\cdot\vec{\sigma})^2 &= (n_1\sigma_1+n_2\sigma_2+n_3\sigma_3)^2\\ &= \sum_i n_i^2\sigma_i^2 + \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}}(n_in_j\sigma_i\sigma_j+n_jn_i\sigma_j\sigma_i)\\ &= \sum_i n_i^2\sigma_i^2 + \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}}n_in_j(\sigma_i\sigma_j+\sigma_j\sigma_i).\end{aligned}$$

We know that for the Pauli matrices

$$\sigma_i^2=\mathbb{1}$$

and

$$\sigma_i\sigma_j = -\sigma_j\sigma_i \quad (i\neq j).$$

Therefore, by the above relation for $(\hat{n}\cdot\vec{\sigma})^2$, we must have

$$(\hat{n}\cdot\vec{\sigma})^2 = \left(\sum_i n_i^2\right)\mathbb{1} = \mathbb{1}.$$

Thus ρ^2 becomes

$$\rho^2 = \frac{1}{4}[2\cdot\mathbb{1}+2\hat{n}\cdot\vec{\sigma}] = \frac{1}{2}[\mathbb{1}+\hat{n}\cdot\vec{\sigma}] = \rho.$$

This result ($\rho^2=\rho$ when $\vec{p}=\hat{n}$) proves that $\rho=\frac{1}{2}(\mathbb{1}+\hat{n}\cdot\vec{\sigma})$ describes a pure state (and can thus be written as $\rho=|\psi\rangle\langle\psi|$ for some $|\psi\rangle$).[14]

[13] Yes, the Bloch *sphere*, is actually a *ball*. However, in some places the term Bloch sphere is indeed reserved only for the boundary of the ball.

[14] If we define $\hat{n}$ by the spherical angles θ and φ

$$\hat{n} = \sin\theta\cos\varphi\hat{x}+\sin\theta\sin\varphi\hat{y}+\cos\theta\hat{z},$$

To complete the proof, we must also prove the opposite direction: if $\rho = |\psi\rangle\langle\psi|$, then there exists a unit vector $\hat{n}$ such that

$$\rho = \frac{1}{2}(\mathbb{1} + \hat{n}\cdot\vec{\sigma}).$$

However, we have actually proved that already. If $\rho = |\psi\rangle\langle\psi|$ then we have a pure state and $\rho^2 = \rho$. If we replace $\hat{n}$ in the above proof with $\vec{p}$, we get

$$\rho^2 = \frac{1}{4}[(1+|\vec{p}|^2)\mathbb{1} + 2\vec{p}\cdot\vec{\sigma}].$$

This will give back $\rho^2 = \rho = \frac{1}{2}[\mathbb{1} + \vec{p}\cdot\vec{\sigma}]$ only if $|\vec{p}| = 1$, thus completing the proof. □

So far we have concentrated on qubits and the two dimensional Hilbert space. For an N dimensional space we can use, instead of the Pauli matrices, $N^2 - 1$ (linearly independent) Hermitian matrices h_i with zero trace ($i = 1, 2, \dots, N^2 - 1$). These matrices are the generators of the $SU(N)$ group (recall that in N dimensions the density matrix has $N^2 - 1$ independent parameters). Using these generators the density matrix can be written as[15]

$$\rho_N = \frac{1}{N}\mathbb{1} + \frac{1}{2}\eta_i h_i,$$

where

$$\eta_i = \langle h_i\rangle = \mathrm{Tr}(\rho_N h_i).$$

The allowed combinations of the η_i's define a region in an $N^2 - 1$ dimensional space. If we denote by λ_i the $N^2 - 1$ eigenvalues of ρ_N, then the region V of allowed η_i's, is the region for which all the eigenvalues are positive (and add up to 1, which is immediate since $\mathrm{Tr}\,\rho = 1$), i.e.

$$V = \{\eta_i, \quad i = 1, 2, \dots, N^2 - 1 | \sum\lambda_i = 1, \lambda_i \geq 0\}.$$

The points on the boundary of this region are those point where at least one eigenvalue is zero (beyond this, some have to be negative which is not allowed for density matrices).

So far, all we have just said is true for any N, including $N = 2$. However, there is a big difference between $N = 2$ and $N > 2$. For $N = 2$, the density matrix has only two

then the state $|\psi\rangle$, for which $\rho = |\psi\rangle\langle\psi|$, is (up to a global phase)

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

This is easily seen, since the state $|\psi\rangle$ obeys

$$\hat{n}\cdot\vec{\sigma}|\psi\rangle \equiv \sigma_{\hat{n}}|\psi\rangle = |\psi\rangle,$$

and is therefore an eigenstate of ρ with eigenvalue 1. Since ρ is a density matrix and $|\psi\rangle$ is an eigenvector with eigenvalue 1, then necessarily $\rho = |\psi\rangle\langle\psi|$.

[15] A simple choice for the h_i's is the three types of matrices:

$$h^{(k)}_{ij} = \begin{cases} h_{k,k} = 1 & \\ h_{k+1,k+1} = 1 & \\ h_{i,j} = 0 & \text{otherwise} \end{cases} \qquad (k = 1, 2. \dots N-1),$$

$$h^{(k,l)}_{ij} = \begin{cases} h_{k,l} = 1 & \\ h_{l,k} = 1 & \\ h_{i,j} = 0 & \text{otherwise} \end{cases} \qquad (k \neq l),$$

$$h^{(k,l)}_{ij} = \begin{cases} h_{k,l} = i & \\ h_{l,k} = -i & \\ h_{i,j} = 0 & \text{otherwise} \end{cases} \qquad (k \neq l).$$

Examples of each type, for $N = 3$, are

$$h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad ; \quad h = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad ; \quad h = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

eigenvalues. Since the sum of the two eigenvalues must be 1, then on the boundary, where one of them is zero, the second must be 1, and we therefore have a pure state. However, when we have $N > 2$, we have more than two eigenvalues, and hence if one eigenvalue becomes zero, it does *not* necessarily mean that one of the other eigenvalues becomes 1 while all the rest are zero. Rather, it just means that the sum of all the rest must be 1. Thus, the systems described by the boundary (for $N > 2$), are not pure states necessarily, although all the pure states must be on the boundary (since only there some of the eigenvalues are zero).

Note, that although the density matrix defines unambiguously the results of measurements, several different physical systems may give rise to the same density matrix. This is shown in the next subsection.

2.3.2. preparation of mixtures. As was mentioned before, there are two basic cases in which we must use density matrices, when we lack information on the system, and when studying only part of the system. We shall now elaborate on the first of these two cases (lack of knowledge), while the second case (studying only part of the system) will be discussed in the next subsection (2.3.3). The lack of knowledge is represented here by the use of probabilities. Since we do not know which state the system is in, we give a proper probability for each possible state to occur.

Assume two sources, A and B, of particles. Source A produces particles in random states[16] $\{p_A, \psi_A\}_{A=1}^{N_A}$, described by the density matrix ρ_A (either pure or not), and source B produces particles in random states $\{q_B, \varphi_B\}_{B=1}^{N_B}$, described by the density matrix ρ_B (again, either pure or not). Now, we want to create a new set of states. We do this by picking states either out of source A with probability λ, or else, out of source B (with probability $1-\lambda$).

As a result of picking states in the above manner, we can now describe the new collection of states as[17]

$$\{\lambda \cdot p_A, \psi_A\} \cup^* \{(1-\lambda)\cdot q_B, \varphi_B\},$$

which, by the definition of the density matrix, gives us[18]

$$\begin{aligned}\rho_{AB} &= \sum_A \lambda \cdot p_A |\psi_A\rangle\langle\psi_A| + \sum_B (1-\lambda)\cdot q_B |\varphi_B\rangle\langle\varphi_B| \\ &= \lambda \sum_A p_A |\psi_A\rangle\langle\psi_A| + (1-\lambda)\sum_B q_B |\varphi_B\rangle\langle\varphi_B|,\end{aligned}$$

which is simply[19]

$$\rho_{AB} = \lambda \cdot \rho_A + (1-\lambda)\cdot \rho_B.$$

Let us now check that the new matrix ρ_{AB} is indeed a density matrix. First, its trace is indeed 1

$$\mathrm{Tr}\,\rho_{AB} = \lambda\,\mathrm{Tr}\,\rho_A + (1-\lambda)\,\mathrm{Tr}\,\rho_B = \lambda\cdot 1 + (1-\lambda)\cdot 1 = 1.$$

[16] Recall, that here $\{p_A, \psi_A\}$, stands for a set of states $\{\psi_A\}$ (A an index), where each state ψ_A has a probability p_A of occurring.

[17] The symbol $\cup^*$ stands for a *disjoint union*. A disjoint union, is a union which keeps track of the set an element came from, and distinguishes between elements also on this basis. Thus, in our case here, even if we have some A and B, such that $\lambda \cdot p_A = (1-\lambda)\cdot q_B$ and $\psi_A = \varphi_B$, our new joined set will include both (once from A and once from B). In a regular union, the two identical occurrences would be reduced to a single occurrence.

[18] In the new collection of states there is a chance $\lambda \cdot p_A$ for states $|\psi_A\rangle$ to occur and a probability $(1-\lambda)\cdot q_B$ for states $|\varphi_B\rangle$ to occur. Since $\sum_A \lambda \cdot p_A + \sum_B (1-\lambda)\cdot q_B = 1$ (because $\sum_A p_A = 1$, $\sum_B p_B = 1$ and $\lambda + (1-\lambda) = 1$), then we can treat the new joined collection of states as a single set of the form $\{p_k, \psi_k\}$, where $p_k \in \{\lambda \cdot p_A\} \cup^* \{(1-\lambda)\cdot p_B\}$ and $\psi_k \in \{\psi_A\} \cup^* \{\varphi_B\}$.

[19] A sum of the form

$$u = \lambda v_1 + (1-\lambda) v_2 \quad (0 \le \lambda \le 1),$$

is called a *convex sum* (of v_1 and v_2). We say that a space is a *convex space*, if all possible convex sums of all possible pairs (of elements in the space), belong to this same space.

It is clearly Hermitian (λ is real)

$$\rho_{AB}^{\dagger} = \lambda\rho_A^{\dagger} + (1-\lambda)\rho_B^{\dagger} = \lambda\rho_A + (1-\lambda)\rho_B = \rho_{AB}.$$

Finally, it is clearly positive, since it is a sum of positive matrices [and $\lambda, (1-\lambda) \geq 0$]

$$\langle\psi|\rho_{AB}|\psi\rangle = \lambda\langle\psi|\rho_A|\psi\rangle + (1-\lambda)\langle\psi|\rho_B|\psi\rangle \geq 0.$$

We therefore see that ρ_{AB} is indeed a density matrix. Thus, we have been able to create a new mixture out of two others. Specifically, we could also use two sources of pure states (each), and create form them a new non-pure mixture, by the method above.

It is important to note, that two *physically different* sources may give the same density matrix. For example, assume that a source A emits particles in state $|0\rangle$ with probability of 50% ($p_0 = 0.5$) and particles in state $|1\rangle$ also with a probability of 50%. We would therefore describe such a system by the density matrix

$$\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\mathbb{1}.$$

Now, assume that we also have a source B which emits particles in state $|+\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ with probability of 50% ($p_+ = 0.5$) and particles in state $|-\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ also with a probability of 50%. We would therefore describe such a system by the density matrix

$$\begin{aligned}
\rho_B &= \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| \\
&= \frac{1}{2}\left[\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|) + \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\frac{1}{\sqrt{2}}(\langle 0| - \langle 1|)\right] \\
&= \frac{1}{2}\left[|0\rangle\langle 0| + |1\rangle\langle 1|\right] \\
&= \frac{1}{2}\mathbb{1}.
\end{aligned}$$

We see that although both sources (A and B) are physically different, we get the same density matrix in both.[20] Finding the same density matrix in both cases means that we cannot distinguish between the two sources (by performing measurements).

The above result is just a particular case of the following general rule: For every *non-pure* density matrix (of any dimension) there is an infinite number of physical systems which give that same density matrix.

PROOF. We shall first prove this for the Bloch sphere (2×2 density matrices). As we have already seen 2×2 density matrices may be represented by a polarization vector $\vec{p}$ according to

$$\rho(\vec{p}) = \frac{1}{2}(\mathbb{1} + \vec{p}\cdot\vec{\sigma}) \quad (|\vec{p}| \leq 1).$$

Now, if $\vec{p}$ can be written as

$$\vec{p} = \lambda\hat{n}_1 + (1-\lambda)\hat{n}_2 \quad (0 \leq \lambda \leq 1),$$

then $\rho(\vec{p})$ also obeys

$$\rho(\vec{p}) = \lambda\rho(\hat{n}_1) + (1-\lambda)\rho(\hat{n}_2) \quad (0 \leq \lambda \leq 1),$$

where $\rho(\hat{n}_i)$, just like $\rho(\vec{p})$, is defined as $\rho(\hat{n}_i) = \frac{1}{2}(\mathbb{1} + \hat{n}_i \cdot \vec{\sigma})$. Note, that since $\hat{n}_i$ are unit vectors (on the surface of the Bloch sphere), then $\rho(\hat{n}_i)$ represent pure states. Therefore, if indeed $\vec{p} = \lambda\hat{n}_1 + (1-\lambda)\hat{n}_2$, then $\rho(\vec{p})$ describes a physical system having the pure states $\rho(\hat{n}_1)$ and $\rho(\hat{n}_2)$ occurring with probabilities λ and $1-\lambda$ respectively.

Now, the question is how many possible sets $\{\hat{n}_1, \hat{n}_2, \lambda\}$ exist such that $\vec{p} = \lambda\hat{n}_1 + (1-\lambda)\hat{n}_2$. It is easy to show that for $|\vec{p}| < 1$ (non-pure states) there is an infinite number of such sets.

[20] In fact every choice of two orthonormal states with equal probability to occur will give us the same density matrix $\rho = \frac{1}{2}\mathbb{1}$.

Showing this will prove that there is indeed an infinite number of different physical systems which give the same non-pure density matrix.

The case $|\vec{p}| < 1$ has an infinite number of sets $\{\hat{n}_1, \hat{n}_2, \lambda\}$, since every $\hat{n}_1$ belongs to such a set. To show this, simply pick an arbitrary unit vector $\hat{n}_1$. If we now draw a straight line passing through $\hat{n}_1$ and $\vec{p}$, the second point at which this line intersects the surface of the Bloch sphere is at $\hat{n}_2$: The point is on the surface of the Bloch sphere, so it must be a unit vector, and since $\vec{p}$ lies on the line between $\hat{n}_1$ and $\hat{n}_2$, then there necessarily is some λ such that $\vec{p} = \lambda\hat{n}_1 + (1-\lambda)\hat{n}_2$. Since, for a given $\vec{p}$, we chose an arbitrary $\hat{n}_1$ and found a matching $\hat{n}_2$, then there must be an infinite number of pairs (one for each possible $\hat{n}_1$).

Note, that for $|\vec{p}| = 1$, the above logic does not apply, since using the above method will give us $\hat{n}_2 = \vec{p}$.

One might claim that the above proof is missing the possibility of more than two unit vectors $\hat{n}_i$ and the possibility of generating $\vec{p}$ using vectors $\vec{p}_i$ with $|\vec{p}_i| < 1$. This is indeed so, however, all we wanted to prove was that there is an infinite number of physical systems which give the same non-pure density matrix. We succeeded in this, even though there are more possibilities than the ones covered in the proof.

Although the proof given so far has been for a two dimensional Hilbert space, it also applies for any higher dimension Hilbert space. This is easily seen, since one may always examine just a two-dimension subspace of the larger Hilbert space and use for it, the result proved here. □

The fact that we have an infinite number of ways to create the same mixture in quantum mechanics, is markedly different from the situation in the classical physics where there is only one possible way.

2.3.3. combined systems, partial trace, and the reduced matrix. We have so far seen that density matrices arise from (random) ensembles of initial states. We shall now see that they can also arise when we study only part of a system which, as a whole, is in a pure state. Before we do this, however, we must know how to describe a state of two (or more) particles.

2.3.3.1. *Tensor product (combining a number systems into one).* Assume two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ (not necessarily of the same dimension), each with its own *basis* of states:

$$|\psi_i^A\rangle \in \mathcal{H}_A \quad (i = 1, 2, \ldots, N_A),$$

$$|\varphi_j^B\rangle \in \mathcal{H}_B \quad (j = 1, 2, \ldots, N_B).$$

We define the *tensor product* (also known as *direct product* or *outer product*) of the two spaces as

$$\mathcal{H}_{A\otimes B} = \mathcal{H}_A \otimes \mathcal{H}_B = \mathrm{span}\{|\psi_i^A\rangle \otimes |\varphi_j^B\rangle\}.$$

If the original spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ had N_A and N_B dimensions respectively, then the new space $\mathcal{H}_{A\otimes B}$ has $N_A \cdot N_B$ dimensions: Any state in the new space is described by the quantum number of $\mathcal{H}_A$ (N_A different possibilities) *and* the quantum number of $\mathcal{H}_B$ (N_B different possibilities for each choice of quantum number from $\mathcal{H}_A$).

To complete the definition of the tensor product, we must give two more of its features:

- The tensor product is linear in the complex coefficients appearing in each space, i.e.
$$[\alpha|\psi_i^A\rangle] \otimes [\beta|\varphi_j^B\rangle] = \alpha\beta[|\psi_i^A\rangle \otimes |\varphi_j^B\rangle].$$
- The tensor product is distributive

$$\begin{aligned}[\alpha_1|\psi_1^A\rangle + \alpha_2|\psi_2^A\rangle] \otimes [\beta_1|\varphi_1^B\rangle + \beta_2|\varphi_2^B\rangle] \quad = \quad & \alpha_1\beta_1|\psi_1^A\rangle \otimes |\varphi_1^B\rangle + \alpha_1\beta_2|\psi_1^A\rangle \otimes |\varphi_2^B\rangle \\ & + \alpha_2\beta_1|\psi_2^A\rangle \otimes |\varphi_1^B\rangle + \alpha_2\beta_2|\psi_2^A\rangle \otimes |\varphi_2^B\rangle.\end{aligned}$$

Note, that we shall usually drop the $\otimes$ symbol between states and simply write

$$|\psi_i^A\rangle|\varphi_j^B\rangle,$$

or more often

$$|\psi_i\rangle_A|\varphi_j\rangle_B,$$

instead of $|\psi_i^A\rangle\otimes|\varphi_j^B\rangle$. In some cases, we shall even drop the indices A,B and use the order of the kets to describe which state belongs to what space.

The tensor product can also be written in matrix form. If for example

$$|\psi\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A \equiv \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

$$|\varphi\rangle_B = a|0\rangle_B + b|1\rangle_B \equiv \begin{pmatrix} a \\ b \end{pmatrix},$$

then

$$|\psi\rangle_A\otimes|\varphi\rangle_B = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\otimes\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha\begin{pmatrix} a \\ b \end{pmatrix} \\ \beta\begin{pmatrix} a \\ b \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \alpha a \\ \alpha b \\ \beta a \\ \beta b \end{pmatrix}.$$

And for operators/matrices, we would have (as an example)

$$A\otimes B = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\otimes\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \alpha\begin{pmatrix} a & b \\ c & d \end{pmatrix} & \beta\begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ \gamma\begin{pmatrix} a & b \\ c & d \end{pmatrix} & \delta\begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \alpha a & \alpha b & \beta a & \beta b \\ \alpha c & \alpha d & \beta c & \beta d \\ \gamma a & \gamma b & \delta a & \delta b \\ \gamma c & \gamma d & \delta c & \delta d \end{pmatrix}.$$

2.3.3.2. *The partial trace and the reduced matrix.* Imagine that we have some state $|\psi_{AB}\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$ and an operator O_{AB} of the form

$$O_{AB} = A\otimes\mathbb{1}_B,$$

where A operates on the degrees of freedom of $\mathcal{H}_A$, and $\mathbb{1}_B$ (the identity in $\mathcal{H}_B$) operates on the degrees of freedom of $\mathcal{H}_B$. Let the density matrix describing the pure system (in the $\mathcal{H}_A\otimes\mathcal{H}_B$ space) be $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$. The result of measuring O_{AB} (according to the usual rules of quantum mechanics and density matrices) is given by

$$\begin{aligned}\langle A\rangle_{AB} &\equiv \langle O_{AB}\rangle = \langle\psi_{AB}|O_{AB}|\psi_{AB}\rangle \\ &= \mathrm{Tr}(O_{AB}\,\rho_{AB}) \\ &= \mathrm{Tr}_A\,\mathrm{Tr}_B\left[O_{AB}\,\rho_{AB}\right] \\ &= \mathrm{Tr}_A\left[A\,\mathrm{Tr}_B(\mathbb{1}_B\,\rho_{AB})\right] \\ &= \mathrm{Tr}_A\left[A\,\mathrm{Tr}_B(\rho_{AB})\right],\end{aligned}$$

where Tr_A and Tr_B mean performing the trace only in $\mathcal{H}_A$ and only in $\mathcal{H}_B$ respectively (explained further below). Note, that after taking the trace over B, the operator $\mathrm{Tr}_B(\rho_{AB})$ now operates solely on $\mathcal{H}_A$, and we can therefore drop the tensor product $\otimes$. If we now define

$$\rho_A \equiv \mathrm{Tr}_B(\rho_{AB}),$$

then we may rewrite the previous equation as

$$\langle A\rangle_{AB} = \mathrm{Tr}_A\left[A\,\rho_A\right].$$

This last result is of the standard form $\langle O\rangle = \mathrm{Tr}\left[O\,\rho\right]$, however limited to the $\mathcal{H}_A$ Hilbert space (instead of $\mathcal{H}_A\otimes\mathcal{H}_B$.

We call the process of tracing over a subspace of our system (Tr_B in this case) a *partial trace*. The resulting density matrix $\rho_A \equiv \mathrm{Tr}_B\,\rho_{AB}$ is called the *reduced density matrix*.

To clarify what a partial trace and what a reduced density matrix are, let us repeat the above calculation more explicitly. We shall work with the two orthonormal bases, $|n\rangle_A$ of $\mathcal{H}_A$ and $|m\rangle_B$ of $\mathcal{H}_B$, thus $\langle A\rangle_{AB}$ is

$$\begin{aligned}\langle A\rangle_{AB} &= \mathrm{Tr}_{AB}\left[(A\otimes \mathbb{1}_B)\rho_{AB}\right] = \sum_{n=1}^{N_A}\sum_{m=1}^{N_B} {}_A\langle n|_B\langle m|\left[(A\otimes \mathbb{1}_B)\rho_{AB}\right]|m\rangle_B|n\rangle_A \\ &= \sum_{n=1}^{N_A} {}_A\langle n|A\left(\sum_{m=1}^{N_B} {}_B\langle m|\rho_{AB}|m\rangle_B\right)|n\rangle_A \\ &\equiv \sum_{n=1}^{N_A} {}_A\langle n|A\left(\mathrm{Tr}_B\,\rho_{AB}\right)|n\rangle_A \\ &\equiv \mathrm{Tr}_A\left[A\left(\mathrm{Tr}_B\,\rho_{AB}\right)\right] \equiv \mathrm{Tr}_A\left[A\rho_B\right].\end{aligned}$$

Another way, equivalent to the last, of viewing this, is to denote ρ_{AB} as having four indices instead of just two:

$$\rho^{AB}_{i,m,j,n} \equiv {}_A\langle i|_B\langle j|\rho_{AB}|m\rangle_A|n\rangle_B.$$

In this case $\rho_A = \mathrm{Tr}_B\,\rho_{Ab}$ simply becomes (note the double index m)

$$(\rho_A)_{i,j} = (\mathrm{Tr}_B\,\rho_{Ab})_{i,j} = \sum_m \rho^{AB}_{i,m,j,m}.$$

We have so far defined a partial trace and a reduced matrix, but what do they give us? What we have seen is that they give us a method to find the expectation values for a subsystem (here subsystem A of AB). Note, that when we start with a pure state in the Hilbert space $\mathcal{H}_A\otimes H_B$, we will generally end up with a *non-pure* density matrix ρ_A. Thus, by studying only part of a system, an actually pure state will generally give rise to a seemingly non-pure state (it is non-pure for all practical purposes for a person living in the Hilbert space $\mathcal{H}_A$). It is the lack of knowledge about the rest of the system (B), which gives rise to a mixed state with respect to our subsystem A. Note, that although we found a density matrix ρ_A to describe A, the physical situation is not that of a random source in system A—unlike the case we had in the previous subsection.[21]

Of course, the whole discussion made here could have started with a mixed state in $\mathcal{H}_A\otimes\mathcal{H}_B$, with hardly any change. However, the point here was to see how starting with a *pure* state and performing a partial trace gives a non-pure density matrix.

2.3.4. Effects of measurements on the reduced matrix (selective and non-selective measurements). Let us now see how the reduced density matrix ρ_A is effected when one performs a measurement in the subsystem $\mathcal{H}_B$. For convenience we shall introduce Alice and Bob again: Alice has access to subsystem A and Bob has access to subsystem B. There are two cases which will be considered here. The first is when Bob makes a *selective measurement* in subsystem B, i.e. Bob makes a measurement, but Alice may measure her subsystem only if Bob got a desired result (we calculate ρ_A just for a certain result of Bob's measurement). The second case is when Bob makes a *nonselective* measurement: he makes a measurement (on $\mathcal{H}_B$), and regardless of the result, Alice may perform measurements on her subsystem A. As we shall see, in the first case (selective measurement) the resulting density matrix ρ_A may differ (from the case of no measurements done by Bob) only if we start with an entangled state (defined later), while in the second case ρ_A will not differ, regardless of the system we start with.

To see that reduced density matrices may be influenced by measurements, let us start with a simple example. Assume a system described by the state

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right),$$

[21] However, as we shall see later, if a non-selective measurement is performed on $\mathcal{H}_B$, then the situation is physically the same as a random source in system A (see 2.3.4).

where $|0\rangle$ and $|1\rangle$ stand respectively for the eigenstates of spin up and spin down in the z-direction (of a spin $\frac{1}{2}$ particle). If no measurement is made, the reduced density ρ_A is simply

$$\begin{aligned} \rho_A &= \mathrm{Tr}_B\, \rho_{AB} = {}_B\langle 0|\rho_{AB}|0\rangle_B + {}_B\langle 1|\rho_{AB}|1\rangle_B \\ &= \frac{1}{2}\left(|0\rangle_{AA}\langle 0| + |1\rangle_{AA}\langle 1|\right), \end{aligned}$$

where

$$\rho_{AB} = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right)\left({}_B\langle 0|_A\langle 0| + {}_B\langle 1|_A\langle 1|\right).$$

Now, let us assume that Bob measures the z-component of the spin in his subsystem. Let us further assumes that he finds the spin in the up direction. In this case the state $|\Psi_{AB}\rangle$ collapses to

$$|\Psi_{AB}\rangle \xrightarrow{|0\rangle_B \text{ measured}} |\Psi_{AB}^{(0)}\rangle = |0\rangle_A|0\rangle_B.$$

The reduced density matrix $\rho_A^{(0)}$ of this new state is easily seen to be

$$\rho_A^{(0)} = |0\rangle_{AA}\langle 0|,$$

which is clearly different from the reduced matrix ρ_A we found previously.

We have just seen that performing a measurement on one subsystem of the complete system AB may influence the reduced density matrix. As we shall prove, the new density matrix $\rho_A^{(b)}$ (when the value b is measured in subsystem B) may differ from the original ρ_A (when no measurement was done) only if the original $|\Psi_{AB}\rangle$ may *not* be written in the form

$$|\Psi_{AB}\rangle = |\psi\rangle_A|\varphi\rangle_B$$

(with no sum on the righthand side). When $|\Psi_{AB}\rangle$ may *not* be written in this form, the state is said to be *entangled*.[22] Otherwise (when $|\Psi_{AB}\rangle = |\psi\rangle_A|\varphi\rangle_B$) it is said to be *nonentangled*. Using this notation, the above claim may be restated as

$$\rho_A^{(\text{no measurement})} \neq \rho_A^{(b\text{ measured})} \Rightarrow |\Psi_{AB}\rangle \text{ entangled}.$$

Note, that the opposite direction does not hold.

PROOF. All we need to prove is that when $|\Psi_{AB}\rangle$ equals $|\psi\rangle_A|\varphi\rangle_B$, then regardless of the measurement made, the reduced matrix is the same.

To see this, assume a state $|\Psi_{AB}\rangle$ of the complete system which is nonentangled, i.e. $|\Psi_{AB}\rangle = |\psi\rangle_A|\varphi\rangle_B$. Now, let us further assume that Bob performs a measurement (in subsystem B) and finds a result b. We know that in this case the original state $|\Psi_{AB}\rangle$ collapses to

$$|\Psi_{AB}^{(b)}\rangle = \frac{1}{\sqrt{{}_B\langle\varphi|\Pi_b|\varphi\rangle_B}}|\psi\rangle_A(\Pi_b|\varphi\rangle_B),$$

where Π_b is a projection onto the subspace of the b eigenvalue. Thus the new density matrix (of the complete system) is

$$\rho_{AB}^{(b)} = \frac{1}{{}_B\langle\varphi|\Pi_b|\varphi\rangle_B}\Pi_b|\varphi\rangle_B|\psi\rangle_{AA}\langle\psi|_B\langle\varphi|\Pi_b.$$

Performing the partial trace on this density matrix is very simple, we just choose an orthonormal basis of $\mathcal{H}_B$ which includes the state $\frac{1}{\sqrt{{}_B\langle\varphi|\Pi_b|\varphi\rangle_B}}(\Pi_b|\varphi\rangle_B)$ as one of its elements. Thus we get

$$\rho_A^{(b)} = |\psi\rangle_{AA}\langle\psi|,$$

which is the same result we would have found if no measurement was made. In other words we found

$$\rho_A^{(b)} = \rho_A \quad (\text{when } |\Psi_{AB}\rangle \text{ nonentangled}).$$

[22] Entanglement is further discussed in 2.4.

This completes our proof: only if $|\Psi_{AB}\rangle$ is entangled ($|\Psi_{AB}\rangle \neq |\psi\rangle_A|\phi\rangle_B$) do we have a chance of finding $\rho_A^{(b)} \neq \rho_A$. □

We have seen above that if Bob makes selective measurements on his part of the system, then this may effect Alice's measurements. We now want to see what would happen if Bob still performs his measurement, but no matter what he gets, he allows Alice to perform her measurements as well. In this case, no matter the state $|\Psi_{AB}\rangle$ which we start with (either entangled or not), Alice won't know the difference, and would find the same density matrix as if Bob made no measurements, i.e.

$$\rho_A^{(\text{no measurement})} = \rho_A^{(\text{nonselective})}.$$

PROOF. The most general case of a (pure) state $|\Psi_{AB}\rangle$ of the complete system is a state of the form

$$|\Psi_{AB}\rangle = \sum_i \alpha_i |\psi_i\rangle_A |\phi_i\rangle_B,$$

where $\{|\psi_i\rangle_A\}$ and $\{|\phi_i\rangle_B\}$ are two arbitrary sets of states (the states in each are not necessarily orthogonal, and might even include repetitions). Let us first find the reduced density matrix appropriate to this state when no measurement is made. Such a state gives the density matrix (not reduced yet)

$$\begin{aligned}\rho_{AB} &= \left(\sum_i \alpha_i |\psi_i\rangle_A |\phi_i\rangle_B\right)\left(\sum_j \alpha_j^* {}_A\langle\psi_j|_B\langle\phi_j|\right)\\ &= \sum_{i,j} \alpha_i\alpha_j^* |\psi_i\rangle_A|\phi_i\rangle_{BB}\langle\phi_j|_A\langle\psi_j|,\end{aligned}$$

and therefore (when no measurement is made) using $\sum |n\rangle_{BB}\langle n| = \mathbb{1}_B$, the reduced matrix ρ_A is

$$\begin{aligned}\rho_A &= \mathrm{Tr}_B\,\rho_{AB} = \sum_n {}_B\langle n|\left(\sum_{i,j}\alpha_i\alpha_j^*|\psi_i\rangle_A|\phi_i\rangle_{BB}\langle\phi_j|_A\langle\psi_j|\right)|n\rangle_B\\ &= \sum_{i,j}\alpha_i\alpha_j^*|\psi_i\rangle_A\left(\sum_n {}_B\langle\phi_j|n\rangle_{BB}\langle n|\phi_i\rangle_B\right){}_A\langle\psi_j|\\ &= \sum_{i,j}\alpha_i\alpha_j^*{}_B\langle\phi_j|\phi_i\rangle_B|\psi_i\rangle_{AA}\langle\psi_j|.\end{aligned}$$

or simply

$$\rho_A^{(\text{no measurement})} = \sum_{i,j}\alpha_i\alpha_j^*{}_B\langle\phi_j|\phi_i\rangle_B|\psi_i\rangle_{AA}\langle\psi_j|.$$

Now, Let us turn to the case of a measurement. Assume that Bob makes a measurement of operator B and gets, with some probability p_b, a result b. In such a case $|\Psi_{AB}\rangle$ will collapse as follows:

$$|\Psi_{AB}\rangle = \sum_i \alpha_i|\psi_i\rangle_A|\phi_i\rangle_B \xrightarrow[(b\text{ measured})]{\text{collapse}} |\Psi_{AB}^{(b)}\rangle = \frac{1}{\sqrt{p_b}}\sum_i \alpha_i|\psi_i\rangle_A(\Pi_b|\phi_i\rangle_B),$$

where, as before, Π_b is a projection onto the subspace of the b eigenvalue, and where $\frac{1}{\sqrt{p_b}}$ is a normalization factor (recall that p_b is the probability to measure b). The reduced matrix $\rho_A^{(b)}$ for the new state $|\Psi_{AB}^{(b)}\rangle$ is then

$$\rho_A^{(b)} = \frac{1}{p_b}\sum_{i,j}\alpha_i\alpha_j^*\left(\sum_n {}_B\langle n|\Pi_b|\phi_i\rangle_{BB}\langle\phi_j|\Pi_b|n\rangle_B\right)|\psi_i\rangle_{AA}\langle\psi_j| \quad (\text{with probability } p_b).$$

Which using $\sum |n\rangle_{BB}\langle n| = \mathbb{1}_B$, simplifies to

$$\rho_A^{(b)} = \frac{1}{p_b}\sum_{i,j}\alpha_i\alpha_j^*\left({}_B\langle\phi_j|\Pi_b|\phi_i\rangle_B\right)|\psi_i\rangle_{AA}\langle\psi_j| \quad (\text{with probability } p_b).$$

Recall however, that we are interested in nonselective measurements. If Bob makes a nonselective measurement, then by definition[23] the density matrix describing the system after the measurement is

$$\rho_{AB}^{(\text{nonselective})} = \sum_b p_b |\Psi_{AB}^{(b)}\rangle\langle\Psi_{AB}^{(b)}|,$$

and the partial trace over this density matrix is simply

$$\rho_{A}^{(\text{nonselective})} = \mathrm{Tr}_B\left(\sum_b p_b |\Psi_{AB}^{(b)}\rangle\langle\Psi_{AB}^{(b)}|\right) = \sum_b p_b \rho_A^{(b)}.$$

Replacing $\rho_A^{(b)}$with the previous result we found, the reduced density matrix $\rho_A^{(\text{nonselective})}$ may be written as

$$\begin{aligned}
\rho_{A}^{(\text{nonselective})} &= \sum_b p_b \left[\frac{1}{p_b}\sum_{i,j}\left(\alpha_i\alpha_j^*{}_B\langle\varphi_j|\Pi_b|\varphi_i\rangle_B\right)|\psi_i\rangle_{AA}\langle\psi_j|\right] \\
&= \sum_{i,j}\alpha_i\alpha_j^*{}_B\langle\varphi_j|\left(\sum_b \Pi_b\right)|\varphi_i\rangle_B|\psi_i\rangle_{AA}\langle\psi_j| \\
&= \sum_{i,j}\left(\alpha_i\alpha_j^*{}_B\langle\varphi_j|\varphi_i\rangle_B\right)|\psi_i\rangle_{AA}\langle\psi_j|.
\end{aligned}$$

If we compare this result to $\rho_A^{(\text{no measurement})}$ which we found earlier (when no measurement was made), we see that we have found, as claimed

$$\rho_{A}^{(\text{nonselective})} = \rho_{A}^{(\text{no measurement})}.$$

□

Before going on one should notice that although we have

$$\rho_{A}^{(\text{nonselective})} = \rho_{A}^{(\text{no measurement})},$$

the two cases are physically different. When measurements are made nonselectively, the physical situation is indeed that of states $|\Psi_{AB}^{(b)}\rangle$ occurring with probabilities p_b. However, when no measurement is made there are no such states (with different probability). In this latter case, the density matrix is simply the result of lack of knowledge about system B.

2.3.5. The GHJW[24] theorem[25].

As we saw before, physically different systems may give rise to the same density matrix ρ. We shall now see that all such systems (described by the same ρ, but having some complexity limit—see below) may all be derived from the *same* pure state.

Let there be two sources of states, one emitting states $|\psi_i\rangle$ with probability p_i ($\{p_i, |\psi_i\rangle\}_{i=1}^{n_1}$) and a second emitting states $|\varphi_j\rangle$ with probability q_j ($\{q_j, |\varphi_j\rangle\}_{j=1}^{n_2}$). We shall say that the

[23] Given the set of probabilities and states $\{p_b, |\psi_b\rangle\}$, the density matrix is defined as

$$\rho \equiv \sum_b p_b|\psi_b\rangle\langle\psi_b|.$$

Here the probability of measuring b is p_b and after measuring b the state of the complete system is $|\Psi_{AB}^{(b)}\rangle$. Thus, an observer standing after the measurement apparatus sees states $|\Psi_{AB}^{(b)}\rangle$ with probability p_b.

[24] GHJW stands for Gisin, Hughston, Jozsa and Wooters.

[25] This subsection was originally taught after "entanglement" and the "Schmidt decomposition" were taught. It was moved here, because the material presented seemed to conceptually fit better right after the discussion of density matrices and the partial trace. As a consequence, the use of the Schmidt decomposition is given without proof. The proof and further material are given later when entanglement and the Schmidt decomposition are discussed (subsection 2.4). Of course, the necessary traits (for this subsection) of the Schmidt decomposition are described here.

two sources/systems are two different *realizations* (of the density matrix ρ_A) if both sources give the *same* density matrix ρ_A, i.e. if

$$\sum_{i=1}^{n_1} p_i|\psi_i\rangle\langle\psi_i| = \sum_{j=1}^{n_2} q_j|\varphi_j\rangle\langle\varphi_j| = \rho_A \quad \text{(two realizations of } \rho_A\text{)}.$$

Note that the two sets $\{|\psi_i\rangle\}$ and $\{|\varphi_j\rangle\}$ are not necessarily sets of orthogonal states (within the sets, or between sets), nor necessarily have the same number of elements.

Gisin, Hughston, Jozsa and Wooters have shown (the *GHJW theorem*) that all realizations of the same density matrix ρ_A, consisting of up to n pure states ($i = 1,\dots,n$),[26] may be produced from a single pure state $|\Psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_B$ is at least n dimensional. The different realizations are produced by measuring $|\Psi\rangle_{AB}$ nonselectively, using (for each realization) a suitable observable in the $\mathcal{H}_B$ space.

PROOF. Assume two realizations of the *same* density matrix ρ_A

$$\{p_i,|\psi_i\rangle\}_{i=1}^{n_1} \Rightarrow \rho_A = \sum_{i=1}^{n_1} p_i|\psi_i\rangle\langle\psi_i|,$$

$$\{q_i,|\varphi_i\rangle\}_{i=1}^{n_2} \Rightarrow \rho_A = \sum_{i=1}^{n_2} q_i|\varphi_i\rangle\langle\varphi_i|.$$

We perform a so called "*purification*" of the two by enlarging our Hilbert space to $\mathcal{H}_A \otimes \mathcal{H}_B$

$$\{p_i,|\psi_i\rangle\}_{i=1}^{n_1} \xrightarrow{\text{purification}} |\Psi\rangle_{AB} = \sum_{i=1}^{n_1} \sqrt{p_i}|\psi_i\rangle_A|\beta_i\rangle_B,$$

$$\{q_i,|\varphi_i\rangle\}_{i=1}^{n_2} \xrightarrow{\text{purification}} |\Phi\rangle_{AB} = \sum_{i=1}^{n_2} \sqrt{q_i}|\varphi_i\rangle_A|\beta_i\rangle_B,$$

where in both cases we use the same orthonormal states $|\beta_i\rangle_B \in \mathcal{H}_B$

$${}_B\langle\beta_i|\beta_j\rangle_B = \delta_{ij} \quad (i,j = 1,2,\dots,n;\, n \geq \max(n_1,n_2)\,).$$

The outline of the proof form here on is as follows. The proof will consist of two steps, the first acting as a motivation to the next. We shall start by proving that given a unitary transformation U_B such that $U_B|\Psi\rangle_{AB} = |\Phi\rangle_{AB}$, then by measuring $|\Psi\rangle_{AB}$ nonselectively we may reproduce (using the proper measurement) either of the two realizations $\{p_i,|\psi_i\rangle\}_{i=1}^{n_1}$ and $\{q_i,|\varphi_i\rangle\}_{i=1}^{n_2}$. Having shown that, we shall go on to prove in the next step that such a transformation indeed always exists, thus completing the proof: Since $|\Phi\rangle_{AB}$ arose from an arbitrary[27] realization of ρ_A, and further more, a proper measurement of $|\Psi\rangle_{AB}$ gave the source/realization corresponding to $|\Phi\rangle_{AB}$, then an appropriate measurement exists for *any* realization (of up to n possible states) of ρ_A.

LEMMA (first step). *If there exists a unitary operator U_B (operating on $\mathcal{H}_B$), such that*

$$U_B|\Psi\rangle_{AB} = |\Phi\rangle_{AB},$$

then by a proper nonselective measurement of $|\Psi\rangle_{AB}$ one may reproduce the source corresponding to $|\Phi\rangle_{AB}$ (i.e. $\{q_i,|\varphi_i\rangle\}_{i=1}^{n_2}$).

PROOF. Assume an observable B (operating on $\mathcal{H}_B$) whose eigenstates are the states $|\beta_i\rangle_B$ and whose eigenvalues are *nondegenerate* (there is an infinite choice of such B's—just choose one). Clearly, performing a *nonselective* measurement (see 2.3.4) of the states $|\Psi\rangle_{AB}$ and $|\Phi\rangle_{AB}$ using this operator B will give results physically equivalent (for an observer in A) to the two sources ($\{p_i,|\psi_i\rangle\}$ and $\{q_i,|\varphi_i\rangle\}$ respectively) with which we started.

[26] This limit on n is the complexity limit mentioned above (at the start of the subsection).

[27] Arbitrary, apart for the restriction $n_2 \leq n$.

Now, by the assumption of the lemma, let U_B be a unitary state such that

$$U_B|\Psi\rangle_{AB} = |\Phi\rangle_{AB},$$

or equivalently

$$|\Psi\rangle_{AB} = U_B^{-1}|\Phi\rangle_{AB}.$$

Expanding the last equation using $\{|\varphi_i\rangle\}$ and $\{|\psi_j\rangle\}$ this gives

$$|\Psi\rangle_{AB} = \sum_i \sqrt{p_i}|\psi_i\rangle_A|\beta_i\rangle_B = \sum_j \sqrt{q_j}|\varphi_j\rangle_A U_B^{-1}|\beta_j\rangle,$$

or simply

$$|\Psi\rangle_{AB} = \sum_j \sqrt{q_j}|\varphi_j\rangle_A|\gamma_i\rangle_B,$$

where we have defined

$$|\gamma_i\rangle_B \equiv U_B^{-1}|\beta_i\rangle \quad (i = 1,\dots,n \geq \max(n_1,n_2)).$$

Since U_B is unitary and $|\beta_i\rangle_B$ is an orthonormal set, then necessarily $|\gamma_i\rangle_B$ is also an orthonormal set, i.e.

$${}_B\langle\gamma_i|\gamma_j\rangle_B = \delta_{ij}.$$

Thus, the new expression we got for $|\Psi\rangle_{AB}$ is similar in form to the one we started with:

$$|\Psi\rangle_{AB} = \sum_i \sqrt{p_i}|\psi_i\rangle_A|\beta_i\rangle_B \text{ is of the same form as } |\Psi\rangle_{AB} = \sum_j \sqrt{q_j}|\varphi_j\rangle_A|\gamma_i\rangle_B.$$

It is therefore clear that if we measure $|\Psi\rangle_{AB}$ nonselectively, using an observable (operating on $\mathcal{H}_B$) whose eigenstates are the states $|\gamma_i\rangle_B$ (instead of $|\beta_i\rangle_B$) and whose eigenvalues are nondegenerate, then the result will be physically equivalent to the (second) realization $\{q_j, |\psi_j\rangle\}_{j=1}^m$. One observable which obeys the above requirements is

$$U_B^{-1}BU_B,$$

whose eigenstates are $U_B|\beta_i\rangle_B \equiv |\gamma_i\rangle_B$ (since $|\beta_i\rangle_B$ are eigenstates of B), and has the same eigenvalues as B (and therefore nondegenerate as for B).

We have thus shown that if there exists U_B such that $U_B|\Psi\rangle_{AB} = |\Phi\rangle_{AB}$, then there exists an observable (e.g. $U_B^{-1}BU_B$) which will reduce $|\Psi\rangle_{AB}$ to the realization $\{q_j, |\varphi_j\rangle\}_{j=1}^{n_2}$ (while we already know that B will reduce $|\Psi\rangle_{AB}$ to the realization $\{p_i, |\psi_i\rangle\}_{i=1}^{n_1}$). □

Having proved the first step, we may now go on to prove the second.

LEMMA (second step). *If the partial trace (over $\mathcal{H}_B$) of both*

$$|\Psi\rangle_{AB} = \sum_{i=1}^{n_1} \sqrt{p_i}|\psi_i\rangle_A|\beta_i\rangle_B \quad \textit{and} \quad |\Phi\rangle_{AB} = \sum_{i=1}^{n_2} \sqrt{q_i}|\varphi_i\rangle_A|\beta_i\rangle_B,$$

give the same density matrix ρ_A, then there exists a unitary operator (acting only on $\mathcal{H}_B$) such that

$$U_B|\Psi\rangle_{AB} = |\Phi\rangle_{AB}.$$

PROOF. To prove this we use the Schmidt decomposition (see subsection 2.4 for more details). According to the Schmidt decomposition, $|\Psi\rangle_{AB}$ and $|\Phi\rangle_{AB}$ may always be written as

$$|\Psi\rangle_{AB} = \sum_{i=1}^{m_1} \sqrt{\lambda_i}|a_i\rangle_A|b_i\rangle_B,$$

and

$$|\Phi\rangle_{AB} = \sum_{i=1}^{m_2} \sqrt{\lambda_i'}|a_i'\rangle_A|b_i'\rangle_B,$$

where the four sets $\{|a_i\rangle_A\}$, $\{|b_i\rangle_B\}$, $\{|a_i'\rangle_A\}$ and $\{|b_i'\rangle_B\}$ are all orthonormal sets:

$${}_A\langle a_i|a_j\rangle_A = \delta_{ij} \qquad {}_B\langle b_i|b_j\rangle_B = \delta_{ij},$$

$$_A\langle a'_i|a'_j\rangle_A = \delta_{ij} \qquad _B\langle b'_i|b'_j\rangle_B = \delta_{ij}.$$

Further more, in this case, we must have

$$\lambda'_i = \lambda_i,$$

and may choose a basis such that

$$|a'_i\rangle = |a_i\rangle.$$

The reason for these requirements is that the partial trace of both states give the same *diagonal* density matrix. The elements on the diagonal of this matrix are $\lambda_i|a_i\rangle\langle a_i|$ for $|\Psi\rangle_{AB}$ and $\lambda'_i|a'_i\rangle\langle a'_i|$ for $|\Phi\rangle_{AB}$, but the diagonalization of the the density matrix is unique (up to the order of the eigenvalues and eigenstates), so we must have $\lambda'_i = \lambda_i$, and $|a'_i\rangle = |a_i\rangle$ (if λ_i, λ'_i are degenerate, then we may have $|a'_i\rangle \neq |a_i\rangle$—not just because of different ordering— but we may always choose a basis which does obey $|a'_i\rangle = |a_i\rangle$) . Thus we may write

$$|\Psi\rangle_{AB} = \sum \sqrt{\lambda_i}|a_i\rangle_A|b_i\rangle_B \quad \begin{pmatrix} _A\langle a_i|a_j\rangle_A = \delta_{ij} \\ _B\langle b_i|b_j\rangle_B = \delta_{ij} \end{pmatrix}$$

$$|\Phi\rangle_{AB} = \sum \sqrt{\lambda_i}|a_i\rangle_A|b'_i\rangle_B \quad \begin{pmatrix} _A\langle a'_i|a'_j\rangle_A = \delta_{ij} \\ _B\langle b'_i|b'_j\rangle_B = \delta_{ij} \end{pmatrix}.$$

Now, since $|b_i\rangle_B$ and $|b'_i\rangle_B$ are orthonormal bases (with the same number of elements) then there must exist a unitary transformation between them, i.e.

$$|b'_i\rangle_B = U_B|b_i\rangle_B \quad (U_B = \sum_i |b'_i\rangle_{BB}\langle b_i|).$$

Using this unitary transformation we find

$$|\Phi\rangle_{AB} = (\mathbb{1}_A \otimes U_B)|\Psi\rangle_{AB},$$

Thus we see, that if $|\Psi\rangle_{AB}$ and $|\Phi\rangle_{AB}$ give the same density matrix ρ_A, then there must exist a unitary transformation such that $|\Psi\rangle_{AB} = U_B|\Phi\rangle_{AB}$. □

Recapping the proof, we started by purifying all possible realization (of up to *n* states) of a given density matrix. We then showed that if there exist unitary transformations (acting on $\mathcal{H}_B$ alone) that transform between the purified states, then a single purified state may be used to generate all possible realizations (by performing appropriate nonselective measurements). Finally we showed that such unitary transformations exist, thus completing the proof. □

2.4. Entanglement and the Schmidt decomposition

We have already, briefly, encountered entanglement and the Schmidt decomposition, earlier. It is now time for a more methodic presentation of these two terms.

2.4.1. Entanglement. In Quantum information, entanglement is both an important tool and a subject of active research in its own right. This section gives the definition of entanglement. Much of the course will be on the uses and traits of entanglement.

A state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ is said to be *entangled* (between A and B) if it *cannot* be decomposed into a tensor product of two states, one in $\mathcal{H}_A$ and the second in $\mathcal{H}_B$ (no matter which basis we choose in $\mathcal{H}_A$ and in $\mathcal{H}_B$). Note, that we must state the spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, since a different partition of the spaces (into $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$—instead of $\mathcal{H}_A \otimes \mathcal{H}_B$) might change the state's attribute of being entangled/nonentangled.

As an example let us examine the state

$$|\psi_1\rangle_{AB} = \alpha|0\rangle_A|0\rangle_B + \beta|1\rangle_A|0\rangle_B.$$

This state is *not* entangled since we can write it as a tensor product of the state $\alpha|0\rangle_A + \beta|1\rangle_A$ in $\mathcal{H}_A$ and the state $|0\rangle_B$ in $\mathcal{H}_B$:

$$|\psi_1\rangle_{AB} = \Big(\alpha|0\rangle_A + \beta|1\rangle_A\Big) \otimes |0\rangle_B \quad \text{(nonentangled)}.$$

On the other hand, the state

$$|\psi_2\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)$$

is entangled. To see this let us first write the most general *nonentangled* state of two spin $\frac{1}{2}$ particles. Such a state (since being nonentangled means that it is a tensor product of two states) may be written as

$$\begin{aligned} |\psi\rangle_{AB} &= \Big(a|0\rangle_A + b|1\rangle_A\Big) \otimes \Big(\alpha|0\rangle_B + \beta|1\rangle_B\Big) \\ &= a\alpha|0\rangle_A|0\rangle_B + a\beta|0\rangle_A|1\rangle_B + b\alpha|1\rangle_A|0\rangle_B + b\beta|1\rangle_A|1\rangle_B. \end{aligned}$$

From this expression we see that for a nonentangled state if $a\alpha \neq 0$ and $b\beta \neq 0$, then $a\beta$ and $b\alpha$ must also be non-zero. This condition, however, is not obeyed by our state $|\psi_2\rangle_{AB}$ (our state includes $|0\rangle_A|0\rangle_B$ and $|1\rangle_A|1\rangle_B$ but not $|0\rangle_A|1\rangle_B$ and $|1\rangle_A|0\rangle_B$). As a consequence $|\psi_2\rangle_{AB}$ is not nonentangled, i.e it is entangled:

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B) \quad \text{(entangled)}.$$

The Schmidt decomposition, shown below, gives a systematic way of determining whether a state is entangled or not (see 2.4.2.2).

2.4.2. The Schmidt decomposition. The Schmidt decomposition is an often used tool in the study of quantum information and entanglement. It is basically a standardized and convenient form of writing (pure) states.

THEOREM (Schmidt decomposition). *For any state $|\Psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ there is an* orthonormal *basis $|i\rangle_A$ ($i = 1,\dots,N_A$) of $\mathcal{H}_A$ and an* orthonormal *basis $|\tilde{i}\rangle_B$ ($i = 1,\dots,N_B$) of $\mathcal{H}_B$ such that $|\Psi\rangle_{AB}$ may be written as*

$$|\Psi\rangle_{AB} = \sum_{i=1}^{N\leq\min(N_a,N_B)} \sqrt{\lambda_i}|i\rangle_A|\tilde{i}\rangle_B \quad (\lambda_i > 0),$$

$$\big({}_A\langle i|j\rangle_A = \delta_{ij} \quad ; \quad {}_B\langle\tilde{i}|\tilde{j}\rangle_B = \delta_{ij}\big)\,.$$

The coefficients λ_i ($\lambda_i > 0$) in the above expression are called the Schmidt coefficients, *and the decomposition itself is called the* Schmidt decomposition.

Before giving the proof of the theorem, let us note some relevant points:

(1) The theorem does not claim that the decomposition is unique (it is unique if and only if all the λ_i's are different—see 2.4.2.1). However, as will be shown, N, the number of elements in the sum, is unique.
(2) Different states (e.g. $|\psi\rangle_{AB}$ and $|\varphi\rangle_{AB}$) require, in general, a different choice of the orthonormal bases used.
(3) Let U_A and U_B be unitary operators operating on $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, and let U_{AB} be the unitary operator defined as $U_{AB} \equiv U_A \otimes U_B$. The Schmidt decomposition of $U_{AB}|\psi\rangle_{AB}$ has the same Schmidt *coefficients* as does $|\psi\rangle_{AB}$. However, it uses different bases in $\mathcal{H}_A$ and $\mathcal{H}_B$ ($|i\rangle_A \to U_A|i\rangle_A$ and $|\tilde{i}\rangle_B \to U_B|\tilde{i}\rangle_B$). To see this simply apply U_{AB} onto the Schmidt decomposition of $|\psi\rangle_{AB}$.

PROOF. To prove the theorem we need two auxiliary lemmas first.

LEMMA. *(*Polar decomposition*). Every matrix A may be written as a product of a unitary matrix U and a* positive matrix *H (a Hermitian matrix with only non-negative eigenvalues):*

$$A = UH \quad (H = \sqrt{A^\dagger A}).$$

PROOF. We shall start by proving the lemma for *nonsingular* matrices ($|A| \neq 0$). For such matrices we may always (as explained next) write A as

$$A = A\frac{1}{\sqrt{A^\dagger A}}\sqrt{A^\dagger A} \quad (|A| \neq 0).$$

To see this we must note that $A^\dagger A$ is a *positive* (Hermitian) matrix. It is clearly Hermitian since $(A^\dagger A)^\dagger = A^\dagger A$. It is positive, since for a given state $|\psi\rangle$ we may define

$$c|\varphi\rangle \equiv A|\psi\rangle,$$

where $|\varphi\rangle$ is some *normalized* state and c is some complex number ($A|\psi\rangle$ isn't necessarily normalized). Thus we have

$$\langle\psi|(A^\dagger A)|\psi\rangle = (\langle\psi|A^\dagger)(A|\psi\rangle) = |c|^2\langle\varphi|\varphi\rangle \geq 0.$$

This is true for any state $|\psi\rangle$ and therefore $A^\dagger A$, by definition, is positive (all eigenvalues are non-negative or equivalently $\langle\psi|(A^\dagger A)|\psi\rangle \geq 0$ for all $|\psi\rangle$).

Now, since we assume that A is nonsingular ($|A| \neq 0$), then so is $A^\dagger A$, and thus we also know that the eigenvalues (of $A^\dagger A$) must be *definite* positive, i.e. $\lambda_i > 0$ (if we had a zero eigenvalue, the determinant would also be zero). Therefore, $A^\dagger A$ may be written as

$$A^\dagger A = \sum_i \lambda_i |i\rangle\langle i| \quad (\lambda_i > 0).$$

By definition (of functions of matrices)[28] we may write

$$\sqrt{A^\dagger A} = \sum_i \sqrt{\lambda_i}|i\rangle\langle i|,$$

and indeed using this definition we have

$$\left(\sqrt{A^\dagger A}\right)^2 = \left(\sum_i \sqrt{\lambda_i}|i\rangle\langle i|\right)^2 = \sum_i \lambda_i|i\rangle\langle i| = A^\dagger A.$$

Clearly (when $\lambda_i > 0$), the inverse of $\sqrt{A^\dagger A}$ is

$$\left(\sqrt{A^\dagger A}\right)^{-1} \equiv \frac{1}{\sqrt{A^\dagger A}} = \sum_i \frac{1}{\sqrt{\lambda_i}}|i\rangle\langle i|,$$

which is easily checked:

$$\frac{1}{\sqrt{A^\dagger A}}\sqrt{A^\dagger A} = \left(\sum_i \frac{1}{\sqrt{\lambda_i}}|i\rangle\langle i|\right)\left(\sum_j \sqrt{\lambda_j}|j\rangle\langle j|\right) = \sum_i |i\rangle\langle i| = \mathbb{1}.$$

Therefore we can write, as we did,

$$A = A\frac{1}{\sqrt{A^\dagger A}}\sqrt{A^\dagger A} \quad (|A| \neq 0).$$

Now, let us define

$$U \equiv A\frac{1}{\sqrt{A^\dagger A}}$$

[28] If an operator/matrix A is diagonalizable, i.e. if it may be written in the form

$$A = \sum_i \lambda_i|i\rangle\langle i|,$$

then $f(A)$ is defined as

$$f(A) \equiv \sum_i f(\lambda_i)|i\rangle\langle i|.$$

This definition may be used even if A is not Hermitian (when λ_i are not necessarily real).

If a Taylor expansion of $f(x)$ exists ($f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$), then we may also define $f(A)$ as the Taylor expansion in A:

$$f(A) = a_0\mathbb{1} + a_1 A + a_2 A^2 + \cdots.$$

The two definitions are not always available for use (there may be no Taylor expansion around $x = 0$, or A may not be diagonalizable). However, if the two are possible, then they coincide.

and

$$H \equiv \sqrt{A^\dagger A}.$$

Clearly U is unitary ($UU^\dagger = \mathbb{1}$) and H is Hermitian ($H^\dagger = H$), so assuming that A is nonsingular, we have what we were looking for

$$A = UH \quad (|A| \neq 0).$$

To complete the proof we must now treat the singular case as well. Let us assume that A is an $N \times N$ matrix. If A is indeed singular, then it must have eigenvectors v_i, $i = 1, \ldots, n$ ($n \leq N$), with eigenvalue zero (since$|A| = 0$, the columns of A are linearly dependent, and so there must exist vectors v_i such that $Av_i = 0$). The vectors v_i define a subspace of dimension n for which we can choose an orthonormal basis e_i, $i = 1, \ldots, n$. Let us now complete the orthonormal basis (of the subspace) to an N dimensional orthonormal basis e_i ($i = 1, \ldots, N$) of the whole space (on which A operates). Since the $\{e_i\}$ form an orthonormal basis and since the e_i, for $i = 1, \ldots, n$, are eigenvectors with eigenvalue zero, then there exists some *unitary* matrix V such that

$$V^\dagger AV = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \tilde{A} \end{pmatrix} \quad (V^\dagger = V^{-1}),$$

where $\tilde{A}$ is an $(N-n) \times (N-n)$ nonsingular matrix, and the number of zeros on the diagonal is n.

Since $\tilde{A}$ is nonsingular we can use the result we found above and write

$$V^\dagger AV = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \tilde{U}\tilde{H} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \tilde{U} \end{pmatrix} \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \tilde{H} \end{pmatrix},$$

where $\tilde{U}$ is a unitary matrix and $\tilde{H}$ is a positive Hermitian matrix (both $(N-n) \times (N-n)$ matrices). From this we easily find that

$$A = V \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \tilde{U} \end{pmatrix} V^\dagger V \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \tilde{H} \end{pmatrix} V^\dagger.$$

Defining

$$U \equiv V \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \tilde{U} \end{pmatrix} V^\dagger,$$

and

$$H \equiv V \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \tilde{H} \end{pmatrix} V^\dagger = \sqrt{A^\dagger A},$$

we see that, as declared,

$$A = UH$$

where H is a positive (Hermitian) matrix and U is a unitary matrix. □

LEMMA. *(*Singular value decomposition*). Every matrix A may be written as a product of a unitary matrix U, a diagonal matrix D, and another unitary matrix V:*

$$A = UDV$$

PROOF. According to the previous lemma, we can always write A as

$$A = U_1 H.$$

Since H is Hermitian, then there is a unitary matrix T which diagonalizes it

$$T^\dagger H T = D$$
$$\Rightarrow H = TDT^\dagger.$$

Therefore, we can write

$$A = U_1 T D T^\dagger.$$

We now define

$$U \equiv U_1 T$$

and

$$V \equiv T^\dagger.$$

These new matrices are clearly unitary, and we therefore have

$$A = UDV.$$

□

Having proved the above two lemmas, we may now prove the Schmidt decomposition. By definition, a state $|\Psi_{AB}\rangle$ can be written in general as[29]

$$|\Psi_{AB}\rangle = \sum_{i,j} a_{ij} |\alpha_i\rangle_A |\beta_j\rangle_B,$$

where $|\alpha_i\rangle$ is an orthonormal basis of $\mathcal{H}_A$ and $|\beta_j\rangle$ is an orthonormal basis of $\mathcal{H}_B$. The coefficients a_{ij} define a matrix A

$$(A)_{ij} \equiv a_{ij}.$$

By the second lemma there are matrices U, D, V such that (since D is diagonal)

$$A = UDV \Rightarrow a_{ij} \equiv A_{ij} = \sum_k U_{ik} D_{kk} V_{kj}.$$

Substituting this into $|\Psi_{AB}\rangle$ gives then

$$\begin{aligned} |\Psi_{AB}\rangle &= \sum_{i,j,k} U_{ik} D_{kk} V_{kj} |\alpha_i\rangle_A |\beta_j\rangle_B \\ &= \sum_k D_{kk} \left(\sum_i U_{ik} |\alpha_i\rangle_A \right) \left(\sum_j V_{kj} |\beta_j\rangle_B \right). \end{aligned}$$

Now, since U and V are unitary matrices then they transform an orthonormal basis into a new orthonormal basis. Thus, we may define two new orthonormal bases

$$|k\rangle_A \equiv \sum_i U_{ik} |\alpha_i\rangle_A$$

and

$$|\tilde{k}\rangle_B \equiv \sum_j V_{kj} |\beta_j\rangle_B.$$

[29] By definition ($|\psi_i\rangle_A$ and $|\varphi_j\rangle_B$ are arbitrary states)

$$|\Psi_{AB}\rangle = \sum_{i,j} \beta_{ij} |\psi_i\rangle_A |\varphi_j\rangle_B.$$

If we expand each of the states $|\psi_i\rangle_A$ using the orthonormal basis $|a_i\rangle_A$ and similarly for $\mathcal{H}_B$, we get the above result.

Using these definitions we now have

$$|\Psi_{AB}\rangle = \sum_k D_{kk}|k\rangle_A|\tilde{k}\rangle_B,$$

which is *almost* the Schmidt decomposition. To have the Schmidt decomposition we must have $D_{kk} = \sqrt{\lambda_k}$ and therefore D_{kk} must be positive. In general D_{kk} can always be written as $\sqrt{\lambda_k}e^{i\theta_k}$. If we push the phase $e^{i\theta_k}$ into the definition of our orthonormal bases, then we finally get the desired form.

□

As an example let us examine two cases. The first is

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B).$$

This state is already in a Schmidt decomposition since $|0\rangle$ and $|1\rangle$ are orthonormal and further more the same ket does not appear in two different elements.

However, if we examine

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle_A|\uparrow_z\rangle_B + |\downarrow_z\rangle_A|\uparrow_x\rangle_B),$$

this is not a Schmidt decomposition since $|\uparrow_z\rangle_B$ is not orthonormal to $|\uparrow_x\rangle_B$.

We may now ask how do we find the Schmidt decomposition appropriate for a given state. When the λ_i's are *nondegenerate* this is quite simple (shown next). Assuming the Schmidt decomposition is of the form

$$|\Psi_{AB}\rangle = \sum_i \sqrt{\lambda_i}|i\rangle_A|\tilde{i}\rangle_B,$$

the reduced density matrices in $\mathcal{H}_A$ and $\mathcal{H}_B$ are

$$\rho_A = \mathrm{Tr}_B\,|\Psi_{AB}\rangle\langle\Psi_{AB}| = \sum_j {}_B\langle\tilde{j}|\,(|\Psi_{AB}\rangle\langle\Psi_{AB}|)\,|\tilde{j}\rangle_B = \sum_i \lambda_i|i\rangle_{AA}\langle i|$$

$$\rho_B = \mathrm{Tr}_A\,|\Psi_{AB}\rangle\langle\Psi_{AB}| = \sum_j {}_A\langle j|\,(|\Psi_{AB}\rangle\langle\Psi_{AB}|)\,|j\rangle_A = \sum_i \lambda_i|\tilde{i}\rangle_{BB}\langle\tilde{i}|.$$

We see that the same λ_i's appear in both density matrices (when they are diagonalized). Further more, λ_i is the coefficient of both $|i\rangle_{AA}\langle i|$ in ρ_A and of $|\tilde{i}\rangle_{BB}\langle\tilde{i}|$ in ρ_B (the same index i in all). Thus, if we diagonalize each of the reduced density matrices we can match eigenstates with identical eigenvalues and so deduce the Schmidt decomposition.

As an example of this method let us return to our previous example

$$|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle_A|\uparrow_z\rangle_B + |\downarrow_z\rangle_A|\uparrow_x\rangle_B).$$

The reduced density matrix ρ_B for this state is

$$\begin{aligned}\rho_B &= \frac{1}{2}\left(|\uparrow_z\rangle_{BB}\langle\uparrow_z| + |\uparrow_x\rangle_{BB}\langle\uparrow_x|\right)\\ &= \frac{2+\sqrt{2}}{4}|0\rangle_{BB}\langle 0| + \frac{2-\sqrt{2}}{4}|1\rangle_{BB}\langle 1|,\end{aligned}$$

where we have defined (by finding the eigenstates of ρ_B)[30]

$$|0\rangle_B = \frac{1}{\sqrt{3+2\sqrt{2}}}\left[(1+\sqrt{2})|\uparrow_z\rangle_B + |\downarrow_z\rangle_B\right],$$

$$|1\rangle_B = \frac{1}{\sqrt{3-2\sqrt{2}}}\left[(1-\sqrt{2})|\uparrow_z\rangle_B + |\downarrow_z\rangle_B\right].$$

For ρ_A we get[31]

$$\begin{aligned}\rho_A &= \frac{1}{2}\left[|\uparrow_z\rangle_{AA}\langle\uparrow_z| + \frac{1}{\sqrt{2}}|\uparrow_z\rangle_{AA}\langle\downarrow_z| + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_{AA}\langle\uparrow_z| + |\downarrow_z\rangle_{AA}\langle\downarrow_z|\right] \\ &= \frac{2+\sqrt{2}}{4}|0\rangle_{AA}\langle 0| + \frac{2-\sqrt{2}}{4}|1\rangle_{AA}\langle 1|,\end{aligned}$$

where we have defined here (again after finding the eigenstates of the density matrix)

$$|0\rangle_A = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle_A + |\downarrow_z\rangle_B\right),$$

$$|1\rangle_A = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle_A - |\downarrow_z\rangle_B\right).$$

Having found the (nondegenerate) eigenstates of the two partial density matrices, we can now finally write down the Schmidt decomposition, as follows:

$$|\Psi_{AB}\rangle = \frac{2+\sqrt{2}}{4}|0\rangle_A|0\rangle_B + \frac{2-\sqrt{2}}{4}|1\rangle_A|1\rangle_B.$$

Note, that if the Schmidt coefficients λ_i are degenerate, then the eigenvalues of the density matrices will also be degenerate and we won't be able to make the one-to-one correspondence between the orthonormal states of $\mathcal{H}_A$ and $\mathcal{H}_B$. This means that we cannot use the density matrices to find the Schmidt decomposition when there is a degeneracy.

[30] Using $|\uparrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle)$, we can write (dropping the index B)

$$\begin{aligned}\rho_B &= \frac{1}{2}\left(|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}[(|\uparrow_z\rangle + |\downarrow_z\rangle)(\langle\uparrow_z| + \langle\downarrow_z|)]\right) \\ &= \frac{3}{4}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{4}|\downarrow_z\rangle\langle\downarrow_z| + \frac{1}{4}|\uparrow_z\rangle\langle\downarrow_z| + \frac{1}{4}|\downarrow_z\rangle\langle\uparrow_z|.\end{aligned}$$

In the z basis, this is the same as the matrix

$$\rho_B = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix},$$

whose eigenstates and eigenvalues are

$$v_\pm = \frac{1}{\sqrt{3\pm 2\sqrt{2}}}\begin{pmatrix} 1\pm\sqrt{2} \\ 1 \end{pmatrix} \quad ,\lambda_\pm = \frac{2\pm\sqrt{2}}{4}.$$

Thus, the (normalized) eigenstates of ρ_B are

$$|\pm\rangle = \frac{1}{\sqrt{3\pm 2\sqrt{2}}}\left[\left(1\pm\sqrt{2}\right)|\uparrow_z\rangle + |\downarrow_z\rangle\right].$$

[31] Using $|\uparrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle)$, we may write

$$\begin{aligned}|\Psi_{AB}\rangle &= \frac{1}{\sqrt{2}}\left[|\uparrow_z\rangle_A|\uparrow_z\rangle_B + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_A(|\uparrow_z\rangle_B + |\downarrow_z\rangle_B)\right] \\ &= \frac{1}{\sqrt{2}}\left[\left(|\uparrow_z\rangle_A + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_A\right)|\uparrow_z\rangle_B + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_A|\downarrow_z\rangle_B\right]\end{aligned}$$

so that (dropping the index A)

$$\begin{aligned}\rho_A &= \frac{1}{2}\left(|\uparrow_z\rangle + \frac{1}{\sqrt{2}}|\downarrow_z\rangle\right)\left(\langle\uparrow_z| + \frac{1}{\sqrt{2}}\langle\downarrow_z|\right) + \frac{1}{4}|\downarrow_z\rangle\langle\downarrow_z| \\ &= \frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2\sqrt{2}}|\uparrow_z\rangle\langle\downarrow_z| + \frac{1}{2\sqrt{2}}|\downarrow_z\rangle\langle\uparrow_z| + \frac{1}{2}|\downarrow_z\rangle\langle\downarrow_z|.\end{aligned}$$

However, we can always find the decomposition using the method described in the proof of the Schmidt decomposition.

2.4.2.1. *Uniqueness of the Schmidt decomposition.* We have claimed at the start that the Schmidt decomposition is unique if and only if the Schmidt coefficients are nondegenerate. Further more, it was claimed that even if the decomposition is not unique, the *number* of elements (in the decomposition) is unique. We shall now prove this.

Showing the uniqueness of the number of elements (in the decomposition) is simple. Assuming the Schmidt decomposition to be given by

$$|\Psi_{AB}\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i}|i\rangle_A|\tilde{i}\rangle_B,$$

it is easy to see that the density matrix describing the state is

$$\rho_{AB} = \sum_{i=1}^{N} \lambda_i(|i\rangle_A|\tilde{i}\rangle_B)({}_B\langle\tilde{i}|_A\langle i|).$$

In this form the density matrix is already diagonal, and it is clear that it has *N nonzero* eigenvalues (*N* is also the number of elements in the Schmidt decomposition). Since the density matrix is unique it is clear that the *number* of elements in the Schmidt decomposition must also be unique.[32]

We must now show that the decomposition is unique if and only if the Schmidt coefficients are nondegenerate. Actually, we have already shown above that for *nondegenerate* coefficients the decomposition is unique: We have shown that by diagonalizing each of the reduced density matrices ρ_A and ρ_B we can find the Schmidt coefficients and the unique bases $|i\rangle_A$ and $|\tilde{i}\rangle_B$ (see above). Therefore, to complete our proof we need only show that for the degenerate case the decomposition cannot be unique.

If the coefficients are degenerate, then the Schmidt decomposition includes *at least* two elements with the same coefficient. Thus, the decomposition may be written (concentrating only on two of the degenerate elements) as

$$|\Psi_{AB}\rangle = \cdots + \lambda|i\rangle_A|\tilde{i}\rangle_B + \cdots + \lambda|j\rangle_A|\tilde{j}\rangle_B + \cdots,$$

or, after renumbering the states, as

$$|\Psi_{AB}\rangle = \lambda\left(|0\rangle_A|\tilde{0}\rangle_B + |1\rangle_A|\tilde{1}\rangle_B\right) + \cdots.$$

To prove that a different Schmidt decomposition exists, it suffices to show that there always exist (nontrivial) *orthonormal* states $|a\rangle_A, |b\rangle_A$ and $|\tilde{a}\rangle_B, |\tilde{b}\rangle_B$ such that

$$|0\rangle_A|\tilde{0}\rangle_B + |1\rangle_A|\tilde{1}\rangle_B = |a\rangle_A|\tilde{a}\rangle_B + |b\rangle_A|\tilde{b}\rangle_B.$$

This is easily shown by defining

$$|a\rangle_A \equiv \cos\alpha|0\rangle_A + \sin\alpha|1\rangle_A,$$
$$|b\rangle_A \equiv e^{i\theta}\sin\alpha|0\rangle_A - e^{i\theta}\cos\alpha|1\rangle_A,$$
$$|\tilde{a}\rangle_B \equiv \cos\alpha|\tilde{0}\rangle_B + \sin\alpha|\tilde{1}\rangle_B,$$

and

$$|\tilde{b}\rangle_B \equiv e^{-i\theta}\sin\alpha|\tilde{0}\rangle_B - e^{-i\theta}\cos\alpha|\tilde{1}\rangle_B.$$

With these definitions we have

$$|0\rangle_A = \cos\alpha|a\rangle_A + e^{i\theta}\sin\alpha|b\rangle_A$$
$$|1\rangle_A = \sin\alpha|a\rangle_A - e^{i\theta}\cos\alpha|b\rangle_A,$$
$$|\tilde{0}\rangle_B = \cos\alpha|\tilde{a}\rangle_B + e^{-i\theta}\sin\alpha|\tilde{b}\rangle_B,$$

[32] Again, since many states can give the same density matrix, then this does not necessarily mean that the Schmidt decomposition is unique.

and

$$|\tilde{1}\rangle_B = \sin\alpha|\tilde{a}\rangle_B - e^{-i\theta}\cos\alpha|\tilde{b}\rangle_B.$$

A straight forward calculation[33] then shows that

$$|0\rangle_A|\tilde{0}\rangle_B + |0\rangle_A|\tilde{0}\rangle_B = |a\rangle_A|\tilde{a}\rangle_B + |b\rangle_A|\tilde{b}\rangle_B,$$

regardless of the angles α, θ chosen. Thus, we see that there is an infinite choice of orthonormal bases which give a valid Schmidt decomposition (when there is a degeneracy in the coefficients).

2.4.2.2. *The Schmidt decomposition and entanglement.* We have seen above that the number of elements in the Schmidt decomposition is unique. As you may recall, an *entangled* state is defined as a state which *cannot* be written as a tensor product of two states. On the other hand, a nonentangled state is one which can be written in this form (e.g. $|\Psi_{AB}\rangle = |0\rangle_A|\tilde{0}\rangle_B$). Thus we see that if the Schmidt decomposition has only a single element ($|\Psi_{AB}\rangle = |0\rangle_A|\tilde{0}\rangle_B$) then the state is nonentangled. Otherwise, the state is entangled.

[33] The calculation is:

$$\begin{aligned}
&|0\rangle_A|\tilde{0}\rangle_B + |1\rangle_A|\tilde{1}\rangle_B = \\
&= \left(\cos\alpha|a\rangle_A + e^{i\theta}\sin\alpha|b\rangle_A\right)\left(\cos\alpha|\tilde{a}\rangle_B + e^{-i\theta}\sin\alpha|\tilde{b}\rangle_B\right) \\
&\quad + \left(\sin\alpha|a\rangle_A - e^{i\theta}\cos\alpha|b\rangle_A\right)\left(\sin\alpha|\tilde{a}\rangle_B - e^{-i\theta}\cos\alpha|\tilde{b}\rangle_B\right) \\
&= (\cos^2\alpha + \sin^2\alpha)|a\rangle_A|\tilde{a}\rangle_B + (\sin^2\alpha + \cos^2\alpha)|b\rangle_A|\tilde{b}\rangle_B \\
&\quad + (e^{-i\theta}\cos\alpha\sin\alpha - e^{-i\theta}\sin\alpha\cos\alpha)|a\rangle_A|\tilde{b}\rangle_B \\
&\quad + (e^{i\theta}\sin\alpha\cos\alpha - e^{i\theta}\cos\alpha\sin\alpha)|b\rangle_A|\tilde{a}\rangle_B \\
&= |a\rangle_A|\tilde{a}\rangle_B + |b\rangle_A|\tilde{b}\rangle_B.
\end{aligned}$$

Part 2

Entanglement

CHAPTER 3

Creating entanglement experimentally

To be completed?

CHAPTER 4

Hidden variables

4.1. The EPR[1] Paradox

Assume a two particle wave function of the form

$$\psi \approx \delta(x_1 - x_2 - L)\delta(p_1 + p_2),$$

where δ are not exactly delta functions but only arbitrarily good, normalizable approximations. The operators $x_1 - x_2$ and $p_1 + p_2$ commute,[2] so we can measure both simultaneously.

From the wave functions we know that

$$x_1 - x_2 \approx L \quad \text{(distance between particles)}$$

$$p_1 + p_2 \approx 0 \quad \text{(total mommentum)}.$$

Clearly, if Alice measures x_1, then she knows that $x_2 \approx x_1 - L$, on the other hand, if she measures p_1 then she knows $p_2 = -p_1$. Let us now assume that the distance L is large enough so that the during time it takes to make a measurement, light cannot travel between the two particles. Thus we assume that particle 2 isn't effected by measurements on particle 1.

Following EPR (Einstein, Podolsky, Rosen) We define an *element of reality* as a quantity which may be predicted with certainty without disturbing the system at all. In our case here both x_2 and p_2 are elements of reality[3] because we may find them without disturbing particle 1 (only particle 2). Since the measurements are done far away from particle 2, then the particle is not effected by them and x_2, p_2 must be both elements of reality.

However, from the uncertainty principle, we cannot know both x_2 and p_2, and thus we find a contradiction with quantum mechanics, which tells us that the theory is incomplete. Further more, since the result of the measurement of the system is unaffected by the measurement on particle 1, one might think that the result of the measurements on 2, where already “written” somewhere. This lead to the thought of hidden variables theory (HV).

We should also mention Bohm's version of the EPR paradox, sometimes known as EPRB. His version is discrete. He uses the entangled state

$$|\psi\rangle = \frac{1}{\sqrt{2}} (\uparrow_1 \downarrow_2 - \downarrow_1 \uparrow_2)$$

which has a total spin of zero. One can now measure the spin (in the z-direction) of particle 1 and deduce that of particle 2. From here on it is similar to the original EPR paradox.

4.2. Bell inequalities

The EPR paradox led to the thought that there might exist hidden variable theories which give the same predictions as quantum mechanics. Bohm (1952) had found the pilot wave interpretation which was a *nonlocal* hidden variable theory. The question remained,

[1] Einstein-Podolsky-Rosen

[2]

$$[x_1 - x_2, p_1 + p_2] = [x_1, p_1] - [x_2, p_2] = 0.$$

[3] We could similarly also choose x_2, p_2 as the elements of reality, but then x_1, p_1 won't be????

however, whether a *local* hidden variable theory might be possible. Finally, in 1964, Bell had shown that for a Hilbert space above 2 dimensions, one cannot have a local hidden variable theory (actually he showed an inequality which such a theory must obey, and quantum mechanically does not obey it experimentally—see below)

4.2.1. A local hidden variables theory for spin $\frac{1}{2}$. Before going on to Bell's inequality let us first show (following Bell) an example where a hidden variable theory *is* possible (a 2 dimensional Hilbert space).

Assume a spin $\frac{1}{2}$ system in a state

$$|\psi_0\rangle = |\uparrow_z\rangle.$$

In quantum mechanics the expectation value of $\sigma_{\hat{n}}$ is[4]

$$\langle\sigma_{\hat{n}}\rangle_{\psi_0} = \hat{n}\cdot\hat{z} \equiv \cos\theta_{nz},$$

where

$$\sigma_{\hat{n}} \equiv \hat{n}\cdot\vec{\sigma},$$

and where θ_{nz} is the angle between $\hat{n}$ and $\hat{z}$ ($\vec{\sigma}$ is the vector of Pauli matrices). We now wish to find whether we can produce a hidden variable theory which will reproduce the same results.

To achieve this let us assume, as a parameter (the hidden variable), a unit vector $\hat{\lambda}$ with equal probability to point anywhere on the upper-half ($z > 0$) of a unit sphere (but zero probability to point towards the lower half). We *define* the the result of measuring $\sigma_{\hat{n}}$ as the value $V_{\sigma_{\hat{n}}}$ given by[5]

$$V_{\sigma_{\hat{n}}}(\hat{\lambda}) \equiv \mathrm{sign}\left(\hat{n}\cdot\hat{\lambda}\right) = \mathrm{sign}\left(\cos\theta_{n\lambda}\right).$$

Since we assume that $\hat{\lambda}$ is (for some unknown reason) uniformly distributed on the upper half of the unit sphere, then within an area of $2\pi\frac{\theta_{nz}}{\pi}$ (of the upper hemisphere) we get negative values for $\hat{n}\cdot\hat{\lambda}$ while in the rest of the upper hemisphere (area of $2\pi - 2\pi\frac{\theta_{nz}}{\pi}$) we get a positive value for $\hat{n}\cdot\hat{\lambda}$ (note that the area of half a sphere is $\frac{1}{2}4\pi r^2$, which for $r = 1$ gives 2π). As a result, the average value we get is

$$\langle V_{\sigma_{\hat{n}}}\rangle = \frac{(-1)\cdot\left(2\pi\frac{\theta_{nz}}{\pi}\right) + (+1)\cdot\left(2\pi - 2\pi\frac{\theta_{nz}}{\pi}\right)}{2\pi} = 1 - \frac{2\theta_{nz}}{\pi},$$

where

$$\cos\theta_{nz} \equiv \hat{n}\cdot\hat{z}.$$

Clearly, this result does not give us the desired quantum result ($\cos\theta_{nz}$). To achieve that we simply use a different $\hat{n}$ in the expression for $V_{\sigma_{\hat{n}}}(\hat{\lambda})$. Instead of $\hat{n}$ we shall use $\hat{n}'$ at an angle θ' (with the z-axis) such that

$$1 - \frac{2\theta'}{\pi} \equiv \cos\theta_{nz} = \hat{n}\cdot\hat{z}.$$

The important point is that we can make a one-to-one a mapping between θ and θ' and therefore we can have a hidden variable theory, as desired. In this new theory $\hat{\lambda}$ is still

[4]

$$\begin{aligned}\langle\sigma_{\hat{n}}\rangle_{\psi_0} &= \langle\uparrow_z|\cos\theta\sigma_z + \sin\theta\cos\phi\sigma_x + \sin\theta\sin\phi\sigma_y|\uparrow_z\rangle\\ &= \cos\theta\langle\uparrow_z|\sigma_z|\uparrow_z\rangle + \sin\theta\cos\phi\langle\uparrow_z|\sigma_x|\uparrow_z\rangle + \sin\theta\sin\phi\langle\uparrow_z|\sigma_y|\uparrow_z\rangle\\ &= \cos\theta + 0 + 0\end{aligned}$$

[5] The function $\mathrm{sign}()$ gives the sign of its operand, i.e.

$$\mathrm{sign}(x) = \begin{cases} +1 & x \geq 0 \\ -1 & x < 0 \end{cases}.$$

uniformly distributed on the upper half of the sphere but the value of a measurement (of $\sigma_{\hat{n}}$) is given by (note that $\hat{n}$ appears only in the first term while $\hat{n}'$ appears in the rest)

$$V_{\sigma_{\hat{n}}}(\hat{\lambda}) \equiv \mathrm{sign}\left(\hat{n}' \cdot \hat{\lambda}\right) = \mathrm{sign}\left(\cos\theta_{n'\lambda}\right),$$

where $\hat{n}'$ obeys

$$\hat{n}' \cdot \hat{z} = \cos\theta' = \cos\left[\frac{\pi}{2}(1-\cos\theta_{nz})\right] = \cos\left[\frac{\pi}{2}(1-\hat{n}\cdot\hat{z})\right].$$

(For example if $\hat{n} = \hat{x}\sin\theta\cos\varphi + \hat{y}\sin\theta\sin\varphi + \hat{z}\cos\theta$ then we can define $\hat{n}' = \hat{x}\sin\theta'\cos\varphi + \hat{y}\sin\theta'\sin\varphi + \hat{z}\cos\theta'$).

Note, that we could have also constructed our hidden variable theory differently. For example, we could have constructed a model with a parameter λ uniformly distributed between 0 and 1, such that

$$\sigma_{\hat{n}} = \begin{cases} 1 & 0 < \lambda < \cos^2\frac{\theta_{nz}}{2} \\ -1 & \cos^2\frac{\theta_{nz}}{2} < \lambda < 1 \end{cases}.$$

This would give us

$$\langle\sigma_{\hat{n}}\rangle = \cos\theta_{nz},$$

where of course

$$\hat{n}\cdot\hat{z} = \cos\theta_{nz}.$$

Before going on to the general case, let us try and see if we can construct a hidden variable theory for a system of more then one spin such as a system of two entangled spins.

Assume a system of two spins with total angular momentum zero. For such a system, if we measure spin number 1 in the z direction and get “up”, then measuring spin number 2, also in the z direction, must give “down”. We need a model to give us this behavior. We cannot use the exact same model as before (for each spin separately), since the λ in each spin would be independent and we won’t get the desired result. Instead, let us try a modified model. In our new model the two spins are in opposite directions, but the direction of spin 1 is random, i.e. particle 1 has spin “up” in the random direction $\hat{j}_1$ and particle 2 has spin “up” in the direction $\hat{j}_2 = -\hat{j}_1$. We shall use the random parameters $\hat{j}_i$ instead of the random parameter $\hat{\lambda}$ in the previous model. Thus, we now have (with some change of notation):

$$V_i(\hat{n}) = \mathrm{sign}(\hat{n}\cdot\hat{j}_i) \quad (\hat{j}_2 = -\hat{j}_1),$$

where i is the index of the particle being measure and $\hat{n}$ is the direction in which the measurement is performed. We shall denote by $\hat{a}$ and $\hat{b}$ the directions in which we measure particles 1 and 2 respectively.

We can easily see, that by our assumption of $\hat{j}_1$ being random, we get, as in QM

$$\langle V_1(\hat{a})\rangle_{\hat{j}_1} = \langle V_2(\hat{b})\rangle_{\hat{j}_1} = 0,$$

where $\langle\cdot\rangle_{\hat{j}_1}$ denotes averaging of $\hat{j}_1$.

Now let us calculate $\langle V_1(\hat{a})V_2(\hat{b})\rangle$. Clearly (since $\hat{j}_2 = -\hat{j}_1$), the case $\hat{a} = \hat{b}$ gives the regular QM result

$$\langle V_1(\hat{a})V_2(\hat{a})\rangle_{\hat{j}_1} = -1.$$

To find the result for general directions $\hat{a}$ and $\hat{b}$ we draw two half-spheres on the unit sphere, one centered around $\hat{a}$ and the second around $\hat{b}$. These half-sphere represent, respectively, the directions $\hat{j}_1$ and $\hat{j}_2$ for which $V_1(\hat{a})$ and $V_2(\hat{b})$ are positive. Thus, when $\hat{j}_1$ either falls in the intersection of the two hemispheres, or when it falls outside of *both* hemisphere, we have

$$V_1(\hat{a})V_2(\hat{b}) = -1 \quad (\text{area of } \frac{2\pi-2\theta}{2\pi}4\pi),$$

while on the rest of the unit sphere we have

$$V_1(\hat{a})V_2(\hat{b}) = +1 \quad (\text{area of } \frac{2\theta}{2\pi}4\pi).$$

If we denote by θ the angle between $\hat{a}$ and $\hat{b}$, then the first case occurs on a surface area of

$$\frac{2\pi - 2\theta}{2\pi}4\pi \quad (V_1(\hat{a})V_2(\hat{b}) = -1)$$

and the second on an area of

$$\frac{2\theta}{2\pi}4\pi \quad (V_1(\hat{a})V_2(\hat{b}) = +1).$$

(It is easier to calculate the area of the second case first.). Taking the average over the areas gives

$$\langle V_1(\hat{a})V_2(\hat{b})\rangle_{\hat{j}_1} = -1 + \frac{2\theta}{\pi},$$

while the result in quantum mechanics is[6]

$$\langle \sigma_{\hat{a}}\sigma_{\hat{b}}\rangle = -\cos\theta_{ab}.$$

Not surprisingly, as before, we got different results. The question, however, is whether we can correct our model (as before) so that we get the correct answer. We shall see (in the next two subsection) that the answer is no. The reason is that the parameter in the results is the angle between $\hat{a}$ and $\hat{b}$. We cannot make a deterministic change, *separately* on each function $V_i(\hat{n})$, so that we will get a correct result when combined. (The change in each function may not depend on the direction of measuring the other particle. Otherwise the theory is nonlocal)

4.2.2. The CHSH[7] inequality[8].

Let us assume that a *local* hidden variables theory exists, where by local we mean that that every particle has its own set of hidden variables which determine its behavior (regardless of what the others do). We shall now see that this assumption requires a certain inequality to always hold. Since quantum mechanics does *not always* obey the inequality, then the only conclusion is that quantum mechanics is not a *local* hidden variable theory.

We shall again study the system of two spin $\frac{1}{2}$ particles with total angular momentum 0. This time however, particle 1 can be measured either in direction $\hat{a}$ or in $\hat{a}'$ and particle

[6] The simplest way to see this is to note that for any direction $\hat{n}$ we have

$$\frac{1}{\sqrt{2}}\left(|\uparrow_{\hat{z}}\downarrow_{\hat{z}}\rangle - |\downarrow_{\hat{z}}\uparrow_{\hat{z}}\rangle\right) = \frac{1}{\sqrt{2}}e^{i\varphi_{nz}}\left(|\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle - |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle\right),$$

where φ_{nz} is some global phase. This is because the state (up to a global phase) is uniquely defined by its *total* angular momentum and the momentum in the $\hat{z}$ direction. On both sides of the equations, these values are zero, so the states must be physically the same.

Using this relation, we can write the state as

$$|\Psi\rangle_{12} = \frac{1}{\sqrt{2}}e^{i\varphi}\left(|\uparrow_{\hat{a}}\downarrow_{\hat{a}}\rangle - |\downarrow_{\hat{a}}\uparrow_{\hat{a}}\rangle\right),$$

where $\hat{a}$ is the direction in which we measure particle 1. Using this basis, we may write

$$\sigma_{\hat{b}} = \cos\theta_{ab}\sigma_{\hat{a}} + \sin\theta_{ab}\sigma_{\perp},$$

where θ_{ab} is the angle between $\hat{a}$ and $\hat{b}$, and $\sigma_{\perp}$ is an operator measuring perpendicular to $\hat{a}$. It is now easy to see that

$$\begin{aligned} &{}_{12}\langle\Psi|\sigma_{\hat{a}}^{(1)}\sigma_{\hat{b}}^{(2)}|\Psi\rangle_{12} = \\ &= \frac{1}{2}\left(\langle\uparrow_{\hat{a}}\downarrow_{\hat{a}}| - \langle\downarrow_{\hat{a}}\uparrow_{\hat{a}}|\right)e^{-i\varphi}\left[\sigma_{\hat{a}}^{(1)}(\cos\theta_{ab}\sigma_{\hat{a}}^{(2)} + \sin\theta_{ab}\sigma_{\perp}^{(2)})\right]e^{i\varphi}\left(|\uparrow_{\hat{a}}\downarrow_{\hat{a}}\rangle - |\downarrow_{\hat{a}}\uparrow_{\hat{a}}\rangle\right) \\ &= -\cos\theta_{ab}. \end{aligned}$$

[7] Clauser Horne Shimony and Holt.

[8] Bell's original inequality is discussed in the next subsection (4.2.3).

2 can be measured either in direction $\hat{b}$ or in $\hat{b}'$. Although we cannot measure both $\hat{a}$ and $\hat{a}'$ simultaneously (nor $\hat{b}$ and $\hat{b}'$), the fact that we have hidden variables allows us to know in advance what the result would be (should we make the measurements). We shall use these "results" to develop the inequality. We shall denote the result of measuring $\sigma_{\hat{a}}$ by a the result of measuring $\sigma_{\hat{a}'}$ by a' and so on.

The possible results of any measurement (of $\sigma_{\hat{n}}$) are only ±1, thus (as explained next) we may write

$$(a+a')b+(a-a')b' = \pm 2.$$

This is because either

$$a+a' = 0 \Rightarrow (a-a') = \pm 2 \quad (b' = \pm 1)$$

or

$$a-a' = 0 \Rightarrow (a+a') = \pm 2 \quad (b = \pm 1).$$

Although we do not know the distribution (of occurrence) of the values +2 and −2 in our hidden variables theory, we can conclude that we must have[9]

$$\left|\langle (a+a')b+(a-a')b'\rangle\right| \leq 2 \quad \text{(for hidden variables)}$$

or equivalently

$$\left|\langle ab+a'b+ab'-a'b'\rangle\right| \leq 2 \quad \text{(for hidden variables)}$$

writing this in standard quantum mechanical form we can write

$$\left|\langle \sigma_{\hat{a}}\sigma_{\hat{b}}+\sigma_{\hat{a}'}\sigma_{\hat{b}}+\sigma_{\hat{a}}\sigma_{\hat{b}'}-\sigma_{\hat{a}'}\sigma_{\hat{b}'}\rangle\right| \leq 2 \quad \text{(for hidden variables)}$$

or

$$\left|\langle \sigma_{\hat{a}}\sigma_{\hat{b}}\rangle+\langle\sigma_{\hat{a}'}\sigma_{\hat{b}}\rangle+\langle\sigma_{\hat{a}}\sigma_{\hat{b}'}\rangle-\langle\sigma_{\hat{a}'}\sigma_{\hat{b}'}\rangle\right| \leq 2 \quad \text{(for hidden variables)}.$$

Each one of the four averages (within the absolute value) can be measured in experiment and then the inequality checked. This inequality is called the *CHSH inequality*.

Let us see if QM (always) obeys this inequality. As an example we shall take the case where all direction are in the same plane such that (see figure 4.2.1)

$$\hat{a}\perp\hat{a}' \quad , \quad \hat{b}\perp\hat{b}' \quad , \quad \hat{a}\cdot\hat{b} = \cos\frac{\pi}{4} \quad \text{and } \hat{a}'\cdot\hat{b}' = \cos\frac{3\pi}{4}.$$

Since in QM $\langle\hat{\sigma}^{(1)}_{\hat{n}}\hat{\sigma}^{(2)}_{\hat{m}}\rangle = \hat{n}\cdot\hat{m}$ then for the above choice we have

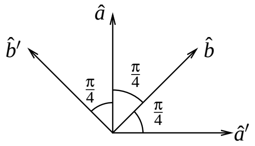


FIGURE 4.2.1. The geometry used for the example of the CHSH inequality violation. The spin of particle 1 is measured either in the $\hat{a}$ or the $\hat{a}'$ directions, while the spin of particle 2 is measured either in the $\hat{b}$ or the $\hat{b}'$ directions.

$$\langle\sigma_{\hat{a}}\sigma_{\hat{b}}\rangle = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2},$$

$$\langle\sigma_{\hat{a}'}\sigma_{\hat{b}}\rangle = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2},$$

[9] If we have probability p_2 of +2 occurring and probability $p_{-2} = 1-p_2$ of −2 occurring, then the average result is

$$\left|\langle (a+a')b+(a'-a)b'\rangle\right| = |(+2)p_2+(-2)(1-p_2)| = |4p_2-2| \leq 2$$

(since $0 \leq p_2 \leq 1$).

$$\langle\sigma_{\hat{a}'}\sigma_{\hat{b}'}\rangle = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2},$$
$$\langle\sigma_{\hat{a}}\sigma_{\hat{b}'}\rangle = \cos\frac{3\pi}{4} = -\frac{\sqrt{2}}{2},$$

and we get

$$\left|\langle\sigma_{\hat{a}}\sigma_{\hat{b}}\rangle + \langle\sigma_{\hat{a}'}\sigma_{\hat{b}}\rangle + \langle\sigma_{\hat{a}'}\sigma_{\hat{b}'}\rangle - \langle\sigma_{\hat{a}}\sigma_{\hat{b}'}\rangle\right| = 2\sqrt{2} \nleq 2.$$

Experiments confirm that QM indeed holds in this case, and therefore QM cannot be a local hidden variables theory. Again, local here means that every particle has its own set of hidden variables which determine its behavior, regardless of what the other particles do (once the hidden variables are determined).

Note, that it can be shown, that for two spins the maximum violation is when the absolute value equals $2\sqrt{2}$ as we got here.

4.2.3. Bell's inequalities. Bell was the first to give a proof that QM contradicts *local* hidden variables. Like CHSH (who came after Bell) he found an inequality which a local hidden variables theory must obey. Since quantum mechanics does not necessarily *always* obey the inequality it cannot be a *local* hidden variables theory. We shall now develop this inequality .

Assume two spins emitted with opposite spins, as before. We measure spin 1, either in direction $\hat{a}$ or in direction $\hat{c}$ and we measure spin 2, either in direction $\hat{b}$ or in direction $\hat{c}$ (the same $\hat{c}$ as a for spin 1). We shall denote the results of such measurements as a, b, c_1, c_2 respectively (c_i the result of measuring spin i in direction $\hat{c}$). Since the spins are in opposite direction we shall use

$$c \equiv c_1 = -c_2.$$

Note, that we can measure either of the pairs (a,b), (a,c) or (b,c), but not all three quantities (a, b and c). However, since we assume hidden variables we can know the results of all three quantities in advance (even if we measure only two of them). Since the spins are in opposite direction (and the measurements take values of ±1) we can write (explained next)

$$\pm a(b - c_2) = (1 + bc_1).$$

This is true since if $b = c_2$, then $b = -c_1$ and both sides give zero. On the other hand, if $b = -c_2$, then both sides give 2, up to a sign. Since the result has a $\pm$ on the left and $-1 < \langle bc_1\rangle < 1$ (and therefore $1 + \langle bc_1\rangle > 0$), then taking the average on all possible hidden variables gives[10]

$$|\langle ab\rangle - \langle ac_2\rangle| \leq 1 + \langle bc_1\rangle,$$

or using $c = c_1 = -c_2$

$$|\langle ab\rangle + \langle ac\rangle| \leq 1 + \langle bc\rangle,$$

The Bell states we already encountered are also called *maximally entangled*. They are maximally entangled in the respect that they give the maximal violation of the Bell inequalities

$$\psi^- = \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B\right)$$
$$\psi^+ = \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right)$$
$$\phi^- = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B - |1\rangle_A|1\rangle_B\right)$$

[10] When we average over $a(b - c_2)$, we sometimes add positive values and sometimes negative values, and so we have

$$|\langle a(b-c_2)\rangle| \leq \langle |a(b-c_2)|\rangle = \langle |1 + bc_1|\rangle = \langle 1 + bc_1\rangle,$$

where the last equality is due to the fact that $1 + bc_1 \geq 0$.

$$\phi^+ = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right),$$

4.3. Contextuality[11]

4.3.1. Definition of Non-Contextuality. Non-contextuality is a hidden variables "theory". Which assumes:

Non-contextuality: The result of a measurement is independent of whether other compatible (i.e. commuting) measurements are made. If $[A,B] = [A,C] = 0$ then measuring A; measuring A and B ; or measuring A and C would all give the same result for A.

Functional consistency: If $[A,B] = 0$ and measuring A,B would give (respectively) α,β, then measuring $f(A,B)$ would give $f(\alpha,\beta)$. The result $f(\alpha,\beta)$ may be assumed to have been measured even if the measurement was never taken

NOTE. Non-contextuality cannot be tested experimentally since one cannot make the different measurements on a state: once measure only A and once both A,B. Once a measurement is made the state collapses other compatible tests, will leave α (the result of measuring A) unchanged.

4.3.1.1. *Mathematical formulation.* Non-contextuality means that one may define a truth function $t(P)$, where $t(P) \in \{0,1\}$ (i.e. a value of either 0 or 1), such that for *every complete* set of orthogonal projections $\{P_i\}$

$$\sum_i P_i = I \quad (P_iP_j = \delta_{ij}P_i, P_i^\dagger = P_i),$$

the truth function obeys

$$\sum_i t(P_i) = 1 \quad (t(P) \in [0,1]).$$

The truth function t tells us in each possible basis (depending on our measuring device) which value we would measure. The difference between the non-contextuality and the standard case, is that in the standard case the truth function t gives the probability and therefore may return any value between 0 and 1, i.e. $t(P) \in [0,1]$.

4.3.2. Contradicting Non-Contextuality. Contradicting non-contextuality is achieved, not by comparing it to experiments, but rather, by showing that it is logically inconsistent (assuming a continuous Hilbert space of 3 dimensions or higher). Two theorems to prove this are the Gleason theorem and the Kochen-Specker theorems (Bell also had one). For a 4-dimensional Hilbert space Mermin

4.3.2.1. *The Gleason theorem.* Gleason replaced the usual axioms of QM by a smaller (more abstract) set of axioms:

(1) Elementary tests (yes-no questions) are represented are represented by projectors in a complex vector space.
(2) Compatible tests (yes-no questions that can be answered simultaneously) correspond to commuting projectors.
(3) If P_u and P_v are orthogonal projectors, then the projector $P_{uv} \equiv P_u + P_v$ has the expectation value

$$\langle P_{uv}\rangle = \langle P_u\rangle + \langle P_v\rangle$$

This new set does not contradict the regular axioms, and therefore any result obtained from it must also be true for the standard set.

[11] Largely based on the book of A. Peres[1].

THEOREM. *The above axioms plus continuity of the vector space require that the expectation value of any Projector P must be of the form*

$$\langle P\rangle = Tr\left(\rho P\right) \quad \Rightarrow \langle A\rangle = Tr\left(\rho A\right),$$

where ρ *is a non-negative operator with unit trace (i.e. a density matrix) which depends only on the state of the system; not on the "quantity" measured.*

If we now assume that a truth function t indeed exists then it must obey $\langle P\rangle = t(P)$. However, contrary to the truth function (in non-contextuality) which returns discrete values (0 or 1), it is clear that the function $\langle P\rangle = \mathrm{Tr}\left(\rho P\right)$ would return a continuous spectrum of values (if the projections P are continuous). This gives a contradiction and therefore non-contextuality contradicts Gleason's axioms and thus also the standard axioms of QM.

4.3.2.2. *The Kochen-Specker theorem.*

THEOREM. *In a Hilbert space of 3-dimensions or higher, it is* impossible *to define a truth function t which associates a value of either* 0 *or* 1 *with every possible projection P such that if*

$$\sum_i P_i = I \quad and \quad [P_i, P_j] = 0,$$

then

$$\sum_i t(P_i) = 1 \quad where \quad t(P_i) \in \{0,1\}$$

PROOF. (Due to Peres)

We start by proving the theorem for the case of 3 dimensions. Instead of referring to projections one may use the vectors defining them: if u is a vector, then it defines the projection $P_u \equiv uu^\dagger$. More precisely, it is sufficient to refer to rays, since the length (including negative lengths) plays no role. A complete set of commuting projections may therefore be defined by a complete set of orthogonal states/vectors/rays. the truth function t associates with each such ray a value of either 0 or 1.

The proof of the theorem has the following general form:

- Choose several complete sets of orthogonal rays, some of them sharing the same rays (but of course not sharing all of the rays). The same ray, in different sets, must of course correspond to the same value of the truth-function t, in all sets.
- Since some sets share rays, this creates constraints on the truth values allowed in different sets. The proof shows, that these constraints cannot all be maintained without a creating a contradiction (for all possible truth functions).

Since the 3-dimensional Hilbert space is isomorphic to $\mathbb{R}^3$ we may work in $\mathbb{R}^3$. We shall study here only 33 different rays[13]. The possible values of the ray components treated will be $0, \pm 1, \pm\sqrt{2}$, where for simplicity of notation $\sqrt{2}$ will be denoted as 2; and -1, $-\sqrt{2}$ will be denoted as $\bar{1}$, $\bar{2}$ respectively. Note that the 33 rays are not all the possible rays one can construct using the given components (for example the ray 111 won't be used). One important feature of the set of rays is that it has the rotation symmetry of a cube. The proof is given in the following table. In each row a set of three orthogonal rays are given under the "Orthogonal triad"; one of these must correspond to a truth value of 1 (referred to as **green**) and the other two must correspond to a truth value of 0 (referred to as red). The **green** (truth value 1) ray is written first in bold-face and then the other two (red — truth value 0). If the red rays have already been mentioned in a previous row they are written in italics. If needed later, more rays, orthogonal to the **green** (truth value 1) ray are also given under the column "Other rays". These extra rays must be red (truth value 0), since they are orthogonal to to the green ray. The third column explains why the first ray was chosen as green.

[13] This is a subset of all possible rays but it suffices to show a contradiction.

Orthogonal triad			Other rays		The first ray is **green** because of
001	100	010	110	$\bar{1}10$	arbitrary choice of *z* axis
101	$\bar{1}01$	*010*			arbitrary choice of *x* vs. $-x$
011	$0\bar{1}1$	*100*			arbitrary choice of *y* vs. $-y$
$\mathbf{1\bar{1}2}$	$\bar{1}12$	*110*	$\bar{2}01$	021	arbitrary choice of *x* vs. *y*
102	$\mathit{\bar{2}01}$	*010*	$\bar{2}11$		orthogonality to 2nd and 3rd rays
211	$\mathit{0\bar{1}1}$	$\mathit{\bar{2}11}$	$\bar{1}02$		orthogonality to 2nd and 3rd rays
201	*010*	$\mathit{\bar{1}02}$	$\bar{1}\bar{1}2$		orthogonality to 2nd and 3rd rays
112	$\mathit{1\bar{1}0}$	$\mathit{\bar{1}\bar{1}2}$	$0\bar{2}1$		orthogonality to 2nd and 3rd rays
012	*100*	$\mathit{0\bar{2}1}$	$1\bar{2}1$		orthogonality to 2nd and 3rd rays
121	$\mathit{\bar{1}01}$	$\mathit{1\bar{2}1}$	$0\bar{1}2$		orthogonality to 2nd and 3rd rays

From the table, the rays 100 (first row), 021 (fourth row), and 012 (last row) are all red (truth value 0). However these three rays are all orthogonal to one another. This creates a contradiction since this gives $\sum t(u) = 0$ instead of $\sum t(u) = 1$, as required by non-contextuality for complete orthogonal rays/vectors/states.

The proof so far has been for 3 dimensions; for higher dimensions $d > 3$ one can use the same proof[14] but add $d-3$ rays which are orthogonal to all the 33 used above (after adding to all the rays here $d-3$ components of 0, in order to make them d-dimensional). The *same* $d-3$ rays are added to the orthogonal sets of each row in the table (making each a set of d orthogonal rays). These new $d-3$ rays are always red (truth value 0) due to the first row. Since, fundamentally, the same table is used, then the same contradiction appears. □

4.3.2.3. *Mermin's proof (4 dimensions).* Mermin has given a simple proof contradicting the premises of non-contextuality in 4 dimensions. He examined the following array of operators[15]

$$\begin{array}{ccc} \mathbb{1}\otimes\sigma_z & \sigma_z\otimes\mathbb{1} & \sigma_z\otimes\sigma_z \\ \sigma_x\otimes\mathbb{1} & \mathbb{1}\otimes\sigma_x & \sigma_x\otimes\sigma_x \\ \sigma_x\otimes\sigma_z & \sigma_z\otimes\sigma_x & \sigma_y\otimes\sigma_y \end{array}.$$

In this array all the operators have eigenvalues of ±1, and the three operators in each row, as well as in each column, commute with each other. Further more, the product of the first two operators (from the left) in each row, and the first two (from the top) in each column, give the third operator in the row/column. The only exception, is in the final column, where the product gives $-\sigma_y\otimes\sigma_y$ instead of $-\sigma_y\otimes\sigma_y$.

Now, if non-contextuality is possible, then by choosing the values (±1) for the four operators determines the values for the rest of the array [e.g. if $\mathbb{1}\otimes\sigma_z$ would return 1 and $\sigma_z\otimes\mathbb{1}$ would return -1, then $\sigma_z\otimes\sigma_z = (\mathbb{1}\otimes\sigma_z)(\sigma_z\otimes\mathbb{1})$ would return $-1 = 1\cdot(-1)$]. However, since the product of the first two operators in the lower row ($\sigma_y\otimes\sigma_y$) gives minus the product of the first two operators in the third column ($-\sigma_y\otimes\sigma_y$), then there is no possible choice of values ±1 which will *not* lead to a contradiction. Mathematically, if

[14] For 4 dimensions there is also a different proof using only 24 rays.

[15] Reminder: The Pauli matrices are

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and they obey the relations

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbb{1}$$

$$\begin{array}{lclcl} \sigma_x\sigma_y = i\sigma_z & ; & \sigma_y\sigma_x = -i\sigma_z & \Rightarrow & [\sigma_x,\sigma_y] = 2i\sigma_z, \\ \sigma_z\sigma_x = i\sigma_y & ; & \sigma_x\sigma_z = -i\sigma_y & \Rightarrow & [\sigma_z,\sigma_x] = 2i\sigma_y, \\ \sigma_y\sigma_z = i\sigma_x & ; & \sigma_z\sigma_y = -i\sigma_x & \Rightarrow & [\sigma_y,\sigma_z] = 2i\sigma_x, \end{array}$$

which may be summarized by

$$\sigma_i\sigma_j = \delta_{ij}\mathbb{1} + i\varepsilon_{ijk}\sigma_k.$$

there exist a value function V which gives the value of the operator that *would* have been measured, then using the assumption of functional consistency premise on the last row gives

$$\begin{aligned} V(\sigma_y \otimes \sigma_y) &= V(\sigma_x \otimes \sigma_z)V(\sigma_z \otimes \sigma_x) = [V(\mathbb{1} \otimes \sigma_z)V(\sigma_x \otimes \mathbb{1})][V(\sigma_z \otimes \mathbb{1})V(\mathbb{1} \otimes \sigma_x)] \\ &= V(\mathbb{1} \otimes \sigma_z)V(\sigma_x \otimes \mathbb{1})V(\sigma_z \otimes \mathbb{1})V(\mathbb{1} \otimes \sigma_x). \end{aligned}$$

On the other hand, using functional consistency on the last column gives

$$\begin{aligned} V(\sigma_y \otimes \sigma_y) &= -V(\sigma_z \otimes \sigma_z)V(\sigma_x \otimes \sigma_x) = -[V(\mathbb{1} \otimes \sigma_z)V(\sigma_z \otimes \mathbb{1})][V(\sigma_x \otimes \mathbb{1})V(\mathbb{1} \otimes \sigma_x)] \\ &= -V(\mathbb{1} \otimes \sigma_z)V(\sigma_x \otimes \mathbb{1})V(\sigma_z \otimes \mathbb{1})V(\mathbb{1} \otimes \sigma_x). \end{aligned}$$

Thus we have found that $V(\sigma_y \otimes \sigma_y) = -V(\sigma_y \otimes \sigma_y)$ which is impossible since the eigenvalues of all our operators (and hence the allowed values to be measured) are ± 1. This contradiction means once again that the assumptions of non-contextuality are contradict quantum mechanics.

CHAPTER 5

Uses of Entanglement

5.1. Encoding information

Recall the four Bell states (which are maximally entangled)

$$\begin{aligned}
\psi^- &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B\right),\\
\psi^+ &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B\right),\\
\phi^- &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B - |1\rangle_A|1\rangle_B\right),\\
\phi^+ &= \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right).
\end{aligned}$$

These four states span the whole Hilbert space of two spin $\frac{1}{2}$ particles.

We now define two operators B_1 and B_2

$$\begin{aligned}
B_1 &\equiv \sigma_x^A \sigma_x^B,\\
B_2 &\equiv \sigma_z^A \sigma_z^B,
\end{aligned}$$

which commute[1]

$$[B_1, B_2] = 0.$$

Each of these two operators have two eigenvalues of ± 1

Bell state	eigenvalue B_1	eigenvalue B_2
ψ^+	$+1$	-1
ψ^-	-1	-1
ϕ^+	$+1$	$+1$
ϕ^-	-1	$+1$

We see that measuring a single operator, cannot distinguish between the four Bell states, but measuring both operators (they commute) determines a single state (see which eigenvalues are measured for each operator and compare to the above table). The only problem is that the two operators B_1 and B_2 are both non-local: they operate simultaneously on both particle A and on particle B (even when they are far away).

We saw that we can encode 2 bits of information (the eigenvalue of B_1 and eigenvalue of B_2) in the four Bell states. Now, let us assume that Charlie creates one of the four Bell states and gives one particle (particle A) to Alice and one to Bob (particle B). We might ask whether Alice and Bob can determine the given state using only *local operations* (LO) and *classical communication* (CC) where local operations means that Alice can use any operator which operates only on particle A (and maybe other particles which belong to Alice) and Bob can perform any operation which operates only on particle B (and maybe

[1] They commute since we are dealing with two particles and not just one. Using $\sigma_z\sigma_x = -\sigma_x\sigma_z = i\sigma_y$, we find that

$$\begin{aligned}
[\sigma_x^A\sigma_x^B, \sigma_z^A\sigma_z^B] &= \sigma_x^A\sigma_x^B\sigma_z^A\sigma_z^B - \sigma_z^A\sigma_z^B\sigma_x^A\sigma_x^B\\
&= i^2\sigma_y^A\sigma_y^B - (-i^2)\sigma_y^A\sigma_y^B = 0.
\end{aligned}$$

other particles which belong to Bob).[2] Classical communication means that Alice and Bob may transmit classical bits between them, but not qubits (Alice can't send particle A to Bob, but she can pass a sheet of paper saying what was the result of her measurement $+1$ or -1). The combination of both local operations and classical communication is often denoted as *LOCC*.

Using only local operations, the best that Alice and Bob can do is extract a single bit of information. They can either both measure σ_z on their particles and compare results, or both measure σ_x and compare results. In the σ_z case, if both get a spin in the same direction, they know that the Bell state is either ϕ^+ or ϕ^-. On the other hand if the get results of opposite directions they know that the Bell state is either ψ^+ or ψ^-. If however, they measure in the σ_x direction, then if their results are in the same direction, the Bell state is either ψ^+ or ϕ^+. Otherwise the Bell state is either ψ^- or ϕ^-.[3] Since they are both performing local operations, then σ_x and σ_z do not commute (unlike $\sigma_x^A\sigma_x^B$ and $\sigma_z^A\sigma_z^B$) and therefore they cannot do both types of measurements and find the specific Bell state.

We conclude therefore that using only local operations and classical communication Alice and Bob can extract only a single bit of information.

5.2. Data hiding

Assume that Charlie has one bit of information which he wants to hide from Alice and Bob, where Alice and Bob may only perform local operations and use classical communication. One method of doing this[4] is if Charlie produces n states, each selected randomly (equal probability) from the four Bell states and gives Alice the first qubit of each pair and Bob the second qubit of each pair. Charlie encodes his bit b by making sure that ψ^- appear an odd number of times if $b = 0$ and appear an even number of times if $b = 1$.

It can be shown that by performing measurements, Alice and Bob have a chance of $\left(\frac{1}{2}\right)^n$ to find with *certainty* the bit b.

5.3. Cryptography

Lets assume that Alice and Bob want to communicate (send messages between them), but that Eve wants eavesdrop to their messages. Alice and Bob of course want to prevent this.

The solution to this problem is simple, and is the same classically and QM (we shall see the difference later on). First, Alice and Bob agree (before hand) on a common key K which is a sequence of L bits, e.g.

$$K = 01100\ldots1.$$

[2] In the $\mathcal{H}_A \otimes \mathcal{H}_B$ Hilbert space, a local operation of Alice would be written as

$$U_A \otimes \mathbb{1}_B,$$

and similarly for a local operation of Bob.

[3] Using

$$\begin{array}{l} |\uparrow_x\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle + |\downarrow_z\rangle\right) \\ |\downarrow_x\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle - |\downarrow_z\rangle\right) \end{array} \quad \Rightarrow \quad \begin{array}{l} |\uparrow_z\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle + |\downarrow_x\rangle\right) \\ |\downarrow_z\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle - |\downarrow_x\rangle\right) \end{array}$$

one finds that

$$\begin{aligned} \psi^- &= \frac{1}{\sqrt{2}}\left(|\downarrow_x\rangle_A|\uparrow_x\rangle_B - |\uparrow_x\rangle_A|\downarrow_x\rangle_B\right), \\ \psi^+ &= \frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle_A|\uparrow_x\rangle_B - |\downarrow_x\rangle_A|\downarrow_x\rangle_B\right), \\ \phi^- &= \frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle_A|\downarrow_x\rangle_B + |\downarrow_x\rangle_A|\uparrow_x\rangle_B\right), \\ \phi^+ &= \frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle_A|\uparrow_x\rangle_B + |\downarrow_x\rangle_A|\downarrow_x\rangle_B\right). \end{aligned}$$

[4] Terhal, DiVincenzo and Leung quan-ph/0011042.

Now, in order to encrypt her message M e.g.

$$M = 01010\ldots,$$

Alice performs a xor[5] operation on her message and key K and generates a new message M', e.g.

$$M' = M \oplus K = \frac{\begin{array}{cccccc} 0 & 1 & 0 & 1 & 0 & \ldots \\ 0 & 1 & 1 & 0 & 0 & \ldots \end{array}}{\begin{array}{cccccc} 0 & 0 & 1 & 1 & 0 & \ldots \end{array}}.$$

Since Eve doesn't know the key she cannot decipher the message, however Bob which does know the key may perform another xor on the sent message M' and retrieve the original message M.

The only problem left for Alice and Bob is how to generate the key without Eve learning it as well (they must transmit messages which Eve might intercept). We shall now use quantum mechanics to generate such a key. The problem is known as *quantum key sharing*.

Let us assume that Charlie (not Eve) produces entangled states in the ψ^- Bell state

$$\psi^- = \frac{1}{\sqrt{2}} \left(|0\rangle_A |1\rangle_B - |1\rangle_A |0\rangle_B \right),$$

and each times sends one (qubit) of the pair to Alice and the second (qubit) to Bob. Alice and Bob can measure their qubits in the z-direction, the result will random but correlated (if Alice gets "up" the Bob gets "down" and vice versa) and so they can create their key. However Eve, since she knows the direction Alice and Bob measure in, can learn the key with out Alice and Bob finding out about it. She Basically has to measure the spin in the z-direction of Bob's (or Alice's) particle and then let it pass on to Bob (or Alice). Bob will measure the same result as Eve (due to the collapse) and this result will be correlated to Alice's result).

Let us assume however that making a measurement destroys the particle (but unitary operators, do not), can Eve still measure the spin without Alice and Bob knowing about it? Yes she can. Assume that Eve has her own spin $\frac{1}{2}$ particle in the "up" state. The total state (Alice Bob and Eve) will now be

$$|\psi_0\rangle_{ABE} = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle_A |\downarrow\rangle_B - |\downarrow\rangle_A |\uparrow\rangle_B \right) |\uparrow\rangle_E.$$

We are now looking for a unitary operator such that

$$|\psi_0\rangle_{ABE} \xrightarrow{U_{AE}} \frac{1}{\sqrt{2}} \left[\left(|\uparrow\rangle_A |\uparrow\rangle_E \right) |\downarrow\rangle_B - \left(|\downarrow\rangle_A |\downarrow\rangle_E \right) |\uparrow\rangle_B \right].$$

This transformation leaves particle E unaffected if A is in the spin up state, and it flips the spin of particle E if A is in the down state. It can be written explicitly as[6]

$$\begin{aligned} U_{AE} &= U_{\text{CNOT}} = |\uparrow\rangle_{AA}\langle\uparrow| \otimes \mathbb{1}_E + |\downarrow\rangle_{AA}\langle\downarrow| \otimes \sigma_x^E \\ &= \frac{1}{2}(\mathbb{1}_A + \sigma_z^A) \otimes \mathbb{1}_E + \frac{1}{2}(\mathbb{1}_A - \sigma_z^A) \otimes \sigma_x^E \end{aligned}$$

[5] The xor operation is denoted by $\oplus$ and is addition modulo 2 i.e.

$$0 \oplus 0 = 0$$
$$0 \oplus 1 = 1 \oplus 0 = 1$$
$$1 \oplus 1 = 0.$$

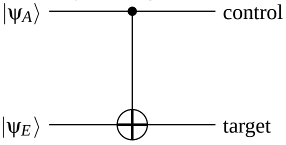
[6] Recall that

$$\sigma_x |\uparrow_z\rangle = |\downarrow_z\rangle \quad ; \quad \sigma_x |\downarrow_z\rangle = |\uparrow_z\rangle.$$

or in matrix form

$$U_{\text{CNOT}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} & 0 \\ 0 & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \end{pmatrix}.$$

This unitary operator is known as a *controlled-not* (CNOT). The particle A which does not change (but determines how E will change) is called the *control*, while particle E which may change is called the *target*. The CNOT is symbolized as



.

Another way to write the CNOT is to use $U_{\text{CNOT}} = \frac{1}{2}(\mathbb{1}_A + \sigma_z^A) \otimes \mathbb{1}_E + \frac{1}{2}(\mathbb{1}_A - \sigma_z^A) \otimes \sigma_x^E$, which can also be written as

$$U_{\text{CNOT}} = \mathbb{1}_A \otimes \mathbb{1}_E - \frac{1}{2}(\mathbb{1}_A - \sigma_z^A)(\mathbb{1}_E - \sigma_x^E).$$

Now, since $\sigma_{\hat{k}}^2 = \mathbb{1}$ then

$$(\mathbb{1} - \sigma_{\hat{k}})^2 = 2(\mathbb{1} - \sigma_{\hat{k}}) \Rightarrow (\mathbb{1} - \sigma_{\hat{k}})^n = 2^{n-1}(\mathbb{1} - \sigma_{\hat{k}}) \quad (n \neq 0).$$

Therefore we can write[7]

$$U_{\text{CNOT}} = e^{-i\frac{\pi}{4}(\mathbb{1}_A - \sigma_z^A)(\mathbb{1}_E - \sigma_x^E)}.$$

We now return to the key sharing problem. We saw that if Eve knows in which direction Alice and Bob measure their spins, then she can find out the key, without them knowing about it.[8] This is true as long as Alice and Bob always measure in the z-axis. If they suddenly switch to measuring in the x-axis (and Eve keeps using the same CNOT) then they will now see that someone is interfering since they will now find that

$$\langle \sigma_x^A \sigma_x^B \rangle = 0$$

if Eve is using the previous CNOT, where as if Eve is not listening then they would find

$$\langle \sigma_x^A \sigma_x^B \rangle = -1.$$

What Alice and Bob can do therefore is to measure their spins in random direction independent of the other: Alice chooses randomly on her side in which direction to measure, x or z, and Bob chooses randomly on his side if to measure in the x or z direction. After performing all the measurements Alice and Bob Publish in the open the direction and result of some of their measurements (but not all). From those measurements which they both performed in the same direction they find the average of the product. If it is -1 then with a good probability, Eve did not listen in, and if it is closer to 0 then Eve did listen in (neglecting noise in the system). If they conclude that Eve did not listen in, they can

[7] Since σ_z^A, σ_x^B each commute with themselves and with each other, we may write

$$U_{\text{CNOT}} = e^{-i\frac{\pi}{4}} e^{i\frac{\pi}{4}\sigma_z^A} e^{i\frac{\pi}{4}\sigma_x^E} e^{-i\frac{\pi}{4}\sigma_z^A \sigma_x^E}.$$

The operators $e^{i\frac{\pi}{4}\sigma_z^A}$ and $e^{i\frac{\pi}{4}\sigma_x^E}$, are local operations, and $e^{-i\frac{\pi}{4}}$ is just a phase. Therefore, up to local operations and a phase we may write that

$$U_{\text{CNOT}} = e^{-i\frac{\pi}{4}\sigma_z^A \sigma_x^E} \quad \text{(up to local operations and a phase)}.$$

[8] We can take the partial trace over Eve's particle and we'll get different reduced density matrices ρ_{AB} if Eve used the CNOT and if she hadn't. However

$$\text{Tr}(\rho_{AB} \sigma_z^A \sigma_z^B) = -1$$

in both cases, so Alice and Bob cannot know that Eve has been listening (and if one gets "up" the second will measure "down" so they cannot in fer from this any change).

publish the rest of the directions they measured in (but this time without the results). From the measurements which they both made in the same direction the can now produce the key K.

Alice and Bob can achieve the previous protocol even without having entangled states between them. Alice can create spins in one of the four states

$$|\uparrow_z\rangle, |\downarrow_z\rangle, |\uparrow_x\rangle, |\downarrow_z\rangle.$$

She then sends them to Bob who measures them randomly in either the x direction or z direction. From here on the protocol is the same as before (except that this time if Alice and Bob measure/create in the same direction they will find the same result, both "up" or both "down" unlike the previous protocol, in which if Alice measured "up" then Bob measured "down" and vice versa).

5.4. teleportation

Assume that Alice and Bob have an entangled state $|\phi^+\rangle_{ab}$ between them

$$|\phi^+\rangle_{ab} = \frac{1}{\sqrt{2}}\left(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b\right).$$

Now, Alice has a third particle A in state $|\psi\rangle$

$$|\psi\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A,$$

and she wants to pass the state itself (not the particle) to Bob.[9] The sate of the whole system (all three particles) is $|\psi\rangle_A|\phi^+\rangle_{ab}$, however it can also be written as[10]

$$\begin{aligned}
|\psi\rangle_A|\phi^+\rangle_{ab} &= (\alpha|0\rangle_A + \beta|1\rangle_A)\frac{1}{\sqrt{2}}\left(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b\right) \\
&= \frac{1}{\sqrt{2}}\left(\alpha|0\rangle_A|0\rangle_a|0\rangle_b + \alpha|0\rangle_A|1\rangle_a|1\rangle_b + \beta|1\rangle_A|0\rangle_a|0\rangle_b + \beta|1\rangle_A|1\rangle_a|1\rangle_b\right) \\
&= \frac{|\phi^+\rangle_{Aa} + |\phi^-\rangle_{Aa}}{2}\alpha|0\rangle_b + \frac{|\psi^+\rangle_{Aa} + |\psi^-\rangle_{Aa}}{2}\alpha|1\rangle_b \\
&\quad + \frac{|\psi^+\rangle_{Aa} - |\psi^-\rangle_{Aa}}{2}\beta|0\rangle_b + \frac{|\phi^+\rangle_{Aa} - |\phi^-\rangle_{Aa}}{2}\beta|1\rangle_b \\
&= \frac{1}{2}|\phi^+\rangle_{Aa}(\alpha|0\rangle_b + \beta|1\rangle_b) + \frac{1}{2}|\psi^+\rangle_{Aa}\sigma_x^b(\alpha|0\rangle_b + \beta|1\rangle_b) \\
&\quad + \frac{1}{2}|\psi^-\rangle_{Aa}(-i)\sigma_y^b(\alpha|0\rangle_b + \beta|1\rangle_b) + \frac{1}{2}|\phi^-\rangle_{Aa}\sigma_z^b(\alpha|0\rangle_b + \beta|1\rangle_b) \\
&= \frac{1}{2}\left[|\phi^+\rangle_{Aa}|\psi\rangle_b + |\psi^+\rangle_{Aa}\sigma_x^b|\psi\rangle_b + |\psi^-\rangle_{Aa}(-i)\sigma_y^b|\psi\rangle_b + |\phi^-\rangle_{Aa}\sigma_z^b|\psi\rangle_b\right].
\end{aligned}$$

We see now that if Alice makes a measurement in the basis of the Bell states of particles a, A then particle B would collapse to one of the states of the form $\sigma_i^b|\psi\rangle_b$ ($i = 0, 1, 2, 3$ where $\sigma_0 = \mathbb{1}$).[11] The result of Alice's measurement is two bits (the two bits needed to determine which of the four Bell states she found). Alice can send the two bits to Bob,

[9] We may assume that particles a, b, A are of different types or of similar types - there is no restriction.

[10] Recall that

$$\begin{aligned}
|\psi^-\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_a - |1\rangle_A|0\rangle_a\right), \\
|\psi^+\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}}\left(|0\rangle_A|1\rangle_a + |1\rangle_A|0\rangle_a\right), \\
|\phi^-\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_a - |1\rangle_A|1\rangle_a\right), \\
|\phi^+\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_a + |1\rangle_A|1\rangle_a\right).
\end{aligned}$$

[11] OK, for $i = 2$ the state is $i\sigma_y^b|\psi\rangle_b$ not $\sigma_y^b|\psi\rangle_b$.

who can then perform on his particle b the appropriate inverse operator (in this case the same σ_i). After this operation Bob will hold in his hand particle b in a state which was previously associated with particle A held by Alice.

the following things should be noted:

- The thing that was passed between Alice and Bob (besides the two bits of information), was a state, not a particle. The state which once described particle A now describes particle b.
- the no-cloning theorem still holds. After the process, the particle A is no longer in its original state but in an entangled state with a.
- The two bits of information we use are completely random (since the collapse is random to one of four possible states). So they are not the ones carrying the information.
- Although Alice sent Bob two bits of information, Bob was able to extract two continuous variables (α and β). However, Bob never knows what these two variables were. He only knows that they were passed correctly.

One consequence of teleportation is that allows one to do non-local operations (assume you have an entangled state). Simply teleport one state to a particle in the vicinity of the second particle, make the measurement there (locally) and then teleport back the new state of the particle.

5.5. Ramsey spectroscopy

5.6. Remote operations

To our list of types of bits we now add the ebit which is simply an entangled state. By changing the basis we choose for each particle (or equivalently performing a unitary operation non each) we can always bring to the Bell state $|\phi^+\rangle$

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

Bennett wrote “equations” which describe the different process. The “equations” were constructed from ebits, qubits and (classical) bits. For example for teleportation one needs an entangled state and to pass two classical bits (the result of Alice’s measurement on her two states). Since teleportation is actually the communication of one qubit (the state, not the particle is passed from Alice to Bob), then it can be written as

$$1\text{ebit}_{AB} + 2\text{bit}_{A\to B} \Rightarrow 1\text{qbit}_{A\to B}.$$

One could also use teleportation to create an entangled state. Simply do a local operation (say a CNOT) on two particles and entangle them. Then teleport the state of one of them to a distant particle and you have two distant entangled particles. This would be written as

$$\text{teleportation} \Rightarrow \text{ebit}_{AB}.$$

Dense coding may also be described in this manner. In dense coding we started with an entangled state, Alice then performed a local operation on her particle (encoded two bits in to it) and then sent the particle to Bob (equivalent to teleportation). In Bennett’s language this would be written as

$$\text{teleportation}_{A\to B} + \text{ebit}_{AB} \Rightarrow 2\text{bit}_{A\to B}.$$

In quantum computation we would like to create interactions between distant particles, i.e. non-local (or remote) operations. However, the rules of the game are as follows:

Locality: Only local operations are allowed. This includes local unitary operators and local measurements. Note that local operation alone cannot create entanglement.

- **Classical communication:** Only classical communication is allowed between our systems (only passing of classical bits). It is not allowed to exchange quantum particles (i.e. not allowed to exchange qubits). Note that classical communication together with local operations, still cannot create entanglement.
- **Entanglement resources:** There are pairs of particles in entangled states, ready to be used. These pairs are given/prepared before the beginning of the calculation and no more pairs may be added after the start of the calculation.

Given the rules of the game we would like to create protocols that will produce the effect of non-local unitary operations. These protocols must give the desired operator regardless of the state we perform the operation on. Such a process will require both the use of entangled pairs (ebits) and the communication of classical bits (usually random ones).

One simple method of doing any non-local operation is with two teleportations. Simply teleport one state to the locality of the second perform a local operation on the two and then teleport the state of one particle back (note that the resultant state will not be back on the original state, but rather on a new one). Such a procedure (two teleportations) will require from us to use 2 ebits and send 4 bits of classical information (1 ebit and 2 bits for each teleportation - see above).

We would like to see if we can do this more efficiently. We shall study the CNOT operation. We shall see that one teleportation and one classical bit suffice to create a CNOT.

CLAIM. A CNOT is equivalent to a teleportation in the sense that

$$\text{CNOT} + 1\ \text{bit}_{A\to B} \Rightarrow \text{teleport}_{A\to B},$$

$$\text{teleport}_{A\to B} + 1\ \text{bit}_{A\to B} \Rightarrow \text{CNOT}.$$

PROOF. To prove the claim we assume an initial state

$$|\psi\rangle_A|0\rangle_B \equiv (\alpha|0\rangle_A + \beta|1\rangle_A)\,|0\rangle_B.$$

If we perform a remote CNOT, with $|\psi\rangle_A$ as the control then we get

$$\begin{aligned}\text{CNOT}\,|\psi\rangle_A|0\rangle_B &\equiv \alpha|0\rangle_A|0\rangle_B + \beta|1\rangle_A|1\rangle_B\\ &= \alpha\frac{|\uparrow_x\rangle_A + |\downarrow_x\rangle_A}{\sqrt{2}}|0\rangle_B + \beta\frac{|\uparrow_x\rangle_A - |\downarrow_x\rangle_A}{\sqrt{2}}|1\rangle_B\\ &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_A\,(\alpha|0\rangle_B - \beta|1\rangle_B) + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_A\,(\alpha|1\rangle_B + \beta|0\rangle_B)\,.\end{aligned}$$

Now if Alice measures σ_x of particle A and sends the result (1 bit) to Bob, then Bob can perform on his particle σ_z if Alice measured “up” or perform σ_x if Alice measured “down”. By doing this particle B will now be in state ψ and we have teleportation, where we have used a remote CNOT and a transfer of 1 classical bit of information (the result of Alice’s measurement).

For how to do the opposite: create a CNOT using teleportation and a single bit see the stator below (creating a stator). □

5.7. State-operators (stators)[12]

We define a *state-operator*, or for short *stator*, as a “creature” which is a combination of states in one Hilbert space and operators in another. Generally speaking it will be written as

$$S = \sum c_i|i\rangle_A \otimes O_i^B,$$

[12] See also: quant-ph/0107143

where the $|i\rangle_A$ are states in the Hilbert space $\mathcal{H}_A$ (of say, particle A), and O_i^B are operators which operate on (states in) the Hilbert space $\mathcal{H}_B$. Thus, when the stator is applied to a state in $\mathcal{H}_B$, the result is a state in $\mathcal{H}_A \otimes \mathcal{H}_B$:

$$S|\psi\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B \quad (|\psi\rangle_B \in \mathcal{H}_B).$$

Here, we shall be interested in pairs of operators which obey[14]

$$AS = BS,$$

where A operates on $\mathcal{H}_A$ and B operates on $\mathcal{H}_B$. Such pairs do not necessarily consist of two Hermitian operators, but we shall be interested in the cases where they do. For example for

$$S = |0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \sigma_z^B,$$

we have

$$\sigma_x^A S = \sigma_z^B S.$$

When a pair of operators A, B does indeed obey the relation

$$AS = BS,$$

then necessarily[15]

$$A^n S = B^n S,$$

and therefore (using a Taylor expansion)

$$f(A)S = f(B)S.$$

Specifically, for A and B which are also Hermitian, we have

$$e^{i\alpha A} S = e^{i\alpha B} S,$$

where $e^{i\alpha A}$ and $e^{i\alpha B}$ are now unitary operators (since A, B are Hermitian).

Now, let us assume that Alice has a (unitary) operator $U_\alpha = e^{i\alpha\sigma_x^A}$, and Bob wants to use the *same* parameter α in applying $e^{i\alpha\sigma_z^B}$ on his particle.[16] To do this we shall use our previous stator

$$S = |0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \sigma_z^B.$$

We start by Alice applying her operator on the stator S, i.e. performing

$$U_\alpha S = e^{i\alpha\sigma_x^A} S.$$

For the specific stator S chosen here, we have $\sigma_x^A S = \sigma_z^B S = S\sigma_z^B$, and therefore we can write

$$e^{i\alpha\sigma_x^A} S = S e^{i\alpha\sigma_z^B}.$$

Thus we get (for any $|\psi\rangle_B$)

$$U_\alpha S|\psi\rangle_B = S e^{i\alpha\sigma_z^B}|\psi\rangle_B = \left(|0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \sigma_z^B\right) e^{i\alpha\sigma_z^B}|\psi\rangle_B.$$

[14] As we shall see, we are actually interested in

$$AS = SB.$$

[15] Since A, B operate on different regions (on different Hilbert spaces), then they necessarily commute. To be more exact, we should write $A \otimes \mathbb{1}_B$ and $\mathbb{1}_A \otimes B$ instead of A and B. With this notation it is clear that

$$(A \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes B) = A \otimes B = (\mathbb{1}_A \otimes B)(A \otimes \mathbb{1}_B).$$

Now, if $AS = BS$ and as we saw $AB = BA$, then we have

$$A^2 S = A(AS) = ABS = B(AS) = B^2 S,$$

and by induction

$$A^n S = B^n S.$$

[16] Note, that Bob does not know the value α which Alice is holding. However, he wants to use the same α for his own operation.

To finish the process, Alice measures her spin and sends the result to Bob. The measurement causes a collapse into one of the two states

$$\left(|0\rangle_A \otimes \mathbb{1}_B\right) e^{i\alpha\sigma_z^B}|\psi\rangle_B = |0\rangle_A \otimes e^{i\alpha\sigma_z^B}|\psi\rangle_B \quad (\text{Alice measured } 0),$$

or

$$\left(|1\rangle_A \otimes \sigma_z^B\right) e^{i\alpha\sigma_z^B}|\psi\rangle_B = |1\rangle_A \otimes \sigma_z^B e^{i\alpha\sigma_z^B}|\psi\rangle_B \quad (\text{Alice measured } 1).$$

Now, if Alice measured 0 (and sent this to Bob), then Bob knows that he has $e^{i\alpha\sigma_z^B}|\psi\rangle_B$ on his side, which is just what he wanted. However, if Alice sent Bob a 1, then Bob knows that he has $\sigma_z^B e^{i\alpha\sigma_z^B}|\psi\rangle_B$ on his side. Bob must therefore fix his state (get rid of the extra σ_z^B operator). He does this by performing another σ_z^B on it (since $\sigma_z^2 = \mathbb{1}_{2\times 2}$), ending again with the desired result $e^{i\alpha\sigma_z^B}|\psi\rangle_B$.

We have thus seen that by using the stator S, together with Alice sending a single classical bit (0 or 1 — which Bob used to perform the right corrections), Bob was able to apply $e^{i\alpha\sigma_z^B}$ on his side, where α is any real number chosen by Alice (and unknown to Bob).

The methods just used may be generalized to also achieve a remote CNOT (up to local operations and a phase)

$$U_{\text{"CNOT"}} = e^{-i\frac{\pi}{4}\sigma_z^A\sigma_x^B} \quad (\text{A CNOT up to local operations and a phase}).$$

To do this we generalize our stator to

$$S = |\uparrow_x\rangle_a \otimes \mathbb{1}_{AB} + |\downarrow_x\rangle_a \otimes \left(\sigma_z^A\sigma_x^B\right),$$

which obeys

$$\sigma_z^a S = \sigma_z^A\sigma_x^B S = S\sigma_z^A\sigma_x^B.$$

Therefore, if Alice applies $e^{-i\frac{\pi}{4}\sigma_z^a}$, we get

$$e^{-i\frac{\pi}{4}\sigma_x^a} S = S e^{-i\frac{\pi}{4}\sigma_z^A\sigma_x^B}.$$

Thus, as in the previous case

$$\begin{aligned} e^{-i\frac{\pi}{4}\sigma_x^a} S|\varphi\rangle_A|\psi\rangle_B &= S e^{-i\frac{\pi}{4}\sigma_z^A\sigma_x^B}|\varphi\rangle_A|\psi\rangle_B \\ &= \left(|\uparrow_x\rangle_a \otimes \mathbb{1}_{AB} + |\downarrow_x\rangle_a \otimes \sigma_z^A\sigma_x^B\right) e^{-i\frac{\pi}{4}\sigma_z^A\sigma_x^B}|\varphi\rangle_A|\psi\rangle_B. \end{aligned}$$

Alice then measures the spin in the x direction of particle a. If she finds "up", then Alice performs σ_z^A on particle A and Bob performs σ_x^B on his particle B. Otherwise they do nothing. In both cases the final result is the operation $e^{-i\frac{\pi}{4}\sigma_z^A\sigma_x^B}$ on the remote (from each other) particles A and B.

5.7.1. Creating a stator.

Assume a general, two-level system, unitary operator U. We wish to construct a stator of the form[17]

$$S = |0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes U^B.$$

We start with three particles, an ancilla b and the two particles A, B. We start with the configuration

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_b + |1\rangle_A|1\rangle_b\right)|\psi\rangle_B,$$

[17] To create the more general stator

$$S = |0\rangle_A \otimes U_{B0} + |1\rangle_A \otimes U_{B1}$$

we simply need

$$S = S' U_{B0} = \left[|0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \left(U_{B1}U_{B0}^{-1}\right)\right] U_{B0}.$$

where particles A, b are entangled in advance. Bob performs a local "CNOT" between particles B and b (particle B is the target) and we get

$$\frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_b+|1\rangle_A|1\rangle_b)|\psi\rangle_B \xrightarrow{\text{CNOT}_{Bb}} \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_b+|1\rangle_A|1\rangle_b U^B)|\psi\rangle_B.$$

We now write particle b in the x basis

$$\begin{aligned}\frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_b+|1\rangle_A|1\rangle_b U^B)|\psi\rangle_B &= \frac{1}{2}(|\uparrow_x\rangle_b+|\downarrow_x\rangle_b)|0\rangle_A|\psi\rangle_B+\frac{1}{2}(|\uparrow_x\rangle_b-|\downarrow_x\rangle_b)|1\rangle_A U^B|\psi\rangle_B \\ &= \frac{1}{2}|\uparrow_x\rangle_b\left(|0\rangle_A+|1\rangle_A U^B\right)|\psi\rangle_B+\frac{1}{2}|\downarrow_x\rangle_b\left(|0\rangle_A-|1\rangle_A U^B\right)|\psi\rangle_B.\end{aligned}$$

Bob, now measures the spin of his ancilla b in the x direction. Thus collapsing the system into one of the states

$$\frac{1}{\sqrt{2}}|\uparrow_x\rangle_b\left(|0\rangle_A+|1\rangle_A U^B\right)|\psi\rangle_B \quad \sigma_x^b=\text{"up"},$$

$$\frac{1}{\sqrt{2}}|\downarrow_x\rangle_b\left(|0\rangle_A-|1\rangle_A U^B\right)|\psi\rangle_B \quad \sigma_x^b=\text{"down"}.$$

Now, Bob sends the result of his measurement to Alice, who accordingly decides, whether to perform a σ_z^A on her system (if $\sigma_x^b=$ "down") or if to do nothing ($\sigma_x^b=$ "up"). We can now disregard the particle b, and since this process was done for any general $|\psi_B\rangle$, then we can say that we performed the stator S.

The creation of a CNOT using a teleportation and a single bit is very similar. The only difference is that instead of starting with an entangled pair, we create it using teleportation (begin with two spins, locally entangle them and then teleport one to Alice/Bob).

5.8. POVM (Positive Operator Valued Measures)

When we perform regular measurements we cause the state of the system to collapse. We would like to avoid this collapse. To do this we use an auxiliary particle, called an *ancilla*, which we first interact with the system, and after words measure it - thus collapsing the the ancilla and not the system.

We shall first review the standard Von Neumann measurements. In these measurements, the operator describing the quantity measured is

$$A=\sum\lambda_i\Pi_i,$$

where Π_i is a projection on one of the orthogonal subspaces i

$$\sum\Pi_i=\mathbb{1},$$

$$\Pi_i\Pi_j=\mathbb{1}\delta_{ij},$$

and where λ_i is the eigenvalue associated with the subspace which Π_i projects on to. When we make a measurement the result is one of the λ_i, and for such a result the state $|\psi\rangle$ of the system collapses to $\Pi_i|\psi\rangle$ times a normalization

$$|\psi\rangle \xrightarrow{\text{measure } A} \frac{\Pi_i|\psi\rangle}{\langle\psi|\Pi_i|\psi\rangle}.$$

If we start with a mixture (a density matrix), then

$$\rho \xrightarrow{\text{measure } A} \frac{\Pi_i\rho\Pi_i}{\text{Tr}(\Pi_i\rho)}.$$

We now turn to the new type of measurements. We start adding an auxiliary particle, an ancilla, in a *known* state

$$|\Psi\rangle_{\text{tot}}=|\psi\rangle_{\text{sys}}|0\rangle_a.$$

We shall assume that the ancilla a belongs to a Hilbert space of dimension N_a, and therefore in some orthonormal base (which $|0\rangle_a$ belongs to)

$$\sum_{\mu=1}^{N_a} |\mu\rangle_{aa}\langle\mu| = \mathbb{1}_a.$$

We now cause the ancilla and our system to interact for a short time. The effect of this interaction may be described by a unitary operator U which operates on both $U|\Psi\rangle_{\text{tot}}$. This can also be written as

$$\begin{aligned} U|\Psi\rangle_{\text{tot}} &= \mathbb{1}_a U|\Psi\rangle_{\text{tot}} = \left(\sum_\mu |\mu\rangle_{aa}\langle\mu|\right) U|0\rangle_a|\psi\rangle_{\text{sys}} \\ &= \sum_\mu \left({}_a\langle\mu|U|0\rangle_a\right)|\mu\rangle_a|\psi\rangle_{\text{sys}}. \end{aligned}$$

If we now define the *Kraus operator*Kraus

$$M_\mu \equiv {}_a\langle\mu|U|0\rangle_a,$$

which operates on the Hilbert space of the system, then the last equation may be written as

$$U|\Psi\rangle_{\text{tot}} = \sum_\mu M_\mu|\mu\rangle_a|\psi\rangle_{\text{sys}}.$$

If we now measure μ for the ancilla, then the state would collapse to a single μ

$$U|\Psi\rangle_{\text{tot}} \xrightarrow{\text{measured } \mu} |\mu\rangle_a M_\mu|\psi\rangle_{\text{sys}},$$

this would occur with a probability prob(μ)

$$\begin{aligned} \text{prob}(\mu) &= {}_{\text{tot}}\langle\Psi|U^\dagger|\mu\rangle_{aa}\langle\mu|U|\Psi\rangle_{\text{tot}} \\ &= {}_{\text{sys}}\langle\psi|M_\mu^\dagger M_\mu|\psi\rangle_{\text{sys}}. \end{aligned}$$

Since the sum of probabilities (for all μ) must be 1, then

$$1 = \sum_\mu \text{prob}(\mu) = {}_{\text{sys}}\langle\psi|\left(\sum_\mu M_\mu^\dagger M_\mu\right)|\psi\rangle_{\text{sys}},$$

or (since this is true for any $|\psi\rangle_{\text{sys}}$) simply[19]

$$\sum_\mu M_\mu^\dagger M_\mu = \mathbb{1}_{\text{sys}}.$$

In analogy to the Von Neumann measurements, we may now write, for measurements using an ancilla

$$\begin{gathered} |\psi\rangle_{\text{sys}} \xrightarrow{\text{measured } \mu} M_\mu|\psi\rangle_{\text{sys}} \quad \text{(not normalized)}, \\ \rho \xrightarrow{\text{measured } \mu} \frac{M_\mu\rho M_\mu^\dagger}{\text{Tr}(M_\mu\rho M_\mu^\dagger)}, \\ F_\mu \equiv M_\mu^\dagger M_\mu \text{ a positive operator} \quad (\sum_\mu F_\mu = \mathbb{1}_{\text{sys}}), \\ \text{prob}(\mu) = \text{Tr}(F_\mu\rho_{\text{sys}}), \end{gathered}$$

where in the first equation we look (after the measurement) only at the system itself and disregard the ancilla, and where $M_\mu^\dagger M_\mu$ is a positive operator since we saw that $\text{prob}(\mu) = {}_{\text{sys}}\langle\psi|M_\mu^\dagger M_\mu|\psi\rangle_{\text{sys}}$. Probability is always non-negative, and $|\psi\rangle_{\text{sys}}$ could be any state, and therefore $M_\mu^\dagger M_\mu$ must be a positive operator ($M_\mu^\dagger M_\mu$ is clearly Hermitian, which is also a necessary condition).

[19] This could also be found directly from the definition of the M_μ

$$\sum_\mu M_\mu^\dagger M_\mu = \sum_\mu {}_a\langle 0|U^\dagger|\mu\rangle_{aa}\langle\mu|U|0\rangle_a = {}_a\langle 0|U^\dagger U|0\rangle_a = \mathbb{1}_{\text{sys}}.$$

Since $F_\mu \equiv M_\mu^\dagger M_\mu$ is a positive operator, then it is called a *positive operator valued measure* or *POVM* for short.

We see that we got a very similar behavior to that of the Von Neumann measurements, where the Kraus operators M_μ replace the projections Π_i. The only difference is that, here, the Kraus operators are not necessarily orthogonal, and as a consequence the number of eigenvalues μ may exceed the number of dimensions of the Hilbert space of the system itself (the dimension N_a of the space of the ancilla is arbitrary).

Note, that it may be shown that if there exists operators M_μ that obey the above rules, then there exists an appropriate ancilla for the system.

An important difference between regular (Von Neumann) measurements and the POVM ones, is that in the latter case, the results are not eigenvalues of an operator and the system alone, but rather of an operator and the system together with the ancilla. However, one can find correlations between the measured μ and the state of the system.

5.8.1. Neumark's theorem (without proof). We have just seen that by adding an ancilla and thus enlarging our Hilbert space we could reach the POVM formalism. The contrary is also true, given an n dimensional Hilbert space with a POVM set of N elements (F_μ, $\mu = 1, \ldots, N$), then we can always realize it as standard measurements in an N dimensional Hilbert space.[20] This theorem is known as *Neumark's theorem*.

5.8.2. Distinguishing between non-orthogonal states. This is especially good for distinguishing between non-orthogonal states of the system, as is shown next.

Assume two non-orthogonal states of a system

$$|\psi_1\rangle = |\uparrow_x\rangle \quad ; \quad |\psi_2\rangle = |\uparrow_z\rangle = \frac{|\uparrow_x\rangle + |\downarrow_x\rangle}{\sqrt{2}}.$$

We know that they have the same probability $\frac{1}{2}$ to occur (there are no other possibilities), and we wish to know which one has occurred (in which state the particle we are holding out of the ensemble is). If we measure σ_x, then we may get two results. If we find $\sigma_x = 1$ (probability $\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}$), we cannot deduce anything since both $|\uparrow_x\rangle$ and $|\uparrow_z\rangle$ have a non zero part which is $|\uparrow_x\rangle$. If however we measure $\sigma_x = -1$ (probability $\frac{1}{2} \cdot \frac{1}{2}$) then we know for certain that the particle was in state $|\uparrow_z\rangle$ (since only it has a non-zero component in the "down" x direction). Thus we see, that by using a standard measurement we will know, for certain, the state of the system only in $\frac{1}{4}$ of the cases.[21] In other words we do not no the answer for certain, in $\frac{3}{4}$ of the measurements.

Now, instead of making standard measurements, let us define

$$F_1 = \lambda |\downarrow_z\rangle\langle\downarrow_z|,$$
$$F_2 = \lambda |\downarrow_x\rangle\langle\downarrow_x|,$$
$$F_3 = \mathbb{1}_{\text{sys}} - F_1 - F_2.$$

Note that in these definitions, F_1 uses a state orthogonal to $|\psi_2\rangle$ and F_2 uses a state orthogonal to $|\psi_1\rangle$. This time, if we measure $\mu = 1$, then we know the system is in state $|\uparrow_x\rangle$ (since the probability of measuring $\mu = 1$ for the case of $|\psi_2\rangle$ is $\langle\psi_2|F_1|\psi_2\rangle = 0$), and if we measure $\mu = 2$, then we know the system is in state $|\uparrow_z\rangle$. If however, we measure $\mu = 3$,

[20] We assume that the Hilbert space dimensionality n is smaller than number N of F_μ's in our POVM. If we have more dimensions (in the Hilbert space), we can always make a change of basis so that only N of them will be relevant to the POVM while the rest will be independent and thus irrelevant to the problem. In Neumark's theorem, we then enlarge the number of relevant dimensions.

[21] We could do the same using σ_z instead. This time we would also know in $\frac{1}{4}$ of the cases which direction the spin was, however this tome those cases will tell us that the particle was in the "up" z direction. Measuring in any other direction (except $\pm\hat{z}$ or $\pm\hat{x}$) will give us now information at all, since they all have non-zero projections on both $\hat{x}$ and $\hat{z}$.

then we cannot know the state of the system. We see that only in the case of $\mu = 3$ we cannot tell the state of the particle. The probability of $\mu = 3$ occurring is simply[22]

$$\text{prob}(\mu = 3) = 1 - \left(\frac{1}{2}\frac{\lambda}{2} + \frac{1}{2}\frac{\lambda}{2}\right) = 1 - \frac{\lambda}{2},$$

which can also be found using the trace

$$\text{prob}(\mu = 3) = \text{Tr}(\rho F_3)$$

where

$$\rho = \frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x| + \frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z|.$$

We would like to find λ such that the probability of not knowing for certain the original state will be minimal (.i.e. $\text{prob}(\mu = 3) = 1 - \frac{\lambda}{2}$ will be minimal).[23] clearly by our definitions F_1 and F_2 are positive operators (if and only if $\lambda > 0$). The condition we require is that F_3 will also be positive (and we are looking for the maximum λ which gives this). Since it is a 2×2 matrix it is enough to require that the trace and determinant both have the same sign. The optimal λ is then

$$\lambda = 2 - \sqrt{2},$$

which gives us the minimum probability of not knowing for certain

$$\text{prob}(\mu = 3) = \frac{1}{\sqrt{2}}.$$

This result is indeed better than the one we had before, with standard measurements, which gave us a chance of failure of $\frac{3}{4}$. This is indeed an improvement although not a very large one in this case.

5.9. Measure of entanglement (Distillation)

Let us assume that we have a system in a state

$$|\Psi\rangle_{ab} = \alpha|0\rangle_a|0\rangle_b + \beta|1\rangle_a|1\rangle_b \quad (|a| \leq |\beta|),$$

where we know α, β and we assume $|a| \leq |\beta|$. Unless $|\alpha|, |\beta|$ are both $\frac{1}{\sqrt{2}}$, the state is not maximally entangled (does not maximally violate the Bell inequality). We now want to distill this state, in order to get the maximally entangled state

$$|\phi^+\rangle_{ab} = \frac{1}{\sqrt{2}}\left(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b\right).$$

[22] There is a probability $\frac{1}{2}$ of state $|\psi_1\rangle$ occurring and a probability $\frac{1}{2}$ of state $|\psi\rangle_2$. The probability of measuring $\mu = 1$ and $\mu = 2$ for $|\psi_1\rangle$ are

$$\text{prob}(\mu = 1)_{|\psi_1\rangle} = \langle\psi_1|F_1|\psi_1\rangle = \frac{\lambda}{2} \quad ; \quad \text{prob}(\mu = 2)_{|\psi_1\rangle} = \langle\psi_1|F_2|\psi_1\rangle = 0$$

and similarly for $|\psi_2\rangle$

$$\text{prob}(\mu = 2)_{|\psi_2\rangle} = \langle\psi_2|F_2|\psi_2\rangle = \frac{\lambda}{2} \quad ; \quad \text{prob}(\mu = 1)_{|\psi_2\rangle} = \langle\psi_2|F_1|\psi_2\rangle = 0.$$

Recalling that each state occurs wit probability $\frac{1}{2}$, the probability of getting $\mu \neq 1, 2$ (i.e. getting $\mu = 3$) is

$$\text{prob}(\mu = 3) = 1 - \left(\frac{1}{2}\frac{\lambda}{2} + \frac{1}{2}\frac{\lambda}{2}\right).$$

[23] Actually to make sure we get the best results we should have used different λ's for positive measure

$$F_1 = \lambda_1|\downarrow_z\rangle\langle\downarrow_z|,$$

$$F_2 = \lambda_2|\downarrow_x\rangle\langle\downarrow_x|.$$

However, after all the optimization, we get the same result.

We want to do this using only local operations. To do this we shall use the *Procrustean method*. If we define the Kraus operator[24]

$$M_0 \equiv \lambda \left(\frac{\beta}{\alpha} |0\rangle_{aa}\langle 0| + |1\rangle_{aa}\langle 1| \right),$$

then whenever we measure $\mu = 0$, the state we will find is

$$|\Psi\rangle_{ab} \xrightarrow{\mu=0} \frac{M_0|\Psi\rangle_{ab}}{\sqrt{_{ab}\langle\Psi|M_0^{\dagger}M_0|\Psi\rangle_{ab}}} = \frac{\lambda}{\sqrt{_{ab}\langle\Psi|M_0^{\dagger}M_0|\Psi\rangle_{ab}}}\beta\left(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b\right),$$

where clearly (because of the normalization) we will find that

$$\frac{\lambda}{\sqrt{_{ab}\langle\Psi|M_0^{\dagger}M_0|\Psi\rangle_{ab}}}\beta = \frac{1}{\sqrt{2}}.$$

We see therefore, that if choosing such a Kraus operator, will distill our state to the Bell state whenever we measure $\mu = 0$. This will occur probability $\text{prob}(\mu = 0)$ given by

$$\begin{aligned}\text{prob}(\mu=0) &= {}_{ab}\langle\Psi|M_0^{\dagger}M_0|\Psi\rangle_{ab} = |\lambda\beta|^2\left({}_b\langle 1|_a\langle 1| + {}_b\langle 0|_a\langle 0|\right)\left(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b\right) \\ &= 2|\lambda\beta|^2.\end{aligned}$$

Clearly, to increase the probability of the desired distillation, we would like $|\lambda|$ to be as large as possible (β is given). However we cannot raise it arbitrarily since we require $M_0^{\dagger}M_0$ to be a positive operator. By our definition

$$M_0^{\dagger}M_0 = |\lambda|^2 \begin{pmatrix} \frac{\beta^*}{\alpha^*} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\beta}{\alpha} & 0 \\ 0 & 1 \end{pmatrix} = |\lambda|^2 \begin{pmatrix} \left|\frac{\beta}{\alpha}\right|^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since, however we must have[25]

$$\sum_{\mu} M_{\mu}^{\dagger}M_{\mu} = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the $M_{\mu}^{\dagger}M_{\mu}$ are all positive operators then necessarily, we must have

$$|\lambda|^2\left|\frac{\beta}{\alpha}\right|^2 \leq 1$$
$$\Rightarrow |\lambda\beta|^2 \leq |\alpha|^2.$$

Using, this last result, we see that the maximum probability possible for distillation (recall that $|a| \leq |\beta|$) is[26]

$$\text{prob}(\mu = 0) \leq 2|\alpha|^2 \leq 1.$$
$$M_1 \equiv U\left(\mathbb{1} - \sqrt{M_0^{\dagger}M_0}\right),$$
$$M_2 \equiv \mathbb{1} - M_1 - M_0,$$

where U is any arbitrary unitary operator.

[24] Note, that although M_0 operates only on the Hilbert space of particle a, the system we consider is both particles a and b. The ancilla used for the POVM measurement is a third particle.

[25] We could also define the "complementary" M_1 of M_0 by

$$\mathbb{1} = M_0^{\dagger}M_0 + M_1^{\dagger}M_1$$
$$\Rightarrow M_1 = U\sqrt{\mathbb{1} - M_0^{\dagger}M_0}.$$

Requiring that it be a positive operator, would give us the same result.

[26] The requirement that $|a| < |\beta|$, comes in the form, that if we had the opposite then $|\alpha|^2$ would be larger than $\frac{1}{2}$, and we would get a probability of finding $\mu = 0$ of $2|\lambda\beta|^2$ which is greater than 1.

5.9.1. Distillation of *n* pairs. Assume that we now have two pairs of non-maximally entangled states, where the two pairs are described by the same state which we know (we know the parameters α, β)

$$|\Psi\rangle^{\otimes 2} = (\alpha|0\rangle_a|0\rangle_b + \beta|1\rangle_a|1\rangle_b)\,(\alpha|0\rangle_{a'}|0\rangle_{b'} + \beta|1\rangle_{a'}|1\rangle_{b'}) \quad (|a| \leq |\beta|).$$

As before we would like to extract a maximally entangled state out of this pair. We can write the above state also as

$$|\Psi\rangle^{\otimes 2} = \alpha^2|0\rangle_a|0\rangle_{a'}|0\rangle_b|0\rangle_{b'} + \beta^2|1\rangle_a|1\rangle_{a'}|1\rangle_b|1\rangle_{b'} + \sqrt{2}\alpha\beta\left(\frac{|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|0\rangle_{a'}|1\rangle_b|0\rangle_{b'}}{\sqrt{2}}\right).$$

If now Alice measures the operator $\sigma_T \equiv \sigma_z^a + \sigma_z^{a'}$ on her two particles a, a', then there are three possible results

$$|\Psi\rangle^{\otimes 2} \xrightarrow{\sigma_T=2} |0\rangle_a|0\rangle_{a'}|0\rangle_b|0\rangle_{b'},$$

$$|\Psi\rangle^{\otimes 2} \xrightarrow{\sigma_T=-2} |1\rangle_a|1\rangle_{a'}|1\rangle_b|1\rangle_{b'},$$

$$|\Psi\rangle^{\otimes 2} \xrightarrow{\sigma_T=0} \frac{1}{\sqrt{2}}\left(|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|0\rangle_{a'}|1\rangle_b|0\rangle_{b'}\right),$$

where the last case, which is of interest to us, has the probability

$$\text{prob}(\mu = 0) = 2|\alpha\beta|^2,$$

to occur. If the original state $|\Psi\rangle^{\otimes 2}$ indeed collapse to this last state, then we almost have a purely entangled state. All that is needed is that both Alice and Bob perform local CNOT operations on their two particles, where the primed particles (a', b') are the targets. As a result we get

$$\begin{aligned}
|\Psi\rangle^{\otimes 2} \quad & \xrightarrow{\sigma_T=0} \quad \frac{1}{\sqrt{2}}\left(|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|0\rangle_{a'}|1\rangle_b|0\rangle_{b'}\right) \\
& \xrightarrow{\text{CNOT}^{a,a'}_{b,b'}} \quad \frac{1}{\sqrt{2}}\left(|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|1\rangle_{a'}|1\rangle_b|1\rangle_{b'}\right) \\
& = \quad \frac{1}{\sqrt{2}}|1\rangle_{a'}|1\rangle_{b'}\left(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b\right).
\end{aligned}$$

We now have two particles a, b in an entangled state and two more in the same “up” state.

Now let us examine a more general case, with n pairs

$$\begin{aligned}
|\Psi\rangle^{\otimes n} &= (\alpha|0\rangle_{a_i}|0\rangle_{b_i} + \beta|1\rangle_{a_i}|1\rangle_{b_i})^{\otimes n} \\
&= \alpha^n \prod_{i=1}^{n} |0\rangle_{a_i}|0\rangle_{b_i} + \alpha^{n-1}\beta\sum_{j=1}^{n}\left(|1\rangle_{a_i}|1\rangle_{b_i}\prod_{i\neq j}|0\rangle_{a_j}|0\rangle_{b_j}\right) + \cdots,
\end{aligned}$$

that is we have a tensor product of n pairs numbered $i = 1, \ldots, n$. In analogy to the previous case we now define

$$\sigma_T^a \equiv \sum \sigma_{z_i}.$$

We can group the elements making up the product above, according to the coefficient $\alpha^m\beta^{n-m}$. Clearly (use the binomial expansion) the coefficient $\alpha^m\beta^{n-m}$ appears $\binom{n}{m} = \frac{n!}{m!(n-m)!}$. When Alice measures σ_T^a she will therefore get result $m - (n - m) = 2m - n$ with a probability of $\binom{n}{m}|\alpha^m\beta^{n-m}|^2$ (similar to the factor of $2|\alpha\beta|^2$ we had for $n = 2$ above)

$$\text{prob}(\sigma_T^a = 2m - n) = \binom{n}{m}\left|\alpha^m\beta^{(n-m)}\right|^2 = \frac{n!}{m!(n-m)!}\left|\alpha^{2m}\beta^{2(n-m)}\right|.$$

If we now examine $n \to \infty$ the probability will be maximal, and approach a delta function at $m = |\alpha|^2 n + O(\sqrt{n})$.[27] Thus for large n we may examine only the case of $m = |\alpha|^2 n$. For a given m all the elements we add are all orthogonal to one another and are each symmetric

[27] This is true regardless of the values of α, β (as long as none of them is zero).

in Alice and Bob's particles, therefore we can make a change of base[28] so that that the system have the form (not normalized) for a given m

$$|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B + \cdots + |\binom{n}{m}\rangle_A|\binom{n}{m}\rangle_B,$$

where A is a new "particle" with $\binom{n}{m}$ states which replace all the a_i particles ($i = 1,\dots,n$) which were originally held by Alice, and similarly for B which replaces Bob's b_i's. Now that for a given m all the A,B pairs are symmetric, let us view the case of k identical, maximally entangled states. This situation will be described as (up to normalization)

$$(|0\rangle_{a_i}|0\rangle_{b_i} + |1\rangle_{a_i}|1\rangle_{b_i})^{\otimes k}.$$

Doing the product we will get 2^k elements which are all orthogonal to each other and with the same coefficient (the case of $\alpha = \beta$ above). Therefore if we have above $\binom{n}{m}$ orthogonal elements each with the same coefficient, then we can deduce that this is equivalent to nH entangled pairs, where we define the function H such that

$$2^{nH} \equiv \binom{n}{m}.$$

We are interested in the most likely case of m, which is $m = |\alpha|^2 n$ and we therefore have

$$\begin{aligned} nH(m = |\alpha|^2 n) &= \log_2 \binom{n}{|\alpha|^2 n} \\ &= \log_2 \left(\frac{n!}{(|\alpha|^2 n)!\,(n - |\alpha|^2 n)!} \right) \\ &= \log_2 \left(\frac{n!}{(|\alpha|^2 n)!\,(|\beta|^2 n)!} \right), \end{aligned}$$

where in the last equality we used the fact that $(|\alpha|^2 + |\beta|^2 = 1)$. Using the Stirling's formula

$$\log n! \approx \frac{1}{2}\log(2\pi n) + n\log n - n\log e \approx n\log n,$$

we get here

$$\begin{aligned} nH(m = |\alpha|^2 n) &\approx n\log n - n|\alpha|^2 \log_2(n|\alpha|^2) - n|\beta|^2 \log_2(n|\beta|^2) \\ &= n\left[\log_2 n - (|\alpha|^2 + |\beta|^2)\log_2 n - |\alpha|^2 \log_2 |\alpha|^2 - |\beta|^2 \log_2 |\beta|^2\right] \\ &= -n\left[|\alpha|^2 \log_2 |\alpha|^2 + |\beta|^2 \log_2 |\beta|^2\right]. \end{aligned}$$

If we define

$$\begin{gathered} p \equiv |\alpha|^2 \\ \Rightarrow (1-p) = |\beta|^2, \end{gathered}$$

Then we may write

$$H = -\left[p\log_2 p + (1-p)\log_2 p\right].$$

To conclude we saw that if we use the scheme of measuring σ_T^a (the sum of spins in the z direction of Alice's particle), then on the average we will get out of initially n non-maximally entangled states, nH maximally entangled states, where

$$\begin{aligned} nH &= -n\left[|\alpha|^2 \log_2 |\alpha|^2 + |\beta|^2 \log_2 |\beta|^2\right] \\ &= -n\left[p\log_2 p + (1-p)\log_2 p\right]. \end{aligned}$$

[28] It is enough to make a change of names. We simply number all the permutations of m elements out of n and then call the jth permutation $|j\rangle_A$. We do the same for $|j\rangle_B$, and we automatically get

$$|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B + \cdots + |\binom{n}{m}\rangle_A|\binom{n}{m}\rangle_B.$$

Or simply

$$n \text{ non-maxiamlly entangled} \longrightarrow nH \text{ maxiamlly entangled}.$$

The ratio of maximally entangled pairs, out of the original number of pairs is

$$E(\psi) \equiv \frac{\text{maxiamlly entangled pairs}}{\text{non-maxiamlly entangled pairs}} = H = -\left[|\alpha|^2 \log_2 |\alpha|^2 + |\beta|^2 \log_2 |\beta|^2\right] + O\left(\frac{1}{\sqrt{n}}\right),$$

$$\psi \equiv \alpha|0\rangle + \beta|1\rangle \quad \text{(single non-maximally entangled particle)}$$

We can therefore give $E(\psi)$ the meaning of *measure of entanglement* of a single pair of particles (because on average we can extract $H < 1$ pairs of maximally entangled pairs).

The result we found here may be generalized further.[29] If $|\Psi\rangle_{AB}$ has the *Schmidt decomposition*

$$|\Psi\rangle_{AB} = \sum_{k=1}^{n} \sqrt{p_k}|k\rangle_A|k\rangle_B,$$

then one gets the *Shannon entropy*

$$H(\Psi) \equiv -\sum_k p_k \log_2 p_k \quad \text{(Shannon entropy)}.$$

We then say that

$$E(\Psi) \equiv H(\Psi) = -\sum_k p_k \log_2 p_k$$

is the entanglement associated with $|\Psi\rangle_{AB}$. If two systems have no correlations between them then we simply add their entanglement

$$E(\psi_1 \otimes \psi_2) = E(\psi_1) + E(\psi_2).$$

This is true, since the Schmidt decomposition in such a case is

$$|\psi_1 \otimes \psi_2\rangle = \sum_i \sqrt{p_i}|i\rangle_{A_1}|i\rangle_{B_1} \sum_j \sqrt{q_j}|j\rangle_{A_2}|j\rangle_{B_2} = \sum_{i,j} \sqrt{p_i q_j}|i\rangle_{A_1}|i\rangle_{B_1}|j\rangle_{A_2}|j\rangle_{B_2},$$

where the last element is also in a Schmidt decomposition form. Using the formula for the Shannon entropy we get

$$\begin{aligned} H(\psi_1 \otimes \psi_2) &= -\sum_{i,j} p_i q_j \log_2(p_i q_j) = -\sum_{i,j} p_i q_j \left(\log_2 p_i + \log_2 q_j\right) \\ &= -\sum_i p_i \sum_j q_j \log_2 q_j - \sum_j p_j \sum_i q_i \log_2 p_i = -\sum_j q_j \log_2 q_j - \sum_i q_i \log_2 p_i \\ &= H(\psi_1) + H(\psi_2). \end{aligned}$$

Since the Shannon entropy may be added then so can the entanglement.

Note, that the entanglement measure we defined is a good measure in the sense that it does not depend on the base we choose *locally*. If we make a local unitary transformation of the form $U_A \otimes U_B$ (unlike a unitary transformation U_{AB} which may be non-local), then the specific orthonormal basis vectors we use in the Schmidt decomposition will change, but the Schmidt coefficients will not (a unitary transformation, transforms an orthonormal basis to an orthonormal basis).

Note also that although we can use POVM's to distinguish between non-orthogonal states, with a better chance than regular measurements, we cannot use it to increase entanglement of the system (on average).

As we shall see, the quantity H has the traits of classical entropy. It is called the *Shanon entropy*.

[29] See also Von Neumann entropy S and entanglement measure.

CHAPTER 6

Quantum information

6.1. Data compression (classical)

Assume that we have a very *long* message of n letters, written in an alphabet of k letters. As in any language some letters appear more often then others. We can therefore describe the language by the probability of each letter to appear (we assume that the probability of appearance is independent of the letter/letters before or after it). The set of letters and probabilities we denote as X_k

$$X_k = \{a_x, p_x\}_{x=1}^{k} \quad (\sum_{x=1}^{k} p_x = 1).$$

This is actually equivalent to a density matrix of states.

We would now like to compress our message before sending it, i.e. send less letters/bits which will convey the same message. Since the message is long, then in a *typical* message of length n, the a_x will appear $p_x n$ times. The number of possible ways to order the letters of a *typical* message are therefore[1]

$$\frac{n!}{\prod_x (np_x)!}.$$

Using the definition of the *Shannon entropy*

$$H \equiv -\sum_x p_x \log_2 p_x$$

we can therefore write (using the Stirling approximation)[2]

$$\#\text{typical messsages} \approx \frac{n!}{\prod_x (np_x)!} \approx 2^{nH}.$$

Thus to encode the the different *typical* messages, we can simply number them $1, 2, \ldots, 2^{nH}$ and then send this number instead. The number of bits we need in order to encode all these number is nH. We therefore say that we can encode a message of n letters using an alphabet of k letters, using just nH bits

$$\begin{array}{c} n \text{ letters} \\ k \text{ letter alphabet} \end{array} \xrightarrow{\text{compression}} nH \text{ bits} \quad (H = -\sum_x p_x \log_2 p_x).$$

This of course holds only for the typical messages, whose weight in the overall ensemble of messages increases as $n \to \infty$.

Another way of reaching the same conclusion, is to examine a single message of length n

$$\text{message} = (x_1, x_2, \ldots, x_n)$$

where in the ith position the letter a_{x_i} appears. The probability of such a message occurring is

$$\text{prob(message)} = \text{prob}(x_1, x_2, \ldots, x_n) = p_{x_1} p_{x_2} \cdots p_{x_n}$$

[1] We use here the same logic as was used above, in determining a measure for entanglement.

[2] Recall that

$$\log(n!) \approx n \log n.$$

or

$$\log_2\left[\text{prob}(x_1,x_2,\ldots,x_n)\right] = \log_2(p_{x_1}p_{x_2}\cdots p_{x_n}) = \sum_i \log_2 p_{x_i}.$$

By the central limit theorem, for $n \to \infty$ we have

$$-\frac{1}{n}\log_2\left[\text{prob}(x_1,x_2,\ldots,x_n)\right] \sim -\langle \log_2 p\rangle \equiv H,$$

where the average on the right is with respect to the probability distribution defined by the p_i's. We thus see that the probability of a typical message to occur is 2^{-nH}. As we saw the number of typical messages is 2^{nH}, and thus we see that the set of all the typical messages occurs with a probability very close to one, so that the case of other messages may be neglected.

More rigorously (without proof), we may write that for any $\varepsilon, \delta > 0$ there exists $n_{\varepsilon,\delta}$ sufficiently large such that for any $n > n_{\varepsilon,\delta}$ the following is true: There is a set of "typical" messages (out of all possible sequences of length n) with a total probability greater than $1-\varepsilon$ to occur, such that each "typical" message has a probability P to occur,[3] which obeys

$$2^{-n(H+\delta)} \leq P \leq 2^{-n(H-\delta)}.$$

Since the total probability of all the typical messages to occur is greater then $1-\varepsilon$, then we can put a bound on the number $N_{\varepsilon,\delta}$ of "typical" messages[4]

$$(1-\varepsilon)2^{n(H-\delta)} \leq N_{\varepsilon,\delta} \leq 2^{n(H-\delta)}.$$

Thus we see that in the limit of $n \to \infty$ only 2^{nH} of the 2^n possible messages will occur, and therefore we can use nH bits to encode these "typical" messages.

To conclude we see that H gives a measure of uncertainty of letters in the message. If $H = 0$ then only one letter appears in the message, and is therefore predetermined. If however,[5] $H = \log_2 n$ then all letters are equally likely to appear and we cannot compress our message. We can also say that H is the information that each letter carries. If we again look at the case of $H = 0$ then all letters are identical and the addition of a new one does not give us new information, if on the other hand we have $H = \log_2 n$, then each added letter gives us new information about the message which requires an extra $\log_2 n$ bits to encode it.

6.2. Data compression (Quantum)

We would now like to do the equivalent of classical data compression in the quantum case. In the quantum case the letters will be replace by pure quantum states, so the ensemble describing the "language" is now replaced by a density matrix. The difference between the classical case and the quantum case arises when the density matrix is constructed of non-orthogonal states ($\rho = \sum p_i|\psi_i\rangle\langle\psi_i|$ where the $|\psi_i\rangle$ are not necessarily orthogonal)[6]. In such a case the different states (which were the letters before) cannot be distinguished between.

[3] Each typical message may have a slightly different probability to occur, but all these different probabilities obey the inequality given.

[4] The bound comes from

$$NP_{\max} > 1-\varepsilon$$
$$NP_{\min} < 1$$

[5] When all letters are likely to appear then

$$H = -\sum_{i=1}^{n}\frac{1}{n}\log_2\frac{1}{n} = -\log_2\frac{1}{n} = \log_2 n.$$

[6] We can of course diagonalize the density matrix and get orthonormal states. However, the real physics (for some reason) is that our source emits non-orthogonal states with different probabilities.

The measure we shall use to determine how good our compression is, will be the *fidelity* F. If our original message is $|\varphi_i\rangle$ and we send instead a message which after decompression is $|\varphi_j\rangle$ then the fidelity is defined by how close the two state vectors are

$$F = |\langle\varphi_i|\varphi_f\rangle|^2.$$

For a random (original) state/message described by a density matrix ρ, coded as $|\varphi\rangle$ the fidelity is defined as

$$F \equiv \langle\varphi|\rho|\varphi\rangle = \mathrm{Tr}(|\varphi\rangle\langle\varphi|\rho),$$

which for a pure state degenerates to the first definition. If both the original message and the decompressed message have different probabilities of occurring then we take the fidelity as the average (weighted by the probabilities).

Let us start with an example, assume a density matrix

$$\rho = \frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x|,$$

we would like to find a state $|\varphi\rangle$ with a maximum fidelity for this density matrix. If we diagonalize the matrix we get

$$\rho = \cos^2\frac{\pi}{8}|\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}| + \sin^2\frac{\pi}{8}|\downarrow_{\hat{n}}\rangle\langle\downarrow_{\hat{n}}|,$$

where

$$\hat{n} = \frac{\hat{x}+\hat{z}}{\sqrt{2}}$$

and

$$|\uparrow_{\hat{n}}\rangle = \cos\frac{\pi}{8}|\uparrow_z\rangle + \sin\frac{\pi}{8}|\downarrow_z\rangle,$$
$$|\downarrow_{\hat{n}}\rangle = \sin\frac{\pi}{8}|\uparrow_z\rangle - \cos\frac{\pi}{8}|\downarrow_z\rangle.$$

It can easily be shown that the maximum fidelity is reached when

$$|\varphi\rangle = |\uparrow_{\hat{n}}\rangle$$

which gives

$$F = \langle\varphi|\rho|\varphi\rangle = \cos^2\frac{\pi}{8} = 0.853\ldots.$$

Now let us assume that Alice has a message made of three particles emitted from a source with the same density matrix ρ as above ($\rho = \frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x|$). Alice wants to send the message to Bob but wants may send only 2 of the particles (qubits) to him. We might think that the best she could do is simply send

$$|\varphi\rangle = \rho\otimes\rho\otimes|\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}|,$$

where the density matrices ρ stand for the original particles and $|\uparrow_{\hat{n}}\rangle$ is the same state we used above for maximum fidelity. Since the first two particles, which Alice sent, are the original ones, then their fidelity will be 1, thus the fidelity will change only because of the last particle and we'll get

$$F = 1\cdot 1\cdot\cos^2\frac{\pi}{8} = 0.853\ldots.$$

However, Alice can increase the fidelity of her message. Since she is sending only two qubits and determines the third then the Hilbert space described by her new state belongs to a subspace of the original Hilbert space. We would like to project the original three qubits onto a subspace which is more probable and thus gives us the maximum fidelity (see the following). If we change our basis to the one in which ρ is diagonal, then for any state $|\psi\rangle$ of the three particles (due to the symmetry of the density matrix with respect to $|\uparrow_z\rangle, |\uparrow_x\rangle$) we get

$$|\langle\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = \cos^6\frac{\pi}{8} = 0.62$$
$$|\langle\uparrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = |\langle\uparrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = |\langle\downarrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = \cos^4\frac{\pi}{8}\sin^2\frac{\pi}{8} = 0.107$$

$$|\langle\uparrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = |\langle\downarrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = |\langle\downarrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = \cos^2\frac{\pi}{8}\sin^4\frac{\pi}{8} = 0.018$$

$$|\langle\downarrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = \sin^6\pi = 0.003$$

The dimension of the Hilbert subspace we can encode with just 2 qubits is 4, thus we are looking for a subspace spanned by 4 of the above combination which will occur in the highest probability. This, condition will be fulfilled by taking the Hilbert subspace $\mathcal{H}_1$ spanned by the first 4 states

$$\mathcal{H}_1 = \mathrm{span}\{|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\uparrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\uparrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle\} \subset \mathcal{H}.$$

If we also define

$$\mathcal{H}_2 = \mathrm{span}\{|\downarrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle, |\uparrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}\rangle, |\downarrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}\rangle\} \subset \mathcal{H},$$

then we have[7]

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2.$$

The procedure we shall use is the following. Alice performs a unitary operation U on the 3-particle state emitted from her sources. The unitary operator is such that

$$U : \mathcal{H}_1 \to |\cdot\rangle|\cdot\rangle|0\rangle,$$

and

$$U : \mathcal{H}_2 \to |\cdot\rangle|\cdot\rangle|1\rangle.$$

Having done that Alice measures the last particle and projects it onto either $\mathcal{H}_1$ with probability of $p_1 = 0.62 + 3\cdot 0.107 = 0.94$ or onto $\mathcal{H}_2$ with probability of $p_2 = 0.003 + 3\cdot 0.018 = 0.06$. If Alice measures 0 the she sends the first two qubits to Bob, Bob adds a third qubit in state $|0\rangle$ and then performs U^{-1} (the same U that Alice used) to decode the message. If however, Alice measured 0, then best she can do is to send the two qubits in a predetermined state such that after Bob decodes (by the same method as before) he gets $|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle$, which is the most probable single state. If we denote by Π_i the projection on subspace $\mathcal{H}_i$ then the messages Bob decodes are

$$\frac{\Pi_1|\psi\rangle}{\sqrt{\langle\psi|\Pi_1|\psi\rangle}} \subset \mathcal{H}_1 \quad (\text{with probability } p_1 = \langle\psi|\Pi_1|\psi\rangle = 0.94),$$

$$|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle \subset \mathcal{H}_1 \quad (\text{with probability } p_2 = \langle\psi|\Pi_2|\psi\rangle = 0.06),$$

where the denominator in the first equation is merely for normalization purposes. The fidelity of Alice's message can now finally be found, by comparing the message Bob decoded and the original one. Since, however, Alice may send different messages, depending on the result of her measurement, then we regard the average fidelity[8]

$$F = 0.94\left|\langle\psi|\frac{\Pi_1|\psi\rangle}{\sqrt{\langle\psi|\Pi_1|\psi\rangle}}\right|^2 + 0.06\,|\langle\psi|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle|^2 = 0.94^2 + 0.06\cdot 0.62 = 0.92.$$

We see, that by ruining the states of all three particles instead of just one, we got a higher fidelity.

[7] Note that we use addition ($\oplus$) of spaces, and not a tensor product. This is because we are dealing with subspaces.

[8] Again note, that due to the special symmetry between the two possible states of the original system $|\uparrow_x\rangle$, $|\uparrow_x\rangle$, we do not treat here the cases differently, however for a general case we would have to.

6.3. Schumacher's noiseless encoding

Having solved the above example let us now generalize it. We would like to have an analogy of Shannon's theorem for the quantum case. Clearly, if our source emits states which are mutually orthogonal, then we can distinguish between them and we can therefore use Shannon's classical theorem for compressing the information. The problem arises when the emitted states are not all mutually orthogonal.

Assume a source of states $|\psi_i\rangle$ (not necessarily all orthogonal to each other, $i = 1, \dots, \tilde{N}$) described by the density matrix

$$\rho = \sum_{i=1}^{\tilde{N}} p_i |\psi_i\rangle\langle\psi_i|.$$

A message of n (uncorrelated) letters will therefore be described by the density matrix ρ_n

$$\rho_n = \underbrace{\rho \otimes \rho \cdots \otimes \rho}_{n} = \rho^{\otimes n}.$$

Similarly to the subset of "typical" messages we had in the classical case, we shall see that here we have a probable/likely subspace of the Hilbert space (for large enough n). To see this we diagonalize our density matrix ρ

$$\rho = \sum_{k=1}^{N} \lambda_k |k\rangle\langle k| \quad (\langle k|k'\rangle = \delta_{k,k'}).$$

Once we do this, we are back to the the classical theorem of Shannon (since the states $|k\rangle$ are all mutually orthonormal and therefore distinguishable). We now have an "alphabet" of N letters each with probability λ_k of appearing. Using Shannon theorem we can compress a message of n such letters to a message of nH ($H = \sum \lambda_k \log_2 \lambda_k$) bits or nH qubits. We now define the *Von Neumann entropy S* as

$$S(\rho) \equiv -\operatorname{Tr}(\rho \log_2 \rho),$$

which is most easily calculated (and actually thus defined) when ρ is diagonal. In this case we get

$$S = -\sum_{k=1}^{N} \lambda_k \log_2 \lambda_k,$$

which is just the Shannon entropy for the diagonalized form of the density matrix (but not of the original form, for which we would have used the p_i's). Thus we can say that the dimension of the "likely" or "probable" Hilbert subspace is

$$\dim \mathcal{H}_{\text{prob}} = 2^{nS(\rho)}.$$

As a consequence Alice can compress her message of n particles/states into nS qubits. Bob receiving the message can then decompress it and find the original n state message. Note, however, that unless the possible states are all mutually orthogonal then Bob cannot know for certain what message he has (although he knows, that it is the same as Alice sent). By Holevo's theorem (see earlier), he can extract only 1 (classical) bit out of every qubit.

Before continuing, it is worth to note the difference between the Von Neumann entropy S and the Shannon entropy H. Given the density matrix above

$$\rho = \sum p_i |\psi_i\rangle\langle\psi_i|,$$

the Shannon entropy treats the different states $|\psi_i\rangle$ as distinguishable even though they are not necessarily mutually orthogonal, and thus

$$H(\rho) = -\sum_{i=1}^{\tilde{N}} p_i \log_2 p_i.$$

On the other hand the Von Neumann entropy is found by first diagonalizing the density matrix which gives

$$S = -\sum_{k=1}^{N} \lambda_k \log_2 \lambda_k.$$

The two definitions coincide when the states $|\psi_i\rangle$ are mutually orthogonal, but do not coincide otherwise. further more the Shannon entropy depends on the way the density matrix was constructed (which states are actually emitted by the sources) and not only on the density matrix itself.

6.3.0.1. *Measure of entanglement.* As we saw before the Shannon entropy H gave us measure for the entanglement E (if we have n non-maximally entangled pairs, then we can can distill from them nE maximally entangled pairs). We found that after writing the state in the Schmidt decomposition form

$$|\Psi\rangle_{AB} = \sum_k \sqrt{p_k}|k\rangle_A|k\rangle_B \quad \text{(Schmidt decomposition)},$$

the measure of entanglement was

$$E(\Psi) \equiv H(\Psi) = -\sum_k p_k \log_2 p_k.$$

Using the Schmidt decomposition we can write the density matrix of the two particles as

$$\rho_{AB} = \sum_k p_k|k\rangle_A|k\rangle_{BB}\langle k|_A\langle k|.$$

We see, that in this case the entanglement is simply the trace of $\rho \log_2 \rho$. Thus we can say that the entanglement of two particles/regions is

$$E = S(\rho_{AB}),$$

where again we use the Von Neumann entropy since it doesn't depend on the basis we use, while the Shannon entropy does. Further more if we take a partial trace of ρ_{AB} over A or B we get

$$\rho_A \equiv \mathrm{Tr}_B\, \rho_{AB} = \sum_{k=1}^{n} p_k|k\rangle_{AA}\langle k|,$$

$$\rho_B \equiv \mathrm{Tr}_A\, \rho_{AB} = \sum_{k=1}^{n} p_k|k\rangle_{BB}\langle k|.$$

We see that the coefficients p_k have not changed from the original ρ_{AB} and thus we can also write that

$$E = S(\rho_{AB}) = S(\rho_A) = S(\rho_B) \quad \text{(measure of entanglement)}.$$

6.3.1. dilution. We have so far discussed only the problem distillation: turning n non-maximally entangled states to nS maximally entangled states. The reverse, *dilution*, is also possible: turning nS maximally entangled states to n non-maximally entangled states. The protocol is very simple. Alice starts with n local pairs in the non-maximal entangled state. To create the non-maximal entangled pairs between her and Bob she must now teleport n particles to him. However, Alice and Bob have only nS EPR pairs between them. To over come this, Alice compresses the n pairs to nS pairs and then teleports (using the nS EPR pairs) one particle from each pair to Bob. Bob and Alice then decompress their particles and finally get the n non-maximally entangled states they wanted.

Note that this protocol, holds only for very large n, since only then the compression will have the efficiency nS.

6.4. Communication with noise (classical)

Assume that Alice wants to send a message to Bob, using an alphabet $X = \{x, p_x)_{x=1}^N$ (total of N letters and letter a_x appears with probability p_x), while Bob uses an alphabet $Y = \{y, q_y\}_{y=1}^N$ (note that the two alphabets have the same number of letters). The problem is that there is noise in the communication channel between them, and thus a letter sent by Alice may change with different probabilities to different letters which Bob receives. We denote the probability that receives the letter y if Alice sent x as $p(y|x)$

$$x \xrightarrow{p(y|x)} y.$$

Now, assuming that Bob knows Alice's p_x and knows the noise behavior $p(y|x)$, what can he deduce from the message he receives?

We denote the probability that Alice sent x and Bob received y as $p(x,y)$. By the above definitions we get

$$p(x,y) = p(y|x)p_x.$$

Similarly we also have (note the exchange of x,y in the last probability)

$$p(x,y) = p_y p(x|y).$$

We further assume that we know the Shannon entropy of x and y

$$H(X) = -\sum p_x \log_2 p_x = -\langle \log_2 p_x \rangle_{p_x},$$

$$H(Y) = -\sum p_y \log_2 p_y = -\langle \log_2 p_y \rangle_{p_y}.$$

We Similarly define the *total entropy* $H(X,Y)$ as

$$H(X,Y) \equiv -\sum_{x,y} p(x,y) \log_2 p(x,y) = -\langle \log_2 p(x,y) \rangle_{p(x,y)},$$

and the *conditional entropy* $H(X|Y)$ as

$$H(X|Y) \equiv -\sum_{x,y} p(x,y) \log_2 p(x|y) = -\langle \log_2 p(x|y) \rangle_{p(x,y)}.$$

By the definition $p(x,y) = p_y p(x|y)$, the last definition can also be written as

$$H(X|Y) = -\langle \log_2 p(x,y) \rangle_{p(x,y)} + \langle \log_2 p(y) \rangle_{p(x,y)} = H(X,Y) - H(Y),$$

and similarly

$$H(Y|X) = H(X,Y) - H(X).$$

Note, that by definition we have

$$H(X|Y), H(Y|X) \geq 0.$$

The meaning of the conditional entropy is that it tells us how much information needs to be sent to Bob in order to convey a message, if he *already knows* the sequence y. If Bob knows that he got a letter y then the probability of it coming from a letter x is $p(x|y)$. Therefore, as far as Bob is concerned, Alice does not use the alphabet $\{x, p_x\}$ but rather the alphabet $\{x, p(x|y)\}$, and therefore in order to convey the message (using Shannon's theorem) it suffices to send him $H(X|Y)$ bits per letter (instead of $H(X)$ bits, when he doesn't know the y's).[9]

We can now define the *mutual information* $I(X;Y)$

$$I(X;Y) \equiv H(X) - H(X|Y).$$

This quantity tells us how correlated the x's and y's are. It tells us how many bits per letter x I have save from sending if I know y. If for example x,y are completely on correlated, then having learned y doesn't help me at all and $H(X|Y) = H(X)$ which will give us $I = 0$.

[9] ??? Note, that we are actually talking about an average, since $p(x|y)$ is different for every different letter y Bob has in his sequence. ???

On the other hand if they are completely correlated (one-to-one) then having learned y I need no more information. In this case $H(X|Y)=0$ and $I=H(X)$.

Note, that the mutual information I is symmetric:

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

6.5. Accessible Information

We now turn to a quantum case. Assume that Alice has source of states which emits particles in state $|\psi_i\rangle$ with probability p_i

$$\rho=\sum p_i|\psi_i\rangle\langle\psi_i|.$$

Now, Bob wants to determine which state has been emitted. For this he may choose any POVM set $\{F_y\}$. The probability that Bob measure y of the particle is in a given state $|\psi_x\rangle$ is given by

$$p(y|x)=\langle\psi_x|F_y|\psi_x\rangle.$$

We define the amount of information Bob can deduce from ρ as the *accessible information* $\mathrm{Acc}(\rho)$

$$\mathrm{Acc}(\rho)\equiv\max_{\{F_y\}}I(X;Y).$$

If the states $|\psi_i\rangle$ are all mutually orthogonal then they are distinguishable (using $F_y=|\psi_y\rangle\langle\psi_y|$) and we are back to the classical case

$$\mathrm{Acc}(\rho)=H(X).$$

If however, the states are not all mutually orthogonal, then there is no general formula but it can be proven that

$$\mathrm{Acc}(\rho)\le S(\rho),$$

where an equality is reached only for very long messages ($n\to\infty$).

6.6. Decoherence and the measurement problem

We call a pure state, a coherent one. We shall see that once the state interacts with an environment, then the reduced density of the state (without the environment) becomes non-pure. This process is called *decoherence* or *dephasing*.

As an example of decoherence, assume a pure state

$$|\psi\rangle=|0\rangle+e^{i\alpha}|1\rangle,$$

which is described by the density matrix

$$\rho=\rho^2=\begin{pmatrix} 1 & e^{i\alpha} \\ e^{-i\alpha} & 1 \end{pmatrix}.$$

We now add an environment to the system which is in an initial state $|\tilde{e}\rangle$

$$|\Psi\rangle_{\text{tot}}=|\psi\rangle|\tilde{e}\rangle=\frac{1}{\sqrt{2}}\left(|0\rangle+e^{i\alpha}|1\rangle\right)|\tilde{e}\rangle.$$

We further assume that the interaction of the system and the environment is very weak and we get after some time of interaction

$$|\Psi\rangle_{\text{tot}}\to\frac{1}{\sqrt{2}}\left(|0\rangle|e_0\rangle+e^{i\alpha}|1\rangle|e_1\rangle\right),$$

where $|e_0\rangle,|e_1\rangle$ are some states of the environment, not necessarily orthogonal. The density matrix of the system alone (the reduced matrix after tracing over the environment) is now

$$\rho\to\frac{1}{2}\begin{pmatrix} \langle e_0|e_0\rangle & e^{i\alpha}\langle e_0|e_1\rangle \\ e^{-i\alpha}\langle e_1|e_0\rangle & \langle e_1|e_1\rangle \end{pmatrix}=\frac{1}{2}\begin{pmatrix} 1 & e^{i\alpha}\langle e_0|e_1\rangle \\ e^{-i\alpha}\langle e_1|e_0\rangle & 1 \end{pmatrix}.$$

Except for very special cases that $|e_0\rangle$ and $|e_1\rangle$ differ only by a phase, the new density matrix is no longer a pure one.

6.6.1. density matrices and entanglement. Let us now examine how this effects entanglement. This time we shall start with a system plus environment in a state

$$|\Psi\rangle_{\text{tot}} = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_A|\uparrow\rangle_B + |\downarrow\rangle_A|\downarrow\rangle_B\right)|\uparrow\rangle_E.$$

We now activate an interaction U_{AE} between particles A and E such that

$$U_{AE}|\Psi\rangle_{\text{tot}} = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_A|\uparrow\rangle_B|\uparrow\rangle_E + |\downarrow\rangle_A|\downarrow\rangle_B|\downarrow\rangle_E\right).$$

We can write this state as a density matrix $\rho_{ABE} = U_{AE}|\Psi\rangle\langle\Psi|U_{AE}^\dagger$. Taking the partial trace over the environment E we get

$$\rho_{AB} = \frac{1}{2}\left(|\uparrow\rangle_{AA}\langle\uparrow|\otimes|\uparrow\rangle_{BB}\langle\uparrow| + |\downarrow\rangle_{AA}\langle\downarrow|\otimes|\downarrow\rangle_{BB}\langle\downarrow|\right).$$

This looks like an entangled state, but is it? If we look at the entangled state

$$\frac{1}{\sqrt{2}}\left(|\uparrow\rangle_A|\uparrow\rangle_B + |\downarrow\rangle_A|\downarrow\rangle_B\right),$$

and write its density matrix, we will get a different result than the above (there will appear mixed elements with both "up" and "down" states).

The criteria for entanglement in density matrices, is slightly different than the one for pure states. Here we say that a density matrix is *entangled* if we *cannot* write it as a sum of product density states. That is

$$\rho_{AB} \neq \sum p_i \rho_A \otimes \rho_B \Rightarrow \text{entangled}.$$

6.6.2. The measurement problem. We saw that interaction with the environment leads to decoherence, and the behavior of a system as if it were described by a density matrix. This seems to explain collapse, but it does not, since first of all it does not explain why the collapse is to a certain state, and it doesn't solve the problem that macroscopically large systems may be in superposition - the system plus the environment are still in a superposition (Schrodinger's cat, both alive and dead).

6.7. Error correction - Shor's algorithm

Assume that we want to send a classical bit over a noisy channel. If we simply send one bit (say 0) it might be corrupted by the noise and the bit received (say 1) will be different than the one sent. When dealing with classical bits it is relatively simple to solve the problem (when the noise is weak). We simply duplicate the bit two extra times and send three identical instead of just one

$$\tilde{0} \equiv 000,$$

$$\tilde{1} \equiv 111.$$

Assuming the noise to be weak, at most one bit of the three will be corrupted, we can then correct the error by using the majority rule method (if one bit differs from the other two it is changed to agree with the two).

We now turn to the quantum case. The problem here is two fold. First, due to the no cloning theorem, we cannot duplicate our qubits; second, if we make measurmentmeasurement to determine what has changed we collapse our state and change it.

Before solving the problem, let us first see what type of errors might occur. We start with a general qubit and an environment

$$|\Psi\rangle_{\text{tot}} = (\alpha|0\rangle + \beta|1\rangle)|\text{Env.}\rangle.$$

The most general unitary operator which couples the environment and the qubit but does not entangle them may be written as

$$U = e^{i\theta_{\text{Env.}}\hat{n}_{\text{Env.}}\vec{\sigma}_{\text{sys}}} = \mathbb{1} + \varepsilon_1\sigma_x + \varepsilon_2\sigma_y + \varepsilon_3\sigma_z,$$

where the ε_i are some constants and the Pauli matrices on the right operate on the qubit. We can write the effect of each element in the sum

$$\begin{array}{ccc} |0\rangle & \xrightarrow{\mathbb{1}} & |0\rangle \\ |1\rangle & & |1\rangle \end{array},$$

$$\begin{array}{ccc} |0\rangle & \xrightarrow{\sigma_x} & |1\rangle \\ |1\rangle & & |0\rangle \end{array} \quad (\text{bit flip}),$$

$$\begin{array}{ccc} |0\rangle & \xrightarrow{\sigma_y} & i|1\rangle \\ |1\rangle & & -i|0\rangle \end{array},$$

$$\begin{array}{ccc} |0\rangle & \xrightarrow{\sigma_z} & |0\rangle \\ |1\rangle & & -|1\rangle \end{array} \quad (\text{phase flip}).$$

We see that there are basically two errors we should treat, the *bit flip* and the *phase flip* (the effect of σ_y can be reproduced by their combination and an extra global phase).

Let us start by treating the bit flip. Although we cannot duplicate qubits, we can use a CNOT (actually two) which will give a similar effect. We add to our qubit two more qubits in a known "up" state, and perform a unitary operator U, which is actually a CNOT of the original qubit with each of the two new ones

$$U\left[\frac{1}{\sqrt{2}}\left(\alpha|0\rangle+\beta|1\rangle\right)|0\rangle|0\rangle\right] = \frac{1}{\sqrt{2}}\left(\alpha|0\rangle|0\rangle|0\rangle+\beta|1\rangle|1\rangle|1\rangle\right).$$

Now, assume that a bit flip occurs in one of the three qubits

$$\frac{1}{\sqrt{2}}\left(\alpha|0\rangle|0\rangle|0\rangle+\beta|1\rangle|1\rangle|1\rangle\right) \xrightarrow{\text{bit flip}} \text{or}\begin{cases} \frac{1}{\sqrt{2}}\left(\alpha|1\rangle|0\rangle|0\rangle+\beta|0\rangle|1\rangle|1\rangle\right) \\ \frac{1}{\sqrt{2}}\left(\alpha|0\rangle|1\rangle|0\rangle+\beta|1\rangle|0\rangle|1\rangle\right) \\ \frac{1}{\sqrt{2}}\left(\alpha|0\rangle|0\rangle|1\rangle+\beta|1\rangle|1\rangle|0\rangle\right) \end{cases}.$$

If we make a measurement we will cause a collapse of the wave function, unless the wave function is already an eigenvector. The possible state are all eigenvalues of the operators $\sigma_z^1\sigma_z^2$ and $\sigma_z^2\sigma_z^3$, but the values measured are different according to which bit has flipped:

$\sigma_z^1\sigma_z^2$	$\sigma_z^2\sigma_z^3$	flipped bit
1	1	non
-1	1	1
1	-1	3
-1	-1	2

Let us now generalize the above procedure to take care of all possible errors. Shor suggested the use of 9 qubits to protect a single one. He suggested to use a unitary operation such that

$$\uparrow \to \tilde{\uparrow} = \frac{1}{2\sqrt{2}}\left(\uparrow\uparrow\uparrow+\downarrow\downarrow\downarrow\right)\left(\uparrow\uparrow\uparrow+\downarrow\downarrow\downarrow\right)\left(\uparrow\uparrow\uparrow+\downarrow\downarrow\downarrow\right),$$

$$\downarrow \to \tilde{\downarrow} = \frac{1}{2\sqrt{2}}\left(\uparrow\uparrow\uparrow-\downarrow\downarrow\downarrow\right)\left(\uparrow\uparrow\uparrow-\downarrow\downarrow\downarrow\right)\left(\uparrow\uparrow\uparrow-\downarrow\downarrow\downarrow\right).$$

If we define

$$|0\rangle \equiv \frac{1}{\sqrt{2}}\left(\uparrow\uparrow\uparrow+\downarrow\downarrow\downarrow\right),$$

$$|1\rangle \equiv \frac{1}{\sqrt{2}}\left(\uparrow\uparrow\uparrow-\downarrow\downarrow\downarrow\right),$$

then the qubit $\psi = \frac{1}{\sqrt{2}}(\alpha\uparrow + \beta\downarrow)$ becomes

$$\psi = \frac{1}{\sqrt{2}}(\alpha\uparrow + \beta\downarrow) \xrightarrow{U} \frac{1}{\sqrt{2}}(\alpha|0\rangle|0\rangle|0\rangle + \beta|1\rangle|1\rangle|1\rangle)\,.$$

We can protect each of the $|0\rangle$ and $|1\rangle$ against bit flip by the same method as above. For protection against a single phase flip, we notice that under a phase flip (of a single qubit) $|0\rangle$ becomes $|1\rangle$ and vice versa $|1\rangle$ becomes $|0\rangle$. Thus, if we treat $|0\rangle$ and $|1\rangle$ as a single two-level "particle", the problem of a phase flip is the same as the problem of a bit flip we had before.

Note, that although Shor's algorithm, was the first quantum error-correction code, it is not the most efficient. The most efficient code requires just 5 qubits (instead of the 9 here) to protect a single one.

Bibliography

[1] Asher Peres, Quantum Theory: Concepts and Methods, Kluwer Academic Publishers, 1995.

[2] John Preskill, Quantum Information and Computation, on-line notes, http://theory.caltech.edu/~preskill/ph229/.

Index