

Lecture notes on Quantum Information.
Delivered on February 2003 at Tel - Aviv
given by Dr. Benni Reznik
Written by Amir Seginer

Contents

Part 1. Introduction	5
Chapter 1. Physics and Information	7
1.1. Quantum information	7
Chapter 2. Basics of quantum information	15
2.1. Basics of quantum mechanics	15
2.2. Spin and the Pauli matrices	16
2.3. Open systems, mixtures and the density matrix	16
2.4. Entanglement and the Schmidt decomposition	33
Part 2. Entanglement	39
Chapter 3. Hidden variables	41
3.1. The EPR Paradox	41
3.2. Bell inequalities	42
3.3. Contextually	45
Chapter 4. Uses of Entanglement	51
4.1. Encoding information	51
4.2. Cryptography	52
4.3. teleportation	55
4.4. Remote operations	56
4.5. State-operators (stators)	58
4.6. POVM (Positive Operator Valued Measures)	60
4.7. Measure of entanglement (Distillation)	63
Chapter 5. Quantum information	69
5.1. Data compression (classical)	69
5.2. Data compression (Quantum)	70
5.3. Shumacher's noiseless encoding	73
5.4. Communication with noise (classical)	75
5.5. Accessible Information	76
5.6. Decoherence and the measurement problem	76
5.7. Error correction - Shor's algorithm	78

Part 1

Introduction

Physics and Information

1.1. Quantum information

We 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	does not influence system	effects 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 than the classical one.

1.1.1. 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.¹ We shall generally denote the two states as $|0\rangle$ and $|1\rangle$,² however the general state of such a system is

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

Since α and β are complex numbers, they are each described by two parameters (real and imaginary parts), which gives us four parameters which describe the state ψ . However, we have the two requirements $|\alpha|^2 + |\beta|^2 = 1$ and $\langle\psi|\psi\rangle = 1$, which reduce us to just two continuous parameters. 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.

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

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

¹The simplest, non-trivial Hilbert space, is a two dimensional one.

²The 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 $|E_1\rangle$.

and the second will simply be the spin up qubit

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

The overlap of these is

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

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 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.³ 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 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

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

while

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

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

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

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

³Recall 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 (not explicitly dependent on time).

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.⁴

1.1.3. One 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 is an integer. Now, in order to find classically whether n is even or odd, 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

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

and⁵

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

The evolution of the spin $|\uparrow\rangle_x$ is then given by⁶

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

Now if we choose λ so that

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

⁴Such an operator for two particles could be

$$U = \sum_i |i\rangle_i\langle i|_i\langle 0|_0 + \sum_{\substack{i,j \\ j \neq i, 0}} |j\rangle_i\langle i|_j + \sum_i |0\rangle_i\langle i|_i,$$

which gives

$$U|0\rangle_i = |i\rangle_i$$

and

$$UU^\dagger = \sum_i |j\rangle_i\langle i|_j = \mathbf{1}.$$

⁵Recall that in the z basis

$$\begin{aligned} |\uparrow\rangle_x &= \frac{1}{\sqrt{2}}(|\uparrow\rangle_z + |\downarrow\rangle_z), \\ |\downarrow\rangle_x &= \frac{1}{\sqrt{2}}(|\uparrow\rangle_z - |\downarrow\rangle_z). \end{aligned}$$

⁶since $\sigma_z^2 = \mathbf{1}$, then $\sigma_z^{2m} = \mathbf{1}$ and $\sigma_z^{2m+1} = \sigma_z$. Therefore the Taylor series for $e^{i\theta\sigma_z}$ can be written as

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

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").

1.1.4. 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 phase $e^{i\theta}$, then two parameters may be dropped giving us $2 \cdot 2^N - 2$. Which means that the number of parameters grows exponentially.⁷

1.1.5. examples.

1.1.5.1. *Deutsch's problem.* Assume a black box which accepts as input a single bit and outputs as a result a single bit. We shall denote the effect of the box as $f(x)$ [if the input bit is x then we get as output $f(x)$]. There are of course 4 different possible functions $f(x)$ which may describe the black box, since 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 balanced function, i.e. $f(0) \neq f(1)$.⁸

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 tell exactly what $f(x)$ and will therefore also tell us if $f(x)$ is constant or balanced. However, as we shall see,

⁷If we assume that we need at least one bit for every such parameter, 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 make simulation impossible very quickly.

⁸Note 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.

using quantum mechanics and the superposition principle we can find in just one run, what type of function $f(x)$ is.

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

The operator U_D we shall use, instead, will both accept and output two qubits of information as follows

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

where \oplus means addition 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

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

and

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

where in the second relation y and y' maybe the same or different. Now, since $\langle 0|1\rangle = 0$, we must have that the new set of states after the transformation U_D are mutually orthogonal, which means they must be a basis. Thus, the transformation U_D took us from one orthogonal basis to another, which means that it must be unitarian.⁹

To find whether our new, quantum, black box is a constant function or a balanced one, we can of course run it twice and see (once putting $x = 0$ and once $x = 1$). However, we can also use the superposition principle to determine this with just one run. 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

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

regardless of the value of $f(0)$, we have $0 \oplus f(0) \neq 1 \oplus f(0)$, and therefore if one of the kets in the parentheses is $|0\rangle$ the other must be $|1\rangle$ (and vice versa).

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

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

⁹If $|e_i\rangle$ is orthonormal basis and $|h_i\rangle$ is a second then the transformation from e to h is

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

which is clearly unitarian.

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)

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

The two new states, we found, are orthogonal to one another, and so may be distinguished in a single measurement [simply measure particle one in the basis $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$]. Thus, by a single *quantum* measurement we can distinguish whether $f(x)$ is constant or balanced, a feat we could not accomplish classically.

Note, that once again we managed 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.

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

1.1.5.3. *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*¹⁰

$$\psi^- = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B)$$

$$\psi^+ = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B)$$

$$\phi^- = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B - |1\rangle_A|1\rangle_B)$$

$$\phi^+ = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B),$$

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 all four states are entangled. The basis is also the set of mutual eigenvectors of the set of

¹⁰Note that the Bell states resemble the singlet and triplet states of two spins:

$$\begin{aligned} &\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \quad \text{singlet} \\ &\begin{cases} \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \\ |\uparrow\uparrow\rangle \\ |\downarrow\downarrow\rangle \end{cases} \quad \text{triplet} \end{aligned}$$

commuting operators¹¹

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

We now define 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,¹² it is easy to see that applying these operators on the Bell state $|\psi^-\rangle$ gives¹³

$$\begin{aligned} U_{00}^{(A)} |\psi^-\rangle &= |\psi^-\rangle, \\ U_{01}^{(A)} |\psi^-\rangle &= -|\phi^-\rangle, \\ U_{10}^{(A)} |\psi^-\rangle &= i|\phi^+\rangle, \\ U_{11}^{(A)} |\psi^-\rangle &= |\psi^+\rangle. \end{aligned}$$

Having constructed our tools, we can now turn to our original problem (communicating two bits of information using a single qubit). Imagine two distant persons, Alice holding particle A and Bob holding particle B , where we (and they) know that the particles are in the Bell state $|\psi^-\rangle$. Alice wishes to communicate to Bob *two* bits i and j ($i = 0, 1$ and $j = 0, 1$) of information. To do this Alice operates locally on her particle with the operator $U_{ij}^{(A)}$ we defined. As a result the two particles A and B together are in one of the orthonormal Bell states (up to a global phase). Now 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.¹⁴ since Bob knows, that the particles were originally in state $|\psi^-\rangle$, he can therefore infer which operator A 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.

Another 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

¹¹For 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

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

¹²Recall that

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

¹³Note, that the operator $U^{(A)}$ operates only on a single particle so 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 ,

¹⁴since 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.

(the density matrix is the Identity). Unlike Bob, any other person who gets the transmitted particle has no extra information and therefore cannot infer from it anything.

Basics of quantum information

2.1. Basics of quantum mechanics

Every theory is defined by the three things:

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

In quantum mechanics one can distinguish between two general cases. The first is the closed system in which we know all the elements of the system and we know how they interact with one another. The second, is the open system, in which besides the elements we deal with there is also an environment which we do *not* know how exactly it interacts with our system.

For a closed system, in quantum mechanics, a closed system is described

Observable quantities are described in quantum mechanics 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

$$|\psi\rangle = \sum_a \Pi_a |\psi\rangle.$$

The result of a measurement of the quantity A for a state $|\psi\rangle$ results in a collapse of the state into one of the subspaces of eigenvalue of A

$$|\psi\rangle \xrightarrow{\text{measure } A} \frac{\Pi_a |\psi\rangle}{\sqrt{\langle \psi | \Pi_a | \psi \rangle}}.$$

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 exist such a projection (for $\langle\psi_1|\psi_2\rangle \neq 0$). We define by a Gram-Schmidt process a new state $|\varphi_2\rangle$ orthonormal to $|\psi_1\rangle$:

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

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

Since $|\psi_1\rangle$ and $|\psi_2\rangle$ are not orthogonal while $|\psi_1\rangle$ and $|\varphi_2\rangle$ are, 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 $\langle\psi_2|\Pi|\psi_2\rangle = 0$ 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 $|\psi_1\rangle$ and $|\varphi_2\rangle$ are orthogonal, then so are their projections, so

$$\langle\psi_1|\Pi|\varphi_2\rangle = \langle\varphi_2|\Pi|\psi_1\rangle \quad (\Leftarrow \langle\psi_1|\varphi_2\rangle = 0),$$

and since Π is hermitian then

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

Thus (together with $|\alpha|^2, |\beta|^2 > 0$ and $\langle\psi_1|\Pi|\psi_1\rangle = 1$) we must have

$$0 = \langle\psi_2|\Pi|\psi_2\rangle > 0,$$

which is a contradiction, and therefore 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 $\langle\psi_1|\psi_2\rangle \neq 0$ □

2.2. Spin and the Pauli matrices

2.3. Open systems, mixtures and the density matrix

We have so far dealt only with closed quantum systems. we shall now deal with open systems. There are two general cases in which we deal with 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.
- (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 the closed system case. We shall see that we have to use mixtures instead of regular states, they will be described by density matrices, and 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 than linear combination of states (superposition) since here each state is treated separately and there is no interference effect.¹

To see the difference between open and closed systems let 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 on state $|\psi_A\rangle$ and in $p_b = 1 - p_a$ they will be on state $|\psi_B\rangle$. therefore the probability to measure $|0\rangle$ will be p_a times the probability to measure it 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 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 markedly different. In the mixture, assuming $a_0, b_0 \neq 0$, then no matter the values 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: we sum amplitudes. However in the mixture case, there is no interference and we sum probabilities.

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

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

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

$$\text{Tr } \rho = \sum_n \langle n | \rho | n \rangle = \sum_n \rho_{nn}.$$

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

¹Recall, 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.

(1) It's trace is 1:

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

(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 $\langle \psi | \rho | \psi \rangle \geq 0$, or equivalently, all its eigenvalues are nonnegative

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

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

$$0 \leq \lambda_k \leq 1.$$

PROOF. (1) To prove that $\text{Tr } \rho = 1$ 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

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

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 ($0 \leq p_i \leq 1$), 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

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

(3) To show that the density matrix ρ is a positive operator we shall use its definition to calculate $\langle \psi | \rho | \psi \rangle$, 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 this necessarily means that

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

If we now take as a special case $|\psi\rangle = |\varphi_k\rangle$, where $|\varphi_k\rangle$ is an eigenvector of ρ with eigenvalue λ_k ($\rho|\varphi_k\rangle = \lambda_k|\varphi_k\rangle$), then from the last result we must have²

$$\begin{aligned} 0 \leq \langle \varphi_k | \rho | \varphi_k \rangle &= \langle \varphi_k | \lambda_k | \varphi_k \rangle = \lambda_k \\ &\Rightarrow \lambda_k \geq 0. \end{aligned}$$

□

The density matrix describes mixtures of states, but it can also describe a regular state. This is case when the mixture includes only a single state with probability $p = 1$. We say that such a mixture is a *pure state* if there ex:

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

The density matrix of a pure state has the special trait that

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

This last trait is true only for pure states. If we diagonalize such a matrix then its diagonal must be (since $\sum \lambda_k = 1$ and $\lambda_k \geq 0$) zero everywhere except a single element which is 1. Thus such a density matrix has a single eigenvector $|\varphi\rangle$ with eigenvalue 1 and all the rest with eigenvalue 0. This special eigenvector $|\varphi\rangle$ defines the density matrix of the pure state

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

The density matrix (of any mixture, pure or not) has one more important trait. For any projection operator $\Pi = |\psi\rangle\langle\psi|$ onto a single state $|\psi\rangle$, the probability of it measuring true (the mixture collapses into state $|\psi\rangle$ after the measurement) is

$$\text{prob}(\Pi = 1) = \text{prob}(|\psi\rangle) = \text{Tr}(\rho\Pi) = \text{Tr}(\Pi\rho).$$

This trait can be generalized, in which case the average eigenvalue $\langle O \rangle$ of an observable O , when measured is

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

PROOF. We shall start by proving the simple form of the trait. By definition

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

²We 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 $\langle \psi | \rho | \psi \rangle \geq 0$.

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

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

which proves the simpler trait.

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

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

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 (not complex) 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 two constraints

$$\rho^\dagger = \rho$$

and

$$\text{Tr} \rho = 1.$$

The first constraint is actually $N^2 + N$ equations since on the diagonal $\rho^\dagger = \rho$ means

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

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

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

where in off diagonal case, the number of equations we wrote, takes into account that each equality is actually two, one for the real part and a second for the imaginary part. On the other hand, we counted only the pairs nm above the diagonal, since those below will give us the same equations again. We did not double the equations for the elements on the diagonal since the equations only mean that the imaginary part is zero, but does not tell us anything about the real part (it equals itself, which is trivial).

To the above constraints we must also add the trace

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

which is only one equation, since we already know that the trace has no imaginary component. Subtracting the number of equations from the total number of parameters (in the case of no constraints) we finally get

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

If the density matrix has so few parameters, we might try and construct it as a linear combinations of other types of matrices with the same number of parameters. For example let us try and construct them from the matrices which form the basis of the $SU(N)$ group.³ More explicitly, let us try for the case $N = 2$

For $N = 2$, one possible basis of $SU(2)$ is the Pauli matrices σ_i . 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

$$\langle A, B \rangle \equiv \text{Tr}(AB).$$

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

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

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 the density matrix can be written as

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

³The $SU(N)$ group (**S**pecial **U**nitary group) is the group of all $N \times N$ unitary matrices, with determinate 1 (unitary matrices could also have a determinant of -1)

$$U \in SU(N) \Rightarrow U^\dagger U = \mathbf{1}, |U| = +1.$$

The $SU(N)$ group has $N^2 - 1$ independent parameters, and therefore has a basis of $N^2 - 1$ matrices.

To find the coefficients a_i we apply the constraints we had on the density matrix. first we apply the constraint on the trace $\text{Tr } \rho = 1$. Since $\text{Tr } \sigma_i = 0$ for the Pauli matrices ($i = 1, 2, 3$), then this condition gives us (recall that $\mathbb{1}$ is a 2×2 matrix)

$$\begin{aligned} 1 &= \text{Tr } \rho = a_0 \text{Tr } \mathbb{1} = 2a_0 \\ &\Rightarrow a_0 = \frac{1}{2}. \end{aligned}$$

Now the second requirement is that ρ be Hermitian ($\rho^\dagger = \rho$) since the Pauli matrices themselves (and $\mathbb{1}$) are Hermitian then we have

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

which means that⁴

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

For convenience we define

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

as a result of which we may write the density matrix as

$$\begin{aligned} \rho &= \frac{1}{2}(\mathbb{1} + \vec{p} \cdot \vec{\sigma}) \quad (\vec{p} \in \mathbb{R}^3) \\ \Rightarrow \rho &= \frac{1}{2} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + p_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + p_3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right] = \frac{1}{2} \begin{pmatrix} 1 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & 1 - p_3 \end{pmatrix}. \end{aligned}$$

The last 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-negative⁵

$$|\rho| \geq 0.$$

From the form we found for ρ this means that

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

Therefore we finally have the general form of the density matrix

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

where the vector \vec{p} is called the *polarization*, since it gives the average direction of the direction/polarization of the spin

$$\langle \sigma_x \rangle = \text{Tr}(\rho \sigma_x) = p_x,$$

$$\langle \sigma_y \rangle = \text{Tr}(\rho \sigma_y) = p_y,$$

$$\langle \sigma_z \rangle = \text{Tr}(\rho \sigma_z) = p_z.$$

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

As a conclusion 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*. We shall

⁴Since the Pauli matrices together with the unit matrix constitute a basis, then there is only one possible choice of coefficients a_i which gives a certain matrix (if there were more the elements of the basis wouldn't be linearly independent). The above condition gives two sets of coefficients $\{a_i\}$ and $\{a_i^*\}$. If these sets are the same one then $a_i = a_i^*$.

⁵The 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 it suffices that the product of the eigenvalues be positive for them both to be positive.

see that the case $|\vec{p}| = 1$ (points on the surface of the sphere) coincides with pure states.

CLAIM. If and only if $\vec{p} = \hat{n}$ (iff \vec{p} is a unit vector), then there exists a state $|\psi\rangle$ such that

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

i.e. $\rho(\hat{n})$ describes a pure state.

PROOF. Let us prove in one direction. If we have $\rho^2 = \rho$, then we know that the state is pure, and can therefore be written in the above form $\rho = |\psi\rangle\langle\psi|$.

By definition

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

and

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

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 we must have

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

and ρ^2 becomes

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

which 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 $|\psi\rangle$).⁶

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

⁶If we define \hat{n} by the spherical angles θ and φ

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

the state $|\psi\rangle$ is

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

since this state 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. If $\rho = |\psi\rangle\langle\psi|$ then only $|\psi\rangle$ (up to a global phase) is eigenvector with eigenvalue of 1.

We have so far concentrated on qubits and the two dimensional Hilbert space. In an N dimensional space we can use the $N^2 - 1$ Hermitian matrices h_i ($i = 1, 2, \dots, N^2 - 1$) with zero trace, which 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

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

where

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

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 where all the eigenvalues are positive and add up to 1 ($\text{Tr} \rho = 1$)

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

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 we do not allow).

All we have just said is true for any N including $N = 2$, but there is a big difference between $N = 2$ and $N > 2$. If $N = 2$ then we have just two eigenvalues for the density matrix. Since their sum 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 therefore when one becomes zero, it doesn't necessarily mean that one of the other eigenvalues becomes 1 and all the rest zero, but rather that the sum of all the rest must be 1. As a result the systems described by the boundary, 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 said before, there are two basic cases in which we must use density matrices, when we lack information or when studying only part of the system. We shall now elaborate on these to cases.

Assume two sources A and B of particles, source A produces particles in random states $\{p_A, \psi_A\}$ described by the density matrix ρ_A (pure or not) and source B produces particles in random states $\{p_B, \psi_B\}$ described by the density matrix ρ_B (again, pure or not). Now, we want to create a new set of states. to do this we pick states out of source A with probability λ and states from 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

$$\{\lambda p_A, \psi_A\} \cup \{(1 - \lambda) p_B, \psi_B\}$$

which, by the definition of the density matrix gives us⁷

$$\rho_{AB} = \sum_A \lambda p_A |\psi_A\rangle \langle \psi_A| + \sum_B (1 - \lambda) p_B |\psi_B\rangle \langle \psi_B| \equiv \lambda \rho_A + (1 - \lambda) \rho_B,$$

or simply⁸

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

⁷In the new collection of states there is a chance λp_A for state $|\psi_A\rangle$ to occur and $(1 - \lambda) p_B$ for state $|\psi_B\rangle$. Since $\sum \lambda p_A + \sum (1 - \lambda) p_B = 1$, then we can treat it just as a set of the form $\{p_i, \psi_i\}$, where $p_i \in \{\lambda p_A\} \cup \{(1 - \lambda) p_B\}$ and $\psi_i \in \{\psi_A\} \cup \{\psi_B\}$.

⁸A sum of the type

$$\vec{u} = \lambda \vec{v}_1 + (1 - \lambda) \vec{v}_2 \quad (0 \leq \lambda \leq 1)$$

Let us now check that the new matrix ρ_{AB} is indeed a density matrix. It's trace is indeed 1

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

It is clearly Hermitian (λ is real)

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

And 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 a density matrix, so 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 from them a new non-pure mixture, by the method above.

Note, that we may have two physically different sources which give the same density matrix. For example assume that 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

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

Now, assume 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

$$\rho_B = \frac{1}{2} |+\rangle\langle +| + \frac{1}{2} |-\rangle\langle -| = \frac{1}{2} \mathbf{1}.$$

We see that although both sources are physically different, we get the same density matrix in both,⁹ which means that we cannot distinguish between the two cases by our measurements.

The above result is more general, and actually every non-pure density matrix (not just $\frac{1}{2}\mathbf{1}$) can be a result of an infinite number of systems.¹⁰ For the Bloch sphere this is easily shown since any vector \vec{p} ($|\vec{p}| \leq 1$) may be written as¹¹

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

for some $0 \leq \lambda \leq 1$ and some unit vectors \hat{n}_1, \hat{n}_2 which are not necessarily orthogonal [$\rho(\vec{p})$ is the density matrix defined by \vec{p} and $\rho(\hat{n}_i)$ are pure-state density

is called a *convex sum*. We say that a space is a *convex space* if for *every* two vectors in the space if any vector $\vec{u} = \lambda \vec{v}_1 + (1 - \lambda) \vec{v}_2$ constructed from them by a convex sum ($0 \leq \lambda \leq 1$), also belongs to the same space.

⁹In fact every choice of two orthonormal states with equally probability to occur will give us the same density matrix $\rho = \frac{1}{2}\mathbf{1}$.

¹⁰We prove it here for a two dimensional Hilbert space, but it must be true for any Hilbert space of higher dimension, since you examine just a two-dimension subspace of the larger Hilbert space and use the result proved here.

¹¹As long as $0 \leq \lambda \leq 1$ and \hat{n}_1, \hat{n}_2 are any unit vectors, this will give us $|\vec{p}| < 1$. To see this find \vec{p}^2 , which after a little calculation gives

$$\vec{p}^2 = 1 - 2\lambda(1 - \lambda)(1 + \hat{n}_1 \cdot \hat{n}_2).$$

It can easily be shown that

$$(1 + \hat{n}_1 \cdot \hat{n}_2) \leq 2$$

and

$$\lambda(1 - \lambda) \leq \frac{1}{4}$$

which gives the desired result ($\vec{p}^2 \leq 1$).

matrices defined by \hat{n}_i .¹² The same vector \vec{p} may be constructed from an infinite choice of unit vectors \hat{n}_1, \hat{n}_2 - simply rotate the two vectors *together* around the vector \vec{p} . The only exception to this is when $\lambda = 0$ or $\lambda = 1$, in which case $\vec{p} = \hat{n}_1$ ($\lambda = 1$) or $\vec{p} = \hat{n}_2$ ($\lambda = 0$) and rotating the vectors about \vec{p} does not change a thing (one vector doesn't change and the other, multiplied by zero, does not contribute to \vec{p}). Since each pair of unit vectors $\{\hat{n}_1, \hat{n}_2\}$ describe a different physical system,¹³ then the same density matrix can be a result of an infinite choice of systems, which we cannot distinguish between.

We have just seen that there is only one way to write pure states in the form

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

however, might we be able to construct it from non-pure states

$$\hat{n} = \lambda \vec{p}_1 + (1 - \lambda) \vec{p}_2 \Leftrightarrow \rho(\hat{n}) = \lambda \rho(\vec{p}_1) + (1 - \lambda) \rho(\vec{p}_2)?$$

We shall now see that the answer is “no”.

If the above equality holds then

$$(\lambda \vec{p}_1 + (1 - \lambda) \vec{p}_2)^2 = 1.$$

Clearly this does not hold when \vec{p}_1 is parallel to \vec{p}_2 , unless we are in one of the trivial cases:

- $\vec{p}_1 = \vec{p}_2 = \hat{n}$
- $\vec{p}_1 = \hat{n}$ and $\lambda = 1$
- $\vec{p}_2 = \hat{n}$ and $\lambda = 0$

We are left to check the cases where \vec{p}_1 and \vec{p}_2 are not parallel. In those case we may write ($p_i = |\vec{p}_i| \leq 1$)

$$\begin{aligned} (\lambda \vec{p}_1 + (1 - \lambda) \vec{p}_2)^2 &= \lambda^2 p_1^2 + (1 - \lambda)^2 p_2^2 + 2\lambda(1 - \lambda) \vec{p}_1 \cdot \vec{p}_2 \\ &< \lambda^2 p_1^2 + (1 - \lambda)^2 p_2^2 + 2\lambda(1 - \lambda) p_1 p_2 \\ &= [\lambda p_1 + (1 - \lambda) p_2]^2 \\ &< [\lambda + (1 - \lambda)]^2 = 1, \end{aligned}$$

where the first inequality is due to the fact that for non-parallel vectors $\vec{p}_1 \cdot \vec{p}_2 < p_1 p_2$, and the second inequality is due to the fact that $p_1, p_2 \leq 1$, but at least one is shorter than unity since otherwise we are back to the case $\vec{p}_1 = \hat{n}_1$ and $\vec{p} = \hat{n}_2$ which we have already treated (and saw that only the trivial cases give a pure state).

We have therefore seen that unless we are in one of the above trivial cases then

$$|\lambda \vec{p}_1 + (1 - \lambda) \vec{p}_2| < 1 \quad \left(\text{unless } \begin{cases} \vec{p}_1 = \vec{p}_2 = \hat{n} \\ \text{or} \\ \vec{p}_1 = \hat{n}, \lambda = 1 \\ \text{or} \\ \vec{p}_2 = \hat{n}, \lambda = 0 \end{cases} \right).$$

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.

¹²Note, that orthogonal states correspond to unit vectors of opposite direction ($\hat{n}_1 = -\hat{n}_2$) and not to orthogonal vectors. $\hat{n}_1 = \hat{z}$ gives the pure state $|\uparrow_z\rangle\langle\uparrow_z| = |0\rangle\langle 0|$ and $\hat{n}_2 = -\hat{z}$ gives the pure state $|\downarrow_z\rangle\langle\downarrow_z| = |1\rangle\langle 1|$, while $\hat{n}_3 = \hat{x}$ gives the pure state $|\uparrow_x\rangle\langle\uparrow_x| = |+\rangle\langle +|$ ($|0\rangle$ and $|1\rangle$ are orthogonal but $|+\rangle$ and $|0\rangle$ are not).

¹³If we define

$$\rho_i = \frac{1}{2}(\mathbb{1} + \hat{n}_i \cdot \vec{\sigma}),$$

then ρ_i is a pure state. A different choice of $\{\hat{n}_1, \hat{n}_2\}$ means that we are creating our new ensemble out of sources which emit different pure states, and are therefore physically different ensembles.

2.3.3. combined systems and partial trace. 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 systems to one).* Assume two Hilbert spaces \mathcal{H}_A and \mathcal{H}_B (not necessarily of the same dimension), each with its own states

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

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

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

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

If the original spaces $\mathcal{H}_A, \mathcal{H}_B$ 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 in every space

$$[\alpha|\psi_i^A\rangle] \otimes [\beta|\varphi_j^B\rangle] = \alpha\beta[|\psi_i^A\rangle \otimes |\varphi_j^B\rangle].$$

- The tensor product is distributive

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

Note, that usually we shall 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 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.* Now imagine that we have some state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and an operator O_{AB} of the form¹⁴

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

where of course A operates on the degrees of freedom of \mathcal{H}_A (and $\mathbb{1}_B$ operates on the degrees of freedom of \mathcal{H}_B). The density matrix describing the system (in the $\mathcal{H}_A \otimes \mathcal{H}_B$ space) is ρ_{AB} (pure or not). The result of measuring O_{AB} is then given by the usual rules of quantum mechanics and density matrices by

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

where Tr_B means taking the trace in \mathcal{H}_B only and \mathcal{H}_A means taking the trace of H_A only. Note, that after taking the trace over B , the operator $\text{Tr}_B \rho_{AB}$ now operates solely on A and we can therefore drop the tensor product (\otimes). If we now define

$$\rho_A \equiv \text{Tr}_B \rho_{AB},$$

then we may rewrite the previous equation as

$$\langle A \rangle_{AB} = \text{Tr}_A [A \rho_A] \quad (\rho_A = \text{Tr}_B \rho_{AB}).$$

Thus we get the regular expression for density matrices on the smaller Hilbert space \mathcal{H}_A .

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

To clarify what a partial trace, and 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 $\langle A \rangle$ is

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

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

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

and then (note the double index m)

$$\rho_A = \sum_m \rho_{i,m,n,m} = \rho_{i,j}^A.$$

We have so far defined a partial trace and a reduced matrix, but what do they give us? What we have seen is that if we are given a state described by ρ_{AB} , and we are interested in the result of measurements on subsystem A alone, then we may take the partial trace of the full system and find a standard density-matrix description of the subsystem with the density matrix $\rho_A = \text{Tr}_B \rho_{AB}$. Note, that

¹⁴Note that the operator, operating on \mathcal{H}_B is the identity.

when we start with a pure state in the $\mathcal{H}_A \otimes H_B$ Hilbert space we might end with a non-pure density matrix ρ_A . Therefore, by looking at only a part of the system, an actually pure state might give rise to a non-pure state (it is non-pure, for all practical purposes, for a person living only in the \mathcal{H}_A Hilbert space).

2.3.3.3. *How do measurements on \mathcal{H}_B influence ρ_A .* One may ask what happens if one (Bob) makes a measurement on his part of the system \mathcal{H}_B before a measurement is performed (by Alice) on the other part of the system \mathcal{H}_A . There are two cases to treat, the first is when Bob makes a selective measurement, i.e. he makes a measurement, and only if he got the result he wanted Alice can make her measurements (we calculate ρ_A just for certain results of Bob's measurement). The second case is when Bob makes his measurement (on \mathcal{H}_B), but regardless of his result Alice makes her measurements and determines ρ_A . As we shall see, in the first case the resulting density matrix ρ_A is not effected only if we start with a non-entangled state, while in the second case, ρ_A is not effected, no matter what type of system we start with.

Let us now see what happens. In the most general case the system is described by the state

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

which gives the density matrix

$$\rho_{AB} = \left(\sum_i \alpha_i |\psi_i\rangle_A |\varphi_i\rangle_B \right) \left(\sum_j \alpha_j^* \langle\psi_j|_B \langle\varphi_j| \right) = \sum_{i,j} \alpha_i \alpha_j^* |\psi_i\rangle_A |\varphi_i\rangle_{BB} \langle\varphi_j|_A \langle\psi_j|$$

and therefore ρ_A is

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

Now, assume that Bob makes a measurement of operator B and gets a result b . This occurs with probability p_b . In such a case all the $|\varphi_i\rangle_B$ which are not orthogonal to $|b\rangle$ will collapse to $|b\rangle$ (and the rest will vanish from our sum). To denote that only part of the state might survive we shall now some of i', j' where the i' are subset of $i = 1, \dots, N_B$ so that

$${}_B \langle b|\varphi_{i'}\rangle \neq 0,$$

and similarly for the j' . Our new state after Bob's measurement is therefore

$$|\Psi'_{AB}\rangle = \sum_{i'} \alpha_{i'} |\psi_{i'}\rangle_A |b\rangle_B,$$

which will appear with probability p_b . The reduced matrix as a result (${}_B \langle\varphi_j|\varphi_i\rangle_B \rightarrow {}_B \langle b|b\rangle_B = 1$) becomes

$$\rho'_A = \sum_{i',j'} \alpha_{i'} \alpha_{j'}^* |\psi_{i'}\rangle_{AA} \langle\psi_{j'}| \quad (\text{appears with probability } p_b).$$

This new density matrix (ρ'_A) will be the same as the above ρ_A if and only if $|\Psi_{AB}\rangle$ is a non-entangled state.

PROOF. To prove the above statement we shall show that if $|\Psi_{AB}\rangle$ is entangled then $\rho'_A \neq \rho_A$ and if $|\Psi_{AB}\rangle$ is not entangled then $\rho'_A = \rho_A$. We start by writing the

original state $|\Psi_{AB}\rangle$ using the Schmidt decomposition (see below). We can always write the state as a sum of the form

$$\sum_i \sqrt{\lambda_i} |\tilde{\psi}_i\rangle_A |\tilde{\varphi}_i\rangle_B,$$

where

$${}_A\langle\tilde{\psi}_i|\tilde{\psi}_j\rangle_A = \delta_{ij} \quad \text{and} \quad {}_B\langle\tilde{\varphi}_i|\tilde{\varphi}_j\rangle_B = \delta_{ij},$$

and

$$\lambda_i \in \mathbb{R}.$$

If we do not make a measurement then using the above result, the density matrix, this time $({}_B\langle\tilde{\varphi}_i|\tilde{\varphi}_j\rangle_B = \delta_{ij})$ will be

$$\rho_A = \sum_i \lambda_i |\tilde{\psi}_i\rangle_{AA} \langle\tilde{\psi}_i|.$$

However, if we do make a measurement, then as above, we will get

$$\rho'_A = \sum_{i',j'} \sqrt{\lambda_{i'}} \sqrt{\lambda_{j'}} |\tilde{\psi}_{i'}\rangle_{AA} \langle\tilde{\psi}_{j'}|,$$

where again the prime in i', j' denote that this may be only a subset of the possible values of i (and j). Since we are now in an orthonormal basis, the two density matrices will be the same only if all the coefficients of the different $|\tilde{\psi}_i\rangle_{AA} \langle\tilde{\psi}_j|$ are the same.¹⁵ Since mixed elements ($|\tilde{\psi}_{i'}\rangle_{AA} \langle\tilde{\psi}_{j'}|$ where $i' \neq j'$) appear only in ρ'_A then we *cannot* have $\rho'_A = \rho_A$. The only exception to this rule is for a non-entangled state where i has only a single possible value ($i = 1$).¹⁶ We have thus shown that $\rho'_A = \rho_A$ (the density matrix after a selective measurement equals the density matrix if no measurement is made) if and only if $|\Psi_{AB}\rangle$ is non-entangled. \square

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

PROOF. Let us assume that Bob measures operator B with eigenvectors $|i\rangle_B$. We can write all the states $|\varphi_i\rangle_B$ as a linear combination of these eigenvectors (the eigenvalues may be degenerate but we need the eigenvectors that span the Hilbert space \mathcal{H}_B - we assume that the index i runs over all the *eigenvectors*). Our initial state can therefore be written as

$$|\Psi_{AB}\rangle = \sum_i \beta_i |\psi_i\rangle_A |i\rangle_B,$$

¹⁵Note, that $|\psi_i\rangle_A$ and $|\varphi_j\rangle_B$ (before taking a subset, and using i', j') don't have to span the whole Hilbert spaces \mathcal{H}_A and \mathcal{H}_B (respectively), but may span only a subspace of each of the Hilbert spaces. Thus to give a complete density matrix we must add more states which orthonormal to the ones we already have. Never the less the element of the density matrix ${}_A\langle\tilde{\psi}_i|\rho_A|\tilde{\psi}_j\rangle_A$ does not depend on the choice of the other (orthonormal) states (this goes for ρ'_A as well).

¹⁶The only possible case in which ρ'_A has no mixed elements, is if it is of the form

$$\rho'_A = \lambda_{i'} |\tilde{\psi}_{i'}\rangle_{AA} \langle\tilde{\psi}_{i'}|,$$

i.e. i' accepts only a single value. The only case in which this would equal the sum $\sum_i |\psi_i\rangle_{AA} \langle\psi_i|$ over all i 's, is if i accepts only a single value.

where the $|\psi_i\rangle_A$ may either be orthogonal or not. As a result, the density matrix will be

$$\rho_{AB} = \sum_{i,j} \beta_i \beta_j^* |\psi_i\rangle_A |i\rangle_{BB} \langle j|_A \langle \psi_j|.$$

Taking the partial trace over the states in \mathcal{H}_B gives then (since $|i\rangle_B$ are orthonormal)

$$\begin{aligned} \rho_A &= \sum_k {}_B \langle k| \left(\sum_{i,j} \beta_i \beta_j^* |\psi_i\rangle_A |i\rangle_{BB} \langle j|_A \langle \psi_j| \right) |k\rangle_B \\ &= \sum_k |\beta_k|^2 |\psi_k\rangle_{AA} \langle \psi_k| \end{aligned}$$

Now, If Bob measures the eigenvalue b , with probability p_b then $|\Psi_{AB}\rangle$ collapses, and only eigenvectors with eigenvalue b survive and we get

$$|\Psi_{AB}\rangle \xrightarrow{b \text{ measured}} \frac{1}{\sqrt{p_b}} \sum_{i_b} \beta_{i_b} |\psi_{i_b}\rangle_A |i_b\rangle_B,$$

where the subscript b means that we take only the values of i for which $|i\rangle_B$ are eigenvectors with eigenvalues b . The factor of $\frac{1}{\sqrt{p_b}}$ outside is merely for normalization.

Using the last result and taking a partial trace we this time

$$\rho_A(b) = \frac{1}{p_b} \sum_{k_b} |\beta_{k_b}|^2 |\psi_{k_b}\rangle_{AA} \langle \psi_{k_b}|.$$

However, since Alice measures all systems, regardless of the result Bob got, then to find the density matrix describing what she sees we must combine the different cases of b . We saw (see beginning of density matrices) that density matrices of different case are joined to one by a linear combination, each with a weight equally to the probability of it occurring

$$\rho_A = \sum_b p_b \rho(b).$$

Since we sum over all possible values of b then we get

$$\rho_A = \sum_b p_b \left(\frac{1}{p_b} \sum_{k_b} |\beta_{k_b}|^2 |\psi_{k_b}\rangle_{AA} \langle \psi_{k_b}| \right) = \sum_k |\beta_k|^2 |\psi_k\rangle_{AA} \langle \psi_k|,$$

which is the same result we got when Bob hadn't made any measurement. Thus if Bob makes a non-selective measurement, Alice's measurements will not be influenced. \square

Before proceeding one should notice one consequence of the proof. If we start with a non-entangled state then the reduced matrix describes a pure state and if we start with an entangled state we end with a non-pure state

$$\begin{aligned} |\Psi_{AB}\rangle &= |\psi\rangle_A |\varphi\rangle_B \Rightarrow \rho_A = |\psi\rangle_{AA} \langle \psi| = \rho_A^2 \\ |\Psi_{AB}\rangle &= \sum_i \alpha_i |\psi_i\rangle_A |\varphi_i\rangle_B \Rightarrow \rho_A = \sum_i |\alpha_i|^2 |\psi_i\rangle_{AA} \langle \psi_i| = \rho_A^2 \end{aligned}$$

or simply

$$\begin{aligned} \text{non-entangled} &\Rightarrow \text{pure} \\ \text{entangled} &\Rightarrow \text{non-pure.} \end{aligned}$$

2.3.3.4. *The GHJW theorem*¹⁷. We saw before that different physical description may bring about the same density matrix. That is we may have a source emitting states $|\psi_i\rangle$ with probability p_i ($\{p_i, |\psi_i\rangle\}$) and we may have a source emitting states $|\varphi_i\rangle$ with probability q_i . If both sources result in the same density matrix

$$\rho_A = \sum p_i |\psi_i\rangle\langle\psi_i| = \sum q_i |\varphi_i\rangle\langle\varphi_i|,$$

then we say that they are two different *realization* of the density matrix ρ_A . Note that the set $\{|\psi_i\rangle\}$ and $\{|\varphi_i\rangle\}$ are not necessarily sets of orthogonal states (within the set, or between the sets).

The GHJW theorem says that all realizations, consisting of up to n pure states ($i = 1, \dots, n$), of the *same* density matrix ρ_A , may be produced from a single pure state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, where \mathcal{H}_B is (at least) n dimensional. To produce the different realizations, one must measure (non-selectively) for each realization a suitable operator B in the \mathcal{H}_B space.

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

$$\{p_i, |\psi_i\rangle\} \Rightarrow \rho_A = \sum p_i |\psi_i\rangle\langle\psi_i|,$$

$$\{q_i, |\varphi_i\rangle\} \Rightarrow \rho_A = \sum q_i |\varphi_i\rangle\langle\varphi_i|.$$

We may perform a “purification” of the two, by enlarging our Hilbert space to $\mathcal{H}_A \otimes \mathcal{H}_B$

$$\{p_i, |\psi_i\rangle\} \xrightarrow{\text{purification}} |\Phi_1\rangle_{AB} = \sum \sqrt{p_i} |\psi_i\rangle_A |\alpha_i\rangle_B \quad ({}_B\langle\alpha_i|\alpha_j\rangle_B = \delta_{ij}),$$

$$\{q_i, |\varphi_i\rangle\} \xrightarrow{\text{purification}} |\Phi_2\rangle_{AB} = \sum \sqrt{q_i} |\varphi_i\rangle_A |\beta_i\rangle_B \quad ({}_B\langle\beta_i|\beta_j\rangle_B = \delta_{ij}),$$

where the states in \mathcal{H}_B in each case are orthonormal (the $|\alpha_i\rangle$ are orthonormal and the $|\beta_i\rangle$ are orthonormal). The result of the “purifications” are the two pure states $|\Phi_1\rangle_{AB}$ and $|\Phi_2\rangle_{AB}$ which after a partial trace, each gives the density matrix ρ_A .

We know (see second proof in previous subsection) that if we measure a pure state *non-selectively*¹⁸ using an operator B with non-degenerate eigenvectors $|\beta_i\rangle$, the resulting density matrix will behave as if our pure system was of the form

$$\sum \sqrt{q_i} |\varphi_i\rangle_A |\beta_i\rangle_B.$$

Thus if we measure it using the operator $U_B B U_B^{-1}$ our eigen values will be $U_B |\beta_i\rangle$ and the resulting density matrix will behave as if our pure system was of the form

$$\sum \sqrt{\tilde{q}_i} |\tilde{\varphi}_i\rangle_A U_B |\beta_i\rangle_B.$$

Therefore, if we prove that

$$|\Phi_1\rangle_{AB} = (\mathbb{1}_A \otimes U_B) |\Phi_2\rangle_{AB},$$

then

$$|\Phi_1\rangle_{AB} \equiv \sum \sqrt{p_i} |\psi_i\rangle_A |\alpha_i\rangle_B = \sum \sqrt{\tilde{q}_i} |\tilde{\varphi}_i\rangle_A U_B |\beta_i\rangle_B,$$

and measuring $|\Phi_2\rangle_{AB}$ using $U_B B$ will give us the same realization as would have $|\Phi_1\rangle_{AB}$.

¹⁷GHJW stands for Gisin, Hughston, Jozsa and Wootters.

¹⁸Non-selectively means that we make the measurement B and then allow another measurement on space \mathcal{H}_A to be performed, in order to determine ρ_A (from the statistics we obtain). As we have already seen, this procedure does not effect the density matrix ρ_A . This is different from the case of the selective measurement where after the measurement B we throw away result which do not give us the eigenvalues we wanted.

Thus, in order to complete the proof we are left to show that there must exist a unitary operator U_B such that

$$|\Phi_1\rangle_{AB} = (\mathbb{1}_A \otimes U_B)|\Phi_2\rangle_{AB}.$$

To prove this we use the Schmidt decomposition (see below). By the Schmidt decomposition $|\Phi_1\rangle_{AB}$ and $|\Phi_2\rangle_{AB}$ may be written as

$$|\Phi_1\rangle_{AB} = \sum \sqrt{\lambda_i} |a_i\rangle_A |b_i\rangle_B \quad \left(\begin{array}{l} {}_A\langle a_i | a_j \rangle_A = \delta_{ij} \\ {}_B\langle b_i | b_j \rangle_B = \delta_{ij} \end{array} \right)$$

$$|\Phi_2\rangle_{AB} = \sum \sqrt{\lambda'_i} |a'_i\rangle_A |b'_i\rangle_B \quad \left(\begin{array}{l} {}_A\langle a'_i | a'_j \rangle_A = \delta_{ij} \\ {}_B\langle b'_i | b'_j \rangle_B = \delta_{ij} \end{array} \right).$$

Furthermore we must have

$$\begin{aligned} |a'_i\rangle &= |a_i\rangle, \\ \lambda'_i &= \lambda_i, \end{aligned}$$

since both states after taking a partial trace of them give the same diagonal density matrix, and the elements on the diagonal are the $\lambda_i |a_i\rangle\langle a_i|$. Thus

$$|\Phi_1\rangle_{AB} = \sum \sqrt{\lambda_i} |a_i\rangle_A |b_i\rangle_B \quad \left(\begin{array}{l} {}_A\langle a_i | a_j \rangle_A = \delta_{ij} \\ {}_B\langle b_i | b_j \rangle_B = \delta_{ij} \end{array} \right)$$

$$|\Phi_2\rangle_{AB} = \sum \sqrt{\lambda_i} |a_i\rangle_A |b'_i\rangle_B \quad \left(\begin{array}{l} {}_A\langle a'_i | a'_j \rangle_A = \delta_{ij} \\ {}_B\langle b'_i | b'_j \rangle_B = \delta_{ij} \end{array} \right).$$

Since $|b_i\rangle_B$ and $|b'_i\rangle_B$ are orthonormal bases then there must exist a unitary transformation between them

$$|b_i\rangle_B = U_B |b'_i\rangle_B,$$

so that

$$|\Phi_1\rangle_{AB} = (\mathbb{1}_A \otimes U_B)|\Phi_2\rangle_{AB}.$$

This completes the proof. \square

2.4. Entanglement and the Schmidt decomposition

We shall say that a state $|\psi_{AB}\rangle$ is *entangled* if we *cannot* decompose it into a tensor product (no matter which basis we choose). For example the state

$$|\psi_{AB}\rangle = \alpha |a\rangle_A |0\rangle_B + \beta |1\rangle_A |0\rangle_B,$$

is not entangled since we can write it as a tensor product

$$|\psi_{AB}\rangle = [\alpha |a\rangle_A + \beta |1\rangle_A] \otimes |0\rangle_B.$$

For two spin $\frac{1}{2}$ particles the most general form of a non-entangled state may be written as

$$\begin{aligned} |\psi_{AB}\rangle &= [a|0\rangle_A + b|1\rangle_A] \otimes [\alpha|0\rangle_B + \beta|1\rangle_B] \\ &= a\alpha|0\rangle_A|0\rangle_B + a\beta|0\rangle_A|1\rangle_B + b\alpha|1\rangle_A|0\rangle_B + b\beta|1\rangle_A|1\rangle_B. \end{aligned}$$

Thus the state

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

is an entangled state. To see this, note that in the general non-entangled state, if $a\alpha \neq 0$ and $b\beta \neq 0$ then $a\beta$ and $b\alpha$ must also be non-zero. Thus if $|\psi_{AB}\rangle$ non-entangled and includes elements $|0\rangle_A|0\rangle_B$ and $|1\rangle_A|1\rangle_B$ it must also include the mixed elements $|0\rangle_A|1\rangle_B$ and $|1\rangle_A|0\rangle_B$. 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 $|1\rangle_A|0\rangle_B$, so it cannot be non-entangled and is therefore entangled.

THEOREM. For any state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ 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 $|\Psi_{AB}\rangle$ can be written as

$$\sum_{i=1}^{N \leq \min(N_A, N_B)} \sqrt{\lambda_i} |i\rangle_A |\tilde{i}\rangle_B.$$

The coefficients λ_i are called the Schmidt coefficients, and the decomposition is called the Schmidt decomposition.

Note, that different states $|\psi_{AB}\rangle$, in general, required a different choice of the bases. Also note, that if define unitary operators U_A and U_B which operate on \mathcal{H}_A and \mathcal{H}_B respectively then $U_{AB} \equiv U_A \otimes U_B$ when operating on $|\psi_{AB}\rangle$, changes the bases, but not the Schmidt coefficients λ_i .

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 (non-negative eigenvalues) Hermitian matrix H :

$$A = UH.$$

PROOF. We can clearly write A as

$$A = A \frac{1}{\sqrt{A^\dagger A}} \sqrt{A^\dagger A}.$$

Now, $A^\dagger A$ is a positive Hermitian matrix,¹⁹ so it may be written as

$$AA^\dagger = \sum_i \lambda_i |i\rangle\langle i|.$$

Thus, we may write

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

and indeed using this definition

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

Clearly, the inverse of $\sqrt{AA^\dagger}$ is

$$\frac{1}{\sqrt{AA^\dagger}} = \sum_i \frac{1}{\sqrt{\lambda_i}} |i\rangle\langle i|,$$

since

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

Now, we define

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

¹⁹It 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.

and

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

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

$$A = UH.$$

□

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

$$\begin{aligned} T^\dagger H T &= D \\ \Rightarrow H &= T D T^\dagger, \end{aligned}$$

therefore we can write

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

We now define

$$U \equiv U_1 T$$

and

$$V \equiv T^\dagger,$$

they are both clearly unitary, and we therefore have

$$A = UDV.$$

□

Having proved these lemmas, we may now prove the Schmidt decomposition. By definition state $|\Psi_{AB}\rangle$ can be written in general as²⁰

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

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 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

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

²⁰By definition

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

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

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

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

and

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

Using these definitions the 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. \square

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 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? If the λ_i are non-degenerate this is quite simple. If the Schmidt decomposition is

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

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

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

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

We see that the same λ_i appear in both density matrices when they are diagonalized. Further more, in the density matrix 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 value for the index i in all). Thus if we diagonalize each of the reduced density matrices we can match equally coefficients on the diagonally and deduce the Schmidt decomposition. if λ_i appears as the coefficient which includes that these are the coefficients of the appropriate.

If we now return to our previous example

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

then

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

where we define

$$|0\rangle_B = \sqrt{\frac{4+2\sqrt{2}}{3+2\sqrt{2}}} \left(|\uparrow_z\rangle_B + \frac{\sqrt{2}}{2+\sqrt{2}} |\downarrow_z\rangle_B \right),$$

$$|1\rangle_B = \sqrt{\frac{4-2\sqrt{2}}{3-2\sqrt{2}}} \left(|\uparrow_z\rangle_B - \frac{\sqrt{2}}{2-\sqrt{2}} |\downarrow_z\rangle_B \right).$$

For ρ_A we get

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

where we define

$$|0\rangle_A = \sqrt{\frac{12+8\sqrt{2}}{29+20\sqrt{2}}} \left(|\uparrow_z\rangle_A + \frac{3+2\sqrt{2}}{2+2\sqrt{2}} |\downarrow_z\rangle_B \right),$$

$$|1\rangle_A = \sqrt{\frac{12-8\sqrt{2}}{29-20\sqrt{2}}} \left(|\uparrow_z\rangle_A - \frac{3-2\sqrt{2}}{2-2\sqrt{2}} |\downarrow_z\rangle_B \right).$$

We can now finally write down the Schmidt decomposition

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

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, but we can still find using the method described in the proof of the Schmidt decomposition.

Part 2

Entanglement

CHAPTER 3

Hidden variables

3.1. The EPR¹ Paradox

Assume a two particle wave function of the form²

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

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

From the wave functions we know that

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

$$p_1 + p_2 \approx 0 \quad (\text{total momentum}),$$

thus, if Alice measures x_1 then she knows $x_2 \approx x_1 - L$ and if she measures p_1 then she knows $p_2 = -p_1$. We assume in all this that the distance L is large enough so that in the time it takes to make a measurement, light can't travel between the two particles, and particle 2 isn't effected by measurements on 1.

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, 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.

¹Einstein-Podolsky-Rosen

²Can also be described by squeezed states

$$\psi \approx \delta(x_1 - x_2 - L)\delta(p_1 + p_2) = \lim_{\alpha \rightarrow 1} \sum \alpha^n |n\rangle_1 |n\rangle_2.$$

3.2. 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 had found (1952) the pilot wave interpretation which was a non-local hidden variable theory. Bell has shown that above 2 dimensional Hilbert space one cannot have a local hidden variable theory (actually he showed an inequality which such a theory must obey, and quantum mechanically doesn't obey it experimentally - see below)

3.2.1. A hidden variables theory for spin $\frac{1}{2}$. Assume a system of spin $\frac{1}{2}$ in a state

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

We know that in quantum mechanics we have³

$$\langle\sigma_{\hat{n}}\rangle_{\psi_0} = \hat{n} \cdot \hat{z} = \cos\theta,$$

where

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

Can we produce a hidden variable theory, which will reproduce the same results?

Let us assume as a parameter a unit vector $\hat{\lambda}$ with equal probability to be on the upper half ($z > 0$) of a unit sphere (and zero probability to be on the lower part). We define the value $V_\sigma(\hat{n})$ of measuring $\sigma_{\hat{n}}$ as

$$V_\sigma(\hat{n}) \equiv \text{sign}(\hat{n} \cdot \hat{\lambda}) = \text{sign}(\cos\theta_{n\lambda}).$$

Since we assume that $\hat{\lambda}$ is (for some unknown reason) uniformly distributed on the upper half unit sphere, then in an area of $2\pi \cdot \frac{\theta_{n\lambda}}{\pi}$ we get negative values for $\hat{n} \cdot \hat{\lambda}$ and in the rest of the area upper half sphere $2\pi - 2\pi \cdot \frac{\theta_{n\lambda}}{\pi}$ 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

$$\langle V_\sigma(\hat{n}) \rangle = \frac{(-1) \cdot \theta_{n\lambda} + (+1) \cdot (2\pi - \theta_{n\lambda})}{2\pi} = 1 - \frac{2\theta_{n\lambda}}{\pi}.$$

This result however, doesn't give us the quantum result. To achieve that we use a different \hat{n} in $V_\sigma(\hat{n})$. we shall use \hat{n}' with an angle θ' with the z -axis (instead of z) such that

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

We can make a mapping of θ and θ' and therefore we got a desired hidden variable theory.

We could have constructed our hidden variable theory differently. We could have constructed a different model, with a parameter $0 \leq \lambda \leq 1$ such that

$$\sigma_{\hat{n}} = \begin{cases} 1 & 0 < \lambda < \cos^2 \frac{\theta}{2} \\ -1 & \cos^2 \frac{\theta}{2} < \lambda < 1 \end{cases} \\ \Rightarrow \langle\sigma_{\hat{n}}\rangle = \cos\theta,$$

where of course

$$\hat{n} \cdot \hat{z} = \cos\theta.$$

We would now like to see if it is possible to construct a hidden variable theory for a system of more than one spin, say a system of two entangled spins. We

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

shall see that we cannot build a *local* hidden variable theory, in this case (which is consistent with quantum mechanics).

Assume a system of two spins with total angular momentum zero. Therefore, if we measure spin 1 in the z direction and get “up”, then measuring spin 2, also in the z direction must give “down”. We need a model to give us this behavior. We can’t use Bell’s previous model, since the λ in each spin will be independent and we won’t get the desired result. We shall therefore try another model

$$V_{\hat{n}_i}(\sigma_{\hat{k}_i}) = \text{sign}(\hat{k}_i \cdot \hat{n}_i),$$

where i is an index of the particles. In our system

$$\hat{n}_1 = -\hat{n}_2$$

and we shall denote

$$\hat{a} = \hat{k}_1 \quad ; \quad \hat{b} = \hat{k}_2.$$

In other words, the directions of the spins now play the role of λ in the previous model. If we assume that \hat{n}_1 is random then we get, as in QM

$$\langle V_1(\hat{a}) \rangle_{\hat{n}_1} = \langle V_2(\hat{b}) \rangle_{\hat{n}_2 = -\hat{n}_1} = 0.$$

Now let us calculate $\langle V_1(\hat{a})V_2(\hat{b}) \rangle_{\hat{n}_1}$. Clearly, if $\hat{a} = \hat{b}$ we get the regular QM result

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

Now for general \hat{a} and \hat{b} , we can draw a sphere and on it two half-spheres, one around \hat{a} and the second around \hat{b} . For \hat{n}_1 which fall in the intersection of the two, or where non-cover the sphere we have

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

and in the rest we have

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

If we denote as θ the angle between \hat{a} and \hat{b} , and take the average over the areas we get

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

The result in QM is

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

Not surprisingly, we got different results. The question is can we correct our model as before, so that we get the correct answer. We shall see that no. The reason is that the parameter in the results is the angle between \hat{a} and \hat{b} . We can not make a deterministic change, separately on each function $V_i(\hat{k})$ (where we now only the direction of one measurement device, not both) so that we will get a correct result when combined.

3.2.2. The CHSH⁴ inequality. Let us assume that a hidden variables theory exists. We 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 \hat{a}' and particle 2 can be measured either in direction \hat{b} or \hat{b}' . Although we cannot measure both \hat{a} and \hat{a}' simultaneously (nor \hat{b} and \hat{b}'), since there are hidden variables, we can know in advance what the result would be, should we make the measurements. We shall use these “results”. We shall denote the result of

⁴Clauser Horne Shimony and Holt.

measuring $\sigma_{\hat{a}}$ by a the result of measuring $\sigma_{\hat{a}'}$ by a' and so on. Recall that each measurement can give only ± 1 , thus we may write that

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

since 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 don't know the occurrence distribution of $+2$ and of -2 as results in our hidden variables theory, we can conclude that we must have

$$|\langle (a + a')b + (a' - a)b' \rangle| \leq 2 \quad (\text{for hidden variables})$$

or

$$|\langle ab + a'b + a'b' - ab' \rangle| \leq 2 \quad (\text{for hidden variables})$$

writing this in standard QM form we can write

$$|\langle \sigma_{\hat{a}}\sigma_{\hat{b}} + \sigma_{\hat{a}'}\sigma_{\hat{b}} + \sigma_{\hat{a}'}\sigma_{\hat{b}'} - \sigma_{\hat{a}}\sigma_{\hat{b}'} \rangle| \leq 2 \quad (\text{for hidden variables})$$

or

$$|\langle \sigma_{\hat{a}}\sigma_{\hat{b}} \rangle + \langle \sigma_{\hat{a}'}\sigma_{\hat{b}} \rangle + \langle \sigma_{\hat{a}'}\sigma_{\hat{b}'} \rangle - \langle \sigma_{\hat{a}}\sigma_{\hat{b}'} \rangle| \leq 2 \quad (\text{for hidden variables}).$$

Each one of the four averages, in 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

$$\hat{a} \perp \hat{a}' \quad ; \quad \hat{b} \perp \hat{b}' \quad ; \quad \hat{a} \cdot \hat{b} = \cos \frac{\pi}{4}.$$

Since in QM $\langle \hat{\sigma}_{\hat{n}}^{(1)} \hat{\sigma}_{\hat{m}}^{(2)} \rangle = \hat{n} \cdot \hat{m}$ then (if we choose correctly)

$$\langle \sigma_{\hat{a}}\sigma_{\hat{b}} \rangle = \cos \frac{\pi}{4} = \frac{\sqrt{2}}{2}$$

$$\langle \sigma_{\hat{a}'}\sigma_{\hat{b}} \rangle = \cos \frac{\pi}{4} = \frac{\sqrt{2}}{2}$$

$$\langle \sigma_{\hat{a}}\sigma_{\hat{b}'} \rangle = \cos \frac{3\pi}{4} = -\frac{\sqrt{2}}{2}$$

$$\langle \sigma_{\hat{a}'}\sigma_{\hat{b}'} \rangle = \cos \frac{\pi}{4} = \frac{\sqrt{2}}{2}$$

and we get

$$|\langle \sigma_{\hat{a}}\sigma_{\hat{b}} \rangle + \langle \sigma_{\hat{a}'}\sigma_{\hat{b}} \rangle + \langle \sigma_{\hat{a}'}\sigma_{\hat{b}'} \rangle - \langle \sigma_{\hat{a}}\sigma_{\hat{b}'} \rangle| = 2\sqrt{2} \not\leq 2.$$

Experiments confirm that QM indeed holds in this case, and therefore QM cannot be a local hidden variables theory. Local here means that every particle has its own set of hidden variables which determine its behavior, regardless of what the others do (after 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.

3.2.3. Bell's inequalities. Bell was the first to give a proof that QM contradicts *local* hidden variables. 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 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 either measure a, b or measure a, c or b, c , but not all three, however since we assume hidden variables we can know the result of all three measurement in advance (even if we make only two of them). Since the spins are in opposite direction (and the measurements take values of ± 1) we can write

$$\pm a(b - c_2) = (1 + bc_1),$$

since if $b = c_2$ then $b = -c_1$ and both sides give zero, and if $b = -c_2$, then both sides give 2 up to a sign. Since the results has a \pm on the left and $\langle bc_1 \rangle > -1$, then taking the average on all possible hidden variables gives

$$|\langle ab \rangle - \langle ac_2 \rangle| \leq 1 + \langle bc_1 \rangle,$$

or using $c = c_1 = -c_2$

$$|\langle ab \rangle + \langle ac \rangle| \leq 1 + \langle bc \rangle,$$

Before going on we should prove the result for $\langle \sigma_{\hat{n}}^{(1)} \sigma_{\hat{m}}^{(2)} \rangle$ of two spins of opposite spin.

The Bell states we already encountered are also called *maximally entangled*. They are maximally entangled in the respect that they give the maximally violation of the Bell inequalities

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

3.3. Contextually

3.3.1. Definition of Non-Contextually. Non-contextually is a hidden variables “theory”. Which assumes:

Non-contextually: The result of a measurement is independent of whether other compatible (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-contextually can not be tested experimentally since one can not 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.

PROBLEM 3.3.1. Is the functional consistency assumption really new, or is more or less included in the non-contextual assumption?

3.3.1.1. *Mathematical formulation.* Non-contextual 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 $\{P_i\}$

$$\sum_i P_i = I \quad (P_i P_j = \delta_{ij} P_i, P_i^\dagger = P_i),$$

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-contextual 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. $t(P) \in [0, 1]$.

3.3.2. Contradicting Non-Contextually. Contradicting non-contextually is achieved, not by comparing it to experiments, but rather, by showing that it is logically inconstant (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

3.3.2.1. *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 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.

THEOREM. *The above axioms plus continuity of the vector space require that the expectation value of any Projector P must be of the form*

$$\langle P \rangle = \text{Tr}(\rho P) \quad \Rightarrow \quad \langle A \rangle = \text{Tr}(\rho A),$$

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-contextual) which returns discrete values (0 or 1), it is clear that the function $\langle P \rangle = \text{Tr}(\rho P)$ would return a continuous spectrum of values (if the projections P are continuous). This gives a contradiction and therefore non-contextually contradicts Gleason's axioms and thus also the standard axioms of QM.

PROBLEM 3.3.2. Is Gleason's theorem really a proof? In the book (Peres) he doesn't seem to treat it as a real proof.

3.3.2.2. *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*

$$\sum_i P_i = I \quad \text{and} \quad [P_i, P_j] = 0,$$

then

$$\sum_i t(P_i) = 1 \quad \text{where} \quad t(P_i) \in \{0, 1\}$$

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 rays⁵. 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.

⁵This 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	110 110	arbitrary choice of z axis
101 101 010		arbitrary choice of x vs. $-x$
011 011 100		arbitrary choice of y vs. $-y$
112 112 110	201 021	arbitrary choice of x vs. y
102 201 010	211	orthogonality to 2nd and 3rd rays
211 011 211	102	orthogonality to 2nd and 3rd rays
201 010 102	112	orthogonality to 2nd and 3rd rays
112 110 112	021	orthogonality to 2nd and 3rd rays
012 100 021	121	orthogonality to 2nd and 3rd rays
121 101 121	012	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 proof⁶ 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. \square

3.3.2.3. *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 operators⁷

$$\begin{array}{ccc} \mathbb{1} \otimes \sigma_z & \sigma_z \otimes \mathbb{1} & \sigma_z \otimes \sigma_z \\ \sigma_x \otimes \mathbb{1} & \mathbb{1} \otimes \sigma_x & \sigma_x \otimes \sigma_x \\ \sigma_x \otimes \sigma_z & \sigma_z \otimes \sigma_x & \sigma_y \otimes \sigma_y \end{array} .$$

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 columns, 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

⁶For 4 dimensions there is also a different proof using only 24 rays.

⁷Reminder: 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

$$\begin{aligned} \sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbb{1} \\ \sigma_x \sigma_y = i\sigma_z & ; \quad \sigma_y \sigma_x = -i\sigma_z & \Rightarrow [\sigma_x, \sigma_y] = 2i\sigma_z, \\ \sigma_z \sigma_x = i\sigma_y & ; \quad \sigma_x \sigma_z = -i\sigma_y & \Rightarrow [\sigma_z, \sigma_x] = 2i\sigma_y, \\ \sigma_y \sigma_z = i\sigma_x & ; \quad \sigma_z \sigma_y = -i\sigma_x & \Rightarrow [\sigma_y, \sigma_z] = 2i\sigma_x, \end{aligned}$$

which may be summarized by

$$\sigma_i \sigma_j = \delta_{ij} \mathbb{1} + i\epsilon_{ijk} \sigma_k.$$

$(-\sigma_y \otimes \sigma_y)$, then there is no possible choice of values ± 1 which will *not* lead to a contradiction. Mathematically, if 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

$$\begin{aligned} V(\sigma_y \otimes \sigma_y) &= V(\sigma_x \otimes \sigma_z)V(\sigma_z \otimes \sigma_x) = [V(\mathbf{1} \otimes \sigma_z)V(\sigma_x \otimes \mathbf{1})][V(\sigma_z \otimes \mathbf{1})V(\mathbf{1} \otimes \sigma_x)] \\ &= V(\mathbf{1} \otimes \sigma_z)V(\sigma_x \otimes \mathbf{1})V(\sigma_z \otimes \mathbf{1})V(\mathbf{1} \otimes \sigma_x). \end{aligned}$$

On the other hand, using functional consistency on the last column gives

$$\begin{aligned} V(\sigma_y \otimes \sigma_y) &= -V(\sigma_z \otimes \sigma_z)V(\sigma_x \otimes \sigma_x) = -[V(\mathbf{1} \otimes \sigma_z)V(\sigma_z \otimes \mathbf{1})][V(\sigma_x \otimes \mathbf{1})V(\mathbf{1} \otimes \sigma_x)] \\ &= -V(\mathbf{1} \otimes \sigma_z)V(\sigma_x \otimes \mathbf{1})V(\sigma_z \otimes \mathbf{1})V(\mathbf{1} \otimes \sigma_x). \end{aligned}$$

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-contextually are contradict quantum mechanics.

Uses of Entanglement

4.1. Encoding information

Recall the four Bell states (which are maximally entangled)

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

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 commute¹

$$[B_1, B_2] = 0.$$

These two operators have two eigenvalues each 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

¹They 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

$$\begin{aligned}[\sigma_x^A \sigma_x^B, \sigma_z^A \sigma_z^B] &= \sigma_x^A \sigma_x^B \sigma_z^A \sigma_z^B - \sigma_z^A \sigma_z^B \sigma_x^A \sigma_x^B \\ &= i^2 \sigma_y^A \sigma_y^B - (-i^2) \sigma_y^A \sigma_y^B = 0.\end{aligned}$$

particle A (and maybe other particles which belong to Alice) and Bob can perform any operation which operates only on particle B (and maybe 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 they 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 ψ^- . Otherwise the Bell state is either ϕ^+ 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.

4.2. Cryptography

Let's assume that Alice and Bob want to communicate (send messages between them), but that Eve wants to eavesdrop on 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.

$$K = 01100 \dots 1.$$

Now, in order to encrypt her message M e.g.

$$M = 01010 \dots,$$

²In 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.

³Using

$$\begin{aligned} |\uparrow_x\rangle &= \frac{1}{\sqrt{2}} (|\uparrow_z\rangle + |\downarrow_z\rangle) & |\uparrow_z\rangle &= \frac{1}{\sqrt{2}} (|\uparrow_x\rangle + |\downarrow_x\rangle) \\ |\downarrow_x\rangle &= \frac{1}{\sqrt{2}} (|\uparrow_z\rangle - |\downarrow_z\rangle) & |\downarrow_z\rangle &= \frac{1}{\sqrt{2}} (|\uparrow_x\rangle - |\downarrow_x\rangle) \end{aligned}$$

one finds that

$$\psi^- = \frac{1}{\sqrt{2}} (|\downarrow_x\rangle_A |\uparrow_x\rangle_B - |\uparrow_x\rangle_A |\downarrow_x\rangle_B),$$

$$\psi^+ = \frac{1}{\sqrt{2}} (|\uparrow_x\rangle_A |\uparrow_x\rangle_B - |\downarrow_x\rangle_A |\downarrow_x\rangle_B),$$

$$\phi^- = \frac{1}{\sqrt{2}} (|\uparrow_x\rangle_A |\downarrow_x\rangle_B + |\downarrow_x\rangle_A |\uparrow_x\rangle_B),$$

$$\phi^+ = \frac{1}{\sqrt{2}} (|\uparrow_x\rangle_A |\uparrow_x\rangle_B + |\downarrow_x\rangle_A |\downarrow_x\rangle_B).$$

Alice performs a xor⁴ operation on her message and key K and generates a new message M' , e.g.

$$M' = M \oplus K = \frac{\begin{array}{cccccc} 0 & 1 & 0 & 1 & 0 & \dots \\ 0 & 1 & 1 & 0 & 0 & \dots \\ \hline 0 & 0 & 1 & 1 & 0 & \dots \end{array}}{.}$$

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

$$\psi^- = \frac{1}{\sqrt{2}} (|0\rangle_A |1\rangle_B - |1\rangle_A |0\rangle_B),$$

and each time 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 be 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 without 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

$$|\psi_0\rangle_{ABE} = \frac{1}{\sqrt{2}} (|\uparrow\rangle_A |\downarrow\rangle_B - |\downarrow\rangle_A |\uparrow\rangle_B) |\uparrow\rangle_E.$$

We are now looking for a unitary operator such that

$$|\psi_0\rangle_{ABE} \xrightarrow{U_{AE}} \frac{1}{\sqrt{2}} [(|\uparrow\rangle_A |\uparrow\rangle_E) |\downarrow\rangle_B - (|\downarrow\rangle_A |\downarrow\rangle_E) |\uparrow\rangle_B].$$

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 as⁵

$$\begin{aligned} U_{AE} &= U_{\text{CNOT}} = |\uparrow\rangle_{AA} \langle \uparrow| \otimes \mathbb{1}_E + |\downarrow\rangle_{AA} \langle \downarrow| \otimes \sigma_x^E \\ &= \frac{1}{2} (\mathbb{1}_A + \sigma_z^A) \otimes \mathbb{1}_E + \frac{1}{2} (\mathbb{1}_A - \sigma_z^A) \otimes \sigma_x^E \end{aligned}$$

⁴The xor operation is denoted by \oplus and is addition modulo 2 i.e.

$$0 \oplus 0 = 0$$

$$0 \oplus 1 = 1 \oplus 0 = 1$$

$$1 \oplus 1 = 0.$$

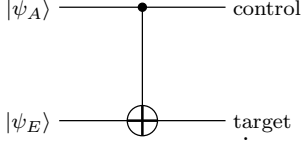
⁵Recall that

$$\sigma_x |\uparrow_z\rangle = |\downarrow_z\rangle \quad ; \quad \sigma_x |\downarrow_z\rangle = |\uparrow_z\rangle.$$

or in matrix form

$$U_{\text{CNOT}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ & & 0 & 1 \\ & & 1 & 0 \end{pmatrix}.$$

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 $\sigma_k^2 = \mathbb{1}$ then

$$(\mathbb{1} - \sigma_k)^2 = 2(\mathbb{1} - \sigma_k) \Rightarrow (\mathbb{1} - \sigma_k)^n = 2^{n-1}(\mathbb{1} - \sigma_k) \quad (n \neq 0).$$

Therefore we can write⁶

$$U_{\text{CNOT}} = e^{-i\frac{\pi}{4}(\mathbb{1}_A - \sigma_z^A)(\mathbb{1}_E - \sigma_x^E)}.$$

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.⁷ 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

⁶Since 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

$$U_{\text{CNOT}} = e^{-i\frac{\pi}{4}\sigma_z^A\sigma_x^E} \quad (\text{up to local operations and a phase}).$$

⁷We 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 infer from this any change).

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 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 they 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

$$|\uparrow_z\rangle, |\downarrow_z\rangle, |\uparrow_x\rangle, |\downarrow_x\rangle.$$

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).

4.3. teleportation

Assume that Alice and Bob have an entangled state $|\phi^+\rangle_{ab}$ between them

$$|\phi^+\rangle_{ab} = \frac{1}{\sqrt{2}} (|0\rangle_a |0\rangle_b + |1\rangle_a |1\rangle_b).$$

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.⁸ The state of the whole system (all three particles) is $|\psi\rangle_A |\phi^+\rangle_{ab}$, however it can also be written as⁹

$$\begin{aligned} |\psi\rangle_A |\phi^+\rangle_{ab} &= (\alpha|0\rangle_A + \beta|1\rangle_A) \frac{1}{\sqrt{2}} (|0\rangle_a |0\rangle_b + |1\rangle_a |1\rangle_b) \\ &= \frac{1}{\sqrt{2}} (\alpha|0\rangle_A |0\rangle_a |0\rangle_b + \alpha|0\rangle_A |1\rangle_a |1\rangle_b + \beta|1\rangle_A |0\rangle_a |0\rangle_b + \beta|1\rangle_A |1\rangle_a |1\rangle_b) \\ &= \frac{|\phi^+\rangle_{Aa} + |\phi^-\rangle_{Aa}}{2} \alpha|0\rangle_b + \frac{|\psi^+\rangle_{Aa} + |\psi^-\rangle_{Aa}}{2} \alpha|1\rangle_b \\ &\quad + \frac{|\psi^+\rangle_{Aa} - |\psi^-\rangle_{Aa}}{2} \beta|0\rangle_b + \frac{|\phi^+\rangle_{Aa} - |\phi^-\rangle_{Aa}}{2} \beta|1\rangle_b \\ &= \frac{1}{2} |\phi^+\rangle_{Aa} (\alpha|0\rangle_b + \beta|1\rangle_b) + \frac{1}{2} |\psi^+\rangle_{Aa} \sigma_x^b (\alpha|0\rangle_b + \beta|1\rangle_b) \\ &\quad + \frac{1}{2} |\psi^-\rangle_{Aa} (-i)\sigma_y^b (\alpha|0\rangle_b + \beta|1\rangle_b) + \frac{1}{2} |\phi^-\rangle_{Aa} \sigma_z^b (\alpha|0\rangle_b + \beta|1\rangle_b) \\ &= \frac{1}{2} [|\phi^+\rangle_{Aa} |\psi\rangle_b + |\psi^+\rangle_{Aa} \sigma_x^b |\psi\rangle_b + |\psi^-\rangle_{Aa} (-i)\sigma_y^b |\psi\rangle_b + |\phi^-\rangle_{Aa} \sigma_z^b |\psi\rangle_b]. \end{aligned}$$

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 $\sigma_i^b |\psi\rangle_b$

⁸We may assume that particles a, b, A are of different types or of similar types - there is no restriction.

⁹Recall that

$$\begin{aligned} |\psi^-\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}} (|0\rangle_A |1\rangle_a - |1\rangle_A |0\rangle_a), \\ |\psi^+\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}} (|0\rangle_A |1\rangle_a + |1\rangle_A |0\rangle_a), \\ |\phi^-\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_a - |1\rangle_A |1\rangle_a), \\ |\phi^+\rangle_{Aa} &\equiv \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_a + |1\rangle_A |1\rangle_a). \end{aligned}$$

($i = 0, 1, 2, 3$ where $\sigma_0 = \mathbb{1}$).¹⁰ 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, 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.

4.4. 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 $|\phi^+\rangle$

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

Bennet 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

$$1\text{ebit}_{AB} + 2\text{bit}_{A \rightarrow B} \Rightarrow 1\text{qbit}_{A \rightarrow B}.$$

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

$$\text{teleportation}_{A \rightarrow B} + \text{ebit}_{AB} \Rightarrow 2\text{bit}_{A \rightarrow B}.$$

¹⁰OK, for $i = 2$ the state is $i\sigma_y^b|\psi\rangle_b$ not $\sigma_y^b|\psi\rangle_b$.

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

$$\text{CNOT} + 1 \text{ bit}_{A \rightarrow B} \Rightarrow \text{teleport}_{A \rightarrow B},$$

$$\text{teleport}_{A \rightarrow B} + 1 \text{ bit}_{A \rightarrow B} \Rightarrow \text{CNOT}.$$

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). \square

4.5. State-operators (stators)

We define a *state-operator*, or for short *stator*, as “creature” which is a combination of states in one Hilbert space and operators in another. Generally speaking it will be written as

$$S = \sum c_i |i\rangle_A \otimes O_i^B,$$

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 on a state in \mathcal{H}_B the result is a state in $\mathcal{H}_A \otimes \mathcal{H}_B$

$$S|\psi\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B \quad (|\psi\rangle_B \in \mathcal{H}_B).$$

We will be interested in pairs of operators which obey

$$AS = BS,$$

where A operates on \mathcal{H}_A and B operates \mathcal{H}_B . Such pairs do not necessarily consist of two Hermitian operators, but we will be interested in the cases where they are. For example if

$$S = |0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \sigma_z^B,$$

then

$$\sigma_x^A S = \sigma_z^B S.$$

If a pair of operators A, B obey the relation

$$AS = BS,$$

then necessarily

$$A^n S = B^n S,$$

and therefore (using a Taylor expansion)

$$f(A)S = f(B)S.$$

Specifically, if A, B are Hermitian then

$$e^{i\alpha A} S = e^{i\beta B} S,$$

where by definition $e^{i\alpha A}$ and $e^{i\beta B}$ are unitary operators.

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 , to perform $e^{i\alpha\sigma_z^B}$ on his particle. To do this we use our previous stator

$$S = |0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \sigma_z^B.$$

We start by Alice performing her operator on the stator S , i.e. performing

$$U_\alpha S = e^{i\alpha\sigma_x^A} S.$$

For our stator 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

$$U_\alpha S |\psi\rangle_B = S e^{i\alpha\sigma_z^B} |\psi\rangle_B = (|0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes \sigma_z^B) e^{i\alpha\sigma_z^B} |\psi\rangle_B.$$

Alice now measures her spin and sends the result to Bob. If Alice found 0 then Bob does nothing, however if she found 1, then Bob must fix his state (get rid of the extra σ_z^B operator) by performing σ_z^B on it.

We can generalize the last operation, in order to achieve a remote CNOT (up to local operations and a phase)

$$U_{\text{CNOT}} = e^{-i\frac{\pi}{4}\sigma_z^A \sigma_x^B} \quad (\text{up to local operations and a phase}).$$

To do this we generalize our stator to

$$S = |\uparrow_x\rangle_a \otimes \mathbb{1}_{AB} + |\downarrow_x\rangle_a \otimes \sigma_z^A \sigma_x^B,$$

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

$$\begin{aligned} e^{-i\frac{\pi}{4}\sigma_x^a} S |\varphi\rangle_A |\psi\rangle_B &= S e^{-i\frac{\pi}{4}\sigma_z^A \sigma_x^B} |\varphi\rangle_A |\psi\rangle_B \\ &= (|\uparrow_x\rangle_a \otimes \mathbb{1}_{AB} + |\downarrow_x\rangle_a \otimes \sigma_z^A \sigma_x^B) e^{-i\frac{\pi}{4}\sigma_z^A \sigma_x^B} |\varphi\rangle_A |\psi\rangle_B. \end{aligned}$$

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 that we were able to perform the operation $e^{-i\frac{\pi}{4}\sigma_z^A \sigma_x^B}$ on the remote particles A, B .

4.5.1. Creating a stator. Assume a general, two-level system, unitary operator U . We wish to construct a stator of the form¹¹

$$S = |0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes U^B.$$

We start with three particles, an ancilla b and the two particles A, B . We start with the configuration

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A |0\rangle_b + |1\rangle_A |1\rangle_b) |\psi\rangle_B,$$

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

$$\begin{aligned} \frac{1}{\sqrt{2}}(|0\rangle_A |0\rangle_b + |1\rangle_A |1\rangle_b U^B) |\psi\rangle_B &= \frac{1}{2}(|\uparrow_x\rangle_b + |\downarrow_x\rangle_b) |0\rangle_A |\psi\rangle_B + \frac{1}{2}(|\uparrow_x\rangle_b - |\downarrow_x\rangle_b) |1\rangle_A U^B |\psi\rangle_B \\ &= \frac{1}{2} |\uparrow_x\rangle_b (|0\rangle_A + |1\rangle_A U^B) |\psi\rangle_B + \frac{1}{2} |\downarrow_x\rangle_b (|0\rangle_A - |1\rangle_A U^B) |\psi\rangle_B. \end{aligned}$$

Bob, now measures the spin of his ancilla b in the x direction. Thus collapsing the system into one of the states

$$\begin{aligned} \frac{1}{\sqrt{2}} |\uparrow_x\rangle_b (|0\rangle_A + |1\rangle_A U^B) |\psi\rangle_B \quad \sigma_x^b = \text{”up”}, \\ \frac{1}{\sqrt{2}} |\downarrow_x\rangle_b (|0\rangle_A - |1\rangle_A U^B) |\psi\rangle_B \quad \sigma_x^b = \text{”down”}. \end{aligned}$$

Now, Bob sends the result of his measurement to Alice, who accordingly decides, whether to perform a σ_z^A on her system (if $\sigma_x^b = \text{”down”}$) or if to do nothing ($\sigma_x^b = \text{”up”}$). 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 .

¹¹To create the more general stator

$$S = |0\rangle_A \otimes U_{B0} + |1\rangle_A \otimes U_{B1}$$

we simply need

$$S = S' U_{B0} = \left[|0\rangle_A \otimes \mathbb{1}_B + |1\rangle_A \otimes (U_{B1} U_{B0}^{-1}) \right] U_{B0}.$$

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).

4.6. 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

$$\begin{aligned} U|\Psi\rangle_{\text{tot}} &= \mathbb{1}_a U|\Psi\rangle_{\text{tot}} = \left(\sum_{\mu} |\mu\rangle_{aa} \langle\mu| \right) U|0\rangle_a |\psi\rangle_{\text{sys}} \\ &= \sum_{\mu} ({}_a\langle\mu|U|0\rangle_a) |\mu\rangle_a |\psi\rangle_{\text{sys}}. \end{aligned}$$

If we now define the *Kraus operator*

$$M_{\mu} \equiv {}_a\langle\mu|U|0\rangle_a,$$

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 $\text{prob}(\mu)$

$$\begin{aligned} \text{prob}(\mu) &= {}_{\text{tot}}\langle\Psi|U^\dagger|\mu\rangle_a \langle\mu|U|\Psi\rangle_{\text{tot}} \\ &= {}_{\text{sys}}\langle\psi|M_\mu^\dagger M_\mu|\psi\rangle_{\text{sys}}. \end{aligned}$$

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 $|\psi\rangle_{\text{sys}}$) simply¹²

$$\sum_{\mu} M_\mu^\dagger M_\mu = \mathbb{1}_{\text{sys}}.$$

In analogy to the Von Neumann measurements, we may now write, for measurements using an ancilla

$$|\psi\rangle_{\text{sys}} \xrightarrow{\text{measured } \mu} M_\mu |\psi\rangle_{\text{sys}} \quad (\text{not normalized}),$$

$$\rho \xrightarrow{\text{measured } \mu} \frac{M_\mu \rho M_\mu^\dagger}{\text{Tr}(M_\mu \rho M_\mu^\dagger)},$$

$$F_\mu \equiv M_\mu^\dagger M_\mu \text{ a positive operator} \quad \left(\sum_{\mu} F_\mu = \mathbb{1}_{\text{sys}}\right),$$

$$\text{prob}(\mu) = \text{Tr}(F_\mu \rho_{\text{sys}}),$$

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).

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 Krause operators M_μ replace the projections Π_i . The only difference is that, here, the Krause 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.

¹²This could also be found directly from the definition of the M_μ

$$\sum_{\mu} M_\mu^\dagger M_\mu = \sum_{\mu} {}_a\langle 0|U^\dagger|\mu\rangle_a \langle\mu|U|0\rangle_a = {}_a\langle 0|U^\dagger U|0\rangle_a = \mathbb{1}_{\text{sys}}.$$

4.6.1. 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 , $\mu = 1, \dots, N$), then we can always realize it as standard measurements in an N dimensional Hilbert space. This theorem is known as *Neumark's theorem*.

4.6.2. 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.¹³ In other words we do not know the answer for certain, in $\frac{3}{4}$ of the measurements.

Now, instead of making standard measurements, let us define

$$\begin{aligned} F_1 &= \lambda |\downarrow_z\rangle\langle\downarrow_z|, \\ F_2 &= \lambda |\downarrow_x\rangle\langle\downarrow_x|, \\ F_3 &= \mathbb{1}_{\text{sys}} - F_1 - F_2. \end{aligned}$$

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 $\mu = 3$, 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 simply¹⁴

$$\text{prob}(\mu = 3) = 1 - \left(\frac{1}{2} \frac{\lambda}{2} + \frac{1}{2} \frac{\lambda}{2} \right) = 1 - \frac{\lambda}{2},$$

¹³We 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 time 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} .

¹⁴There is a probability $\frac{1}{2}$ of state $|\psi_1\rangle$ occurring and a probability $\frac{1}{2}$ of state $|\psi_2\rangle$. 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 with probability $\frac{1}{2}$, the probability of getting $\mu \neq 1, 2$ (i.e. getting $\mu = 3$) is

$$\text{prob}(\mu = 3) = 1 - \left(\frac{1}{2} \frac{\lambda}{2} + \frac{1}{2} \frac{\lambda}{2} \right).$$

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).¹⁵ 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.

4.7. Measure of entanglement (Distillation)

Let us assume that we have a system in a state

$$|\Psi\rangle_{ab} = \alpha|0\rangle_a|0\rangle_b + \beta|1\rangle_a|1\rangle_b \quad (|\alpha| \leq |\beta|),$$

where we know α, β and we assume $|\alpha| \leq |\beta|$. 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

$$|\phi^+\rangle_{ab} = \frac{1}{\sqrt{2}}(|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b).$$

We want to do this using only local operations. To do this we shall use the *Procrustean method*. If we define the Krause operator¹⁶

$$M_0 \equiv \lambda \left(\frac{\beta}{\alpha} |0\rangle_{aa}\langle 0| + |1\rangle_{aa}\langle 1| \right),$$

then whenever we measure $\mu = 0$, the state we will find is

$$|\Psi\rangle_{ab} \xrightarrow{\mu=0} \frac{M_0|\Psi\rangle_{ab}}{\sqrt{{}_{ab}\langle\Psi|M_0^\dagger M_0|\Psi\rangle_{ab}}} = \frac{\lambda}{\sqrt{{}_{ab}\langle\Psi|M_0^\dagger M_0|\Psi\rangle_{ab}}} \beta (|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b),$$

where clearly (because of the normalization) we will find that

$$\frac{\lambda}{\sqrt{{}_{ab}\langle\Psi|M_0^\dagger M_0|\Psi\rangle_{ab}}} \beta = \frac{1}{\sqrt{2}}.$$

¹⁵Actually 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.

¹⁶Note, 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.

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

$$\begin{aligned} \text{prob}(\mu = 0) &= {}_{ab}\langle\Psi|M_0^\dagger M_0|\Psi\rangle_{ab} = |\lambda\beta|^2 ({}_b\langle 1|_a\langle 1| + {}_b\langle 0|_a\langle 0|) (|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b) \\ &= 2|\lambda\beta|^2. \end{aligned}$$

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 have¹⁷

$$\sum_{\mu} M_{\mu}^\dagger M_{\mu} = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the $M_{\mu}^\dagger M_{\mu}$ are all positive operators then necessarily, we must have

$$\begin{aligned} |\lambda|^2 \left|\frac{\beta}{\alpha}\right|^2 &\leq 1 \\ \Rightarrow |\lambda\beta|^2 &\leq |\alpha|^2. \end{aligned}$$

Using, this last result, we see that the maximum probability possible for distillation (recall that $|a| \leq |\beta|$) is¹⁸

$$\text{prob}(\mu = 0) \leq 2|\alpha|^2 \leq 1.$$

$$M_1 \equiv U \left(\mathbb{1} - \sqrt{M_0^\dagger M_0} \right),$$

$$M_2 \equiv \mathbb{1} - M_1 - M_0,$$

where U is any arbitrary unitary operator.

4.7.1. 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

$$|\Psi\rangle^{\otimes 2} \xrightarrow{\sigma_T=2} |0\rangle_a|0\rangle_{a'}|0\rangle_b|0\rangle_{b'},$$

¹⁷We could also define the “complementary” M_1 of M_0 by

$$\begin{aligned} \mathbb{1} &= M_0^\dagger M_0 + M_1^\dagger M_1 \\ \Rightarrow M_1 &= U \sqrt{\mathbb{1} - M_0^\dagger M_0}. \end{aligned}$$

Requiring that it be a positive operator, would give us the same result.

¹⁸The 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.

$$\begin{aligned}
|\Psi\rangle^{\otimes 2} &\xrightarrow{\sigma_T=-2} |1\rangle_a|1\rangle_{a'}|1\rangle_b|1\rangle_{b'}, \\
|\Psi\rangle^{\otimes 2} &\xrightarrow{\sigma_T=0} \frac{1}{\sqrt{2}} (|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|0\rangle_{a'}|1\rangle_b|0\rangle_{b'}),
\end{aligned}$$

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

$$\begin{aligned}
|\Psi\rangle^{\otimes 2} &\xrightarrow{\sigma_T=0} \frac{1}{\sqrt{2}} (|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|0\rangle_{a'}|1\rangle_b|0\rangle_{b'}) \\
&\xrightarrow{\text{CNOT}_{b,b'}^{a,a'}} \frac{1}{\sqrt{2}} (|0\rangle_a|1\rangle_{a'}|0\rangle_b|1\rangle_{b'} + |1\rangle_a|1\rangle_{a'}|1\rangle_b|1\rangle_{b'}) \\
&= \frac{1}{\sqrt{2}} |1\rangle_{a'}|1\rangle_{b'} (|0\rangle_a|0\rangle_b + |1\rangle_a|1\rangle_b).
\end{aligned}$$

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

$$\begin{aligned}
|\Psi\rangle^{\otimes n} &= (\alpha|0\rangle_{a_i}|0\rangle_{b_i} + \beta|1\rangle_{a_i}|1\rangle_{b_i})^{\otimes n} \\
&= \alpha^n \prod_{i=1}^n |0\rangle_{a_i}|0\rangle_{b_i} + \alpha^{n-1}\beta \sum_{j=1}^n \left(|1\rangle_{a_i}|1\rangle_{b_i} \prod_{i \neq j} |0\rangle_{a_j}|0\rangle_{b_j} \right) + \dots,
\end{aligned}$$

that is we have a tensor product of n pairs numbered $i = 1, \dots, n$. 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 binom 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)

$$\text{prob}(\sigma_T^a = 2m - n) = \binom{n}{m} |\alpha^m \beta^{n-m}|^2 = \frac{n!}{m!(n-m)!} |\alpha^{2m} \beta^{2(n-m)}|.$$

If we now examine $n \rightarrow \infty$ the probability will be maximal, and approach a delta function at $m = |\alpha|^2 n + O(\sqrt{n})$.¹⁹ 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 in Alice and Bob’s particles, therefore we can make a change of base²⁰ so that that the system have the form (not normalized) for a given m

$$|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B + \dots + \binom{n}{m} | \rangle_A | \binom{n}{m} \rangle_B,$$

¹⁹This is true regardless of the values of α, β (as long as none of them is zero).

²⁰It is enough to make a change of names. We simply number all the permutations of m elements out of n and then call the j th permutation $|j\rangle_A$. We do the same for $|j\rangle_B$, and we automatically get

$$|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B + \dots + \binom{n}{m} | \rangle_A | \binom{n}{m} \rangle_B.$$

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 like 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

$$\begin{aligned} nH(m = |\alpha|^2 n) &\approx n \log n - n|\alpha|^2 \log_2(n|\alpha|^2) - n|\beta|^2 \log_2(n|\beta|^2) \\ &= n [\log_2 n - (|\alpha|^2 + |\beta|^2) \log_2 n - |\alpha|^2 \log_2 |\alpha|^2 - |\beta|^2 \log_2 |\beta|^2] \\ &= -n [|\alpha|^2 \log_2 |\alpha|^2 + |\beta|^2 \log_2 |\beta|^2]. \end{aligned}$$

If we define

$$\begin{aligned} p &\equiv |\alpha|^2 \\ &\Rightarrow (1 - p) = |\beta|^2, \end{aligned}$$

Then we may write

$$H = -[p \log_2 p + (1 - p) \log_2 p].$$

To conclude we saw that if we use the scheme of measuring σ_z^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

$$\begin{aligned} nH &= -n [|\alpha|^2 \log_2 |\alpha|^2 + |\beta|^2 \log_2 |\beta|^2] \\ &= -n [p \log_2 p + (1 - p) \log_2 p]. \end{aligned}$$

Or simply

$$n \text{ non-maximally entangled} \rightarrow nH \text{ maximally entangled.}$$

The ratio of maximally entangled pairs, out of the original number of pairs is

$$E(\psi) \equiv \frac{\text{maximally entangled pairs}}{\text{non-maximally entangled pairs}} = H = -[|\alpha|^2 \log_2 |\alpha|^2 + |\beta|^2 \log_2 |\beta|^2] + O\left(\frac{1}{\sqrt{n}}\right),$$

$$\psi \equiv \alpha|0\rangle + \beta|1\rangle \quad (\text{single non-maximally entangled particle})$$

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.²¹ 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

$$\begin{aligned} H(\psi_1 \otimes \psi_2) &= - \sum_{i,j} p_i q_j \log_2(p_i q_j) = - \sum_{i,j} p_i q_j (\log_2 p_i + \log_2 q_j) \\ &= - \sum_i p_i \sum_j q_j \log_2 q_j - \sum_j p_j \sum_i q_i \log_2 p_i = - \sum_j q_j \log_2 q_j - \sum_i p_i \log_2 p_i \\ &= H(\psi_1) + H(\psi_2). \end{aligned}$$

Since the Shannon entry 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*.

²¹See also Von Neumann entropy S and entanglement measure.

Quantum information

5.1. 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 than 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

$$X_k = \{a_x, p_x\}_{x=1}^k \quad \left(\sum_{x=1}^k p_x = 1\right).$$

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 therefore¹

$$\frac{n!}{\prod_x (np_x)!}.$$

Using the definition of the *Shannon entropy* we can therefore write (using the Stirling approximation)²

$$\#\text{typical messages} \approx \frac{n!}{\prod_x (np_x)!} \approx 2^{nH}.$$

Thus to encode the the different *typical* messages, we can simply number them $1, 2, \dots, 2^{nH}$ 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

$$\begin{array}{l} n \text{ letters} \\ k \text{ letter alphabet} \end{array} \xrightarrow{\text{compression}} nH \text{ bits} \quad \left(H = -\sum_x p_x \log_2 p_x\right).$$

This of course holds only for the typical messages, whose weight in the overall ensemble of messages increases as $n \rightarrow \infty$.

Another way of reaching the same conclusion, is to examine a single message of length n

$$\text{message} = (x_1, x_2, \dots, x_n)$$

where in the i th position the letter a_{x_i} appears. The probability of such a message occurring is

$$\text{prob}(\text{message}) = \text{prob}(x_1, x_2, \dots, x_n) = p_{x_1} p_{x_2} \cdots p_{x_n}$$

¹We use here the same logic as was used above, in determining a measure for entanglement.

²Recall that

$$\log(n!) \approx n \log n.$$

or

$$\log_2 [\text{prob}(x_1, x_2, \dots, x_n)] = \log_2(p_{x_1} p_{x_2} \cdots p_{x_n}) = \sum_i \log_2 p_{x_i}.$$

By the central limit theorem, for $n \rightarrow \infty$ we have

$$-\frac{1}{n} \log_2 [\text{prob}(x_1, x_2, \dots, x_n)] \sim -\langle \log_2 p \rangle \equiv H,$$

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 than $1 - \varepsilon$, then we can put a bound on the number $N_{\varepsilon, \delta}$ of "typical" messages⁴

$$(1 - \varepsilon)2^{n(H-\delta)} \leq N_{\varepsilon, \delta} \leq 2^{n(H+\delta)}.$$

Thus we see that in the limit of $n \rightarrow \infty$ 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.

5.2. 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 replaced 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 ($\rho = \sum p_i |\psi_i\rangle \langle \psi_i|$)

³Each typical message may have a slightly different probability to occur, but all these different probabilities obey the inequality given.

⁴The bound comes from

$$NP_{\max} > 1 - \varepsilon$$

$$NP_{\min} < 1$$

⁵When all letters are likely to appear then

$$H = - \sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = - \log_2 \frac{1}{n} = \log_2 n.$$

where the $|\psi_i\rangle$ are not necessarily orthogonal)⁶. In such a case one cannot perfectly distinguish the different states (the different letters).

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 after sending and decompressing it the state is $|\varphi_f\rangle$ 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 , encoded as $|\varphi\rangle$ the fidelity is defined as

$$F \equiv \langle\varphi|\rho|\varphi\rangle = \text{Tr}(|\varphi\rangle\langle\varphi|\rho),$$

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

$$\begin{aligned} |\uparrow_{\hat{n}}\rangle &= \cos\frac{\pi}{8}|\uparrow_z\rangle + \sin\frac{\pi}{8}|\downarrow_z\rangle, \\ |\downarrow_{\hat{n}}\rangle &= \sin\frac{\pi}{8}|\uparrow_z\rangle - \cos\frac{\pi}{8}|\downarrow_z\rangle. \end{aligned}$$

It can easily be shown that the maximum fidelity is reached when

$$|\varphi\rangle = |\uparrow_{\hat{n}}\rangle$$

which gives

$$F = \langle\varphi|\rho|\varphi\rangle = \cos^2\frac{\pi}{8} = 0.853\dots$$

Now let us assume that Alice has a message consisting 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 may send only 2 particles (qubits) to him. We might think that the best state for Bob to receive is simply

$$|\varphi\rangle = \rho \otimes \rho \otimes |\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}|,$$

where the density matrices ρ stand for the original particles sent from Alice, and $|\uparrow_{\hat{n}}\rangle$ is the same state we used above for maximum fidelity (the $|\uparrow_{\hat{n}}\rangle$ state is not sent from Alice to Bob. It is prepared locally by Bob based on his prior knowledge of ρ). 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

$$F = 1 \cdot 1 \cdot \cos^2\frac{\pi}{8} = 0.853\dots$$

However, Alice can increase the fidelity of her message. Since she is sending only two qubits, the Hilbert space described by her new state belongs to a subspace

⁶We 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.

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 possible 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

$$|\langle\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = \cos^6 \frac{\pi}{8} = 0.62$$

$$|\langle\uparrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = |\langle\uparrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = |\langle\downarrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = \cos^4 \frac{\pi}{8} \sin^2 \frac{\pi}{8} = 0.107$$

$$|\langle\uparrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = |\langle\downarrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = |\langle\downarrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}|\psi\rangle|^2 = \cos^2 \frac{\pi}{8} \sin^4 \frac{\pi}{8} = 0.018$$

$$|\langle\downarrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}|\psi\rangle|^2 = \sin^6 \pi = 0.003$$

Since the dimension of the Hilbert subspace we can span using just 2 qubits is 4, we are looking for the most probable 4-dimensional subspace spanned by 4 of the above states. This subspace, \mathcal{H}_1 , is just the span of the 4 most probable states, namely

$$\mathcal{H}_1 = \text{span}\{|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\uparrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\uparrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle\} \subset \mathcal{H}.$$

and the remaining subspace is

$$\mathcal{H}_2 = \text{span}\{|\downarrow_{\hat{n}}\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle, |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle, |\uparrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}\rangle, |\downarrow_{\hat{n}}\downarrow_{\hat{n}}\downarrow_{\hat{n}}\rangle\} \subset \mathcal{H},$$

and naturally we have⁷

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2.$$

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

$$U : \mathcal{H}_1 \rightarrow |\cdot\rangle|\cdot\rangle|0\rangle,$$

and

$$U : \mathcal{H}_2 \rightarrow |\cdot\rangle|\cdot\rangle|1\rangle.$$

Having done that, Alice measures the last particle. This measurement has the effect of projection 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, she sends the first two qubits to Bob, Bob adds a third qubit in state $|0\rangle$ and constructs the final decoded message by performing U^{-1} (the same U that Alice used). If, however, Alice measured 1, she sends the two qubits in a predetermined state such that after Bob decodes (by the same method as before) he gets $|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle$ (that is, she sends $U|\uparrow_{\hat{n}}\uparrow_{\hat{n}}\uparrow_{\hat{n}}\rangle$). Note that this 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 the received message can now be found, by comparing the message

⁷Note that we use addition (\oplus) of spaces, and not a tensor product. This is because we are dealing with subspaces.

Bob decodes and the original one. Since, however, Alice may send different messages, depending on the result of her measurement, we regard the average fidelity⁸

$$F = 0.94 \left| \left\langle \psi \left| \frac{\Pi_1 |\psi\rangle}{\sqrt{\langle \psi | \Pi_1 | \psi \rangle}} \right. \right\rangle \right|^2 + 0.06 |\langle \psi | \uparrow_{\hat{n}} \uparrow_{\hat{n}} \uparrow_{\hat{n}} \rangle|^2 = 0.94^2 + 0.06 \cdot 0.62 = 0.92.$$

We see, that Although in our first example Alice had sent two of the three qubits accurately, and in this example, none of them is sent with total accuracy, the total fidelity in the latter case is higher.

5.3. Shumacher'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 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

$$S(\rho) \equiv -\text{Tr}(\rho \log_2 \rho),$$

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

$$\dim \mathcal{H}_{\text{prob}} = 2^{nS(\rho)}.$$

⁸Again note, that due to the special symmetry between the two possible states of the original system $|\uparrow_x\rangle, |\downarrow_x\rangle$ (both with probability $\frac{1}{2}$), we do not treat here the cases differently, however for a general case we would have to compute the fidelity for each possible state on Alice's side, then averaging over these cases with the appropriate probabilities for these states.

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. Furthermore 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.

5.3.0.1. *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 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

$$\rho_{AB} = \sum_{k,l} \sqrt{p_k p_l} |k\rangle_A |l\rangle_B \langle l|_A \langle k|.$$

If we take a partial trace of ρ_{AB} over A or B we get

$$\rho_A \equiv \text{Tr}_B \rho_{AB} = \sum_{k=1}^n p_k |k\rangle_{AA} \langle k|,$$

$$\rho_B \equiv \text{Tr}_A \rho_{AB} = \sum_{k=1}^n p_k |k\rangle_{BB} \langle k|.$$

Thus we can also write that

$$E = H(\Psi) = S(\rho_A) = S(\rho_B) \quad (\text{measure of entanglement}).$$

5.3.1. 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 overcome this, Alice compresses the n particles she wants to send to Bob into nS particles and then teleports them (using the nS EPR pairs) to Bob. Bob then decompresses these back into n particles so finally they have n non-maximally entangled states, just as they wanted.

Note that this protocol holds only for the average case and for very large n , since only then the compression will have the efficiency nS .

5.4. 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 actually consist of the same letters only with different probabilities). 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 Bob 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).

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 much information (per letter) can one gain by knowing y . For example, if x, y are completely not correlated, then having learned y doesn't help at all and $H(X|Y) = H(X)$ which gives $I = 0$, so indeed no information is gained. On the other hand if they are completely correlated (one-to-one) then having learned y we need no more information. In this case $H(X|Y) = 0$ and $I = H(X)$.

Note, that the mutual information I is symmetric:

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

5.5. 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* $\text{Acc}(\rho)$

$$\text{Acc}(\rho) \equiv \max_{\{F_y\}} I(X; Y).$$

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

$$\text{Acc}(\rho) = H(X).$$

If however, the states are not all mutually orthogonal, then there is no general formula but it can be proven that

$$\text{Acc}(\rho) \leq S(\rho),$$

where an equality is reached only for very long messages ($n \rightarrow \infty$).

5.6. 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 $|\tilde{e}\rangle$

$$|\Psi\rangle_{\text{tot}} = |\psi\rangle|\tilde{e}\rangle = \frac{1}{\sqrt{2}} (|0\rangle + e^{i\alpha}|1\rangle) |\tilde{e}\rangle.$$

We further assume that the interaction of the system and the environment is very weak and we get after some time of interaction

$$|\Psi\rangle_{\text{tot}} \rightarrow \frac{1}{\sqrt{2}} (|0\rangle|e_0\rangle + e^{i\alpha}|1\rangle|e_1\rangle),$$

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

$$\rho \rightarrow \frac{1}{2} \begin{pmatrix} \langle e_0|e_0\rangle & e^{i\alpha}\langle e_0|e_1\rangle \\ e^{-i\alpha}\langle e_1|e_0\rangle & \langle e_1|e_1\rangle \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & e^{i\alpha}\langle e_0|e_1\rangle \\ e^{-i\alpha}\langle e_1|e_0\rangle & 1 \end{pmatrix}.$$

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.

5.6.1. 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

$$|\Psi\rangle_{\text{tot}} = \frac{1}{\sqrt{2}} (|\uparrow\rangle_A|\uparrow\rangle_B + |\downarrow\rangle_A|\downarrow\rangle_B) |\uparrow\rangle_E.$$

We now activate an interaction U_{AE} between particles A and E such that

$$U_{AE}|\Psi\rangle_{\text{tot}} = \frac{1}{\sqrt{2}} (|\uparrow\rangle_A|\uparrow\rangle_B|\uparrow\rangle_E + |\downarrow\rangle_A|\downarrow\rangle_B|\downarrow\rangle_E).$$

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

$$\rho_{AB} = \frac{1}{2} (|\uparrow\rangle_A\langle\uparrow| \otimes |\uparrow\rangle_B\langle\uparrow| + |\downarrow\rangle_A\langle\downarrow| \otimes |\downarrow\rangle_B\langle\downarrow|).$$

This looks like an entangled state, but is it? If we look at the entangled state

$$\frac{1}{\sqrt{2}} (|\uparrow\rangle_A|\uparrow\rangle_B + |\downarrow\rangle_A|\downarrow\rangle_B),$$

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

$$\rho_{AB} \neq \sum p_i \rho_A \otimes \rho_B \Rightarrow \text{entangled}.$$

5.6.2. 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).

5.7. 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 measurement 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

$$U = e^{i\theta_{\text{Env.}} \hat{n}_{\text{Env.}} \cdot \vec{\sigma}_{\text{sys}}} = \mathbb{1} + \varepsilon_1 \sigma_x + \varepsilon_2 \sigma_y + \varepsilon_3 \sigma_z,$$

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

$$\begin{array}{c} |0\rangle \\ |1\rangle \end{array} \xrightarrow{\mathbb{1}} \begin{array}{c} |0\rangle \\ |1\rangle \end{array},$$

$$\begin{array}{c} |0\rangle \\ |1\rangle \end{array} \xrightarrow{\sigma_x} \begin{array}{c} |1\rangle \\ |0\rangle \end{array} \quad (\text{bit flip}),$$

$$\begin{array}{c} |0\rangle \\ |1\rangle \end{array} \xrightarrow{\sigma_y} \begin{array}{c} i|1\rangle \\ -i|0\rangle \end{array},$$

$$\begin{array}{c} |0\rangle \\ |1\rangle \end{array} \xrightarrow{\sigma_z} \begin{array}{c} |0\rangle \\ -|1\rangle \end{array} \quad (\text{phase flip}).$$

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

$$U \left[\frac{1}{\sqrt{2}} (\alpha|0\rangle + \beta|1\rangle) |0\rangle|0\rangle \right] = \frac{1}{\sqrt{2}} (\alpha|0\rangle|0\rangle|0\rangle + \beta|1\rangle|1\rangle|1\rangle).$$

Now, assume that a bit flip occurs in one of the three qubits

$$\frac{1}{\sqrt{2}} (\alpha|0\rangle|0\rangle|0\rangle + \beta|1\rangle|1\rangle|1\rangle) \xrightarrow{\text{bit flip}} \text{or} \left\{ \begin{array}{l} \frac{1}{\sqrt{2}} (\alpha|1\rangle|0\rangle|0\rangle + \beta|0\rangle|1\rangle|1\rangle) \\ \frac{1}{\sqrt{2}} (\alpha|0\rangle|1\rangle|0\rangle + \beta|1\rangle|0\rangle|1\rangle) \\ \frac{1}{\sqrt{2}} (\alpha|0\rangle|0\rangle|1\rangle + \beta|1\rangle|1\rangle|0\rangle) \end{array} \right. .$$

If we make a measurement we will cause a collapse of the wave function, unless the wave function is already an eigenvector. The possible states 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

$$\begin{aligned} \uparrow &\rightarrow \tilde{\uparrow} = \frac{1}{2\sqrt{2}} (\uparrow\uparrow\uparrow + \downarrow\downarrow\downarrow) (\uparrow\uparrow\uparrow + \downarrow\downarrow\downarrow) (\uparrow\uparrow\uparrow + \downarrow\downarrow\downarrow), \\ \downarrow &\rightarrow \tilde{\downarrow} = \frac{1}{2\sqrt{2}} (\uparrow\uparrow\uparrow - \downarrow\downarrow\downarrow) (\uparrow\uparrow\uparrow - \downarrow\downarrow\downarrow) (\uparrow\uparrow\uparrow - \downarrow\downarrow\downarrow). \end{aligned}$$

If we define

$$\begin{aligned} |0\rangle &\equiv \frac{1}{\sqrt{2}} (\uparrow\uparrow\uparrow + \downarrow\downarrow\downarrow), \\ |1\rangle &\equiv \frac{1}{\sqrt{2}} (\uparrow\uparrow\uparrow - \downarrow\downarrow\downarrow), \end{aligned}$$

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.