University of Vienna

Faculty of Physics

Boltzmanngasse 5, 1090 Wien

Quantum Theory Based On Information And Entropy

Author:
Martin Pimon

Supervisor:
Ao. Univ.-Prof. i.R. Dr.
Reinhold Bertlmann

July 31, 2015

Contents

I A Foundational Principle **1**

1 Introduction **1**

2 The Elementary System **3**

3 The Inconsistency of Hidden Variables **4**

4 Catalogue of Knowledge **5**

5 Measure of Information **8**

6 Malus' Law **8**

7 Time Evolution of the Information Vector **10**

8 The von Neumann Equation **12**

9 Superposition **13**

10 Entanglement **13**

11 A General Bell Inequality **16**

12 The Non-Dichotomic Case **18**

13 Measurement **18**

II Entropy **19**

14 Shannon Entropy **19**

15 Von Neumann Entropy **21**

16 The Entropy of Entangled Systems **23**

17 Entropic Bell Inequalities **26**

18 Entropy in Quantum Measurement **28**

III Summary and Conclusion 32

Part I
A Foundational Principle

1 Introduction

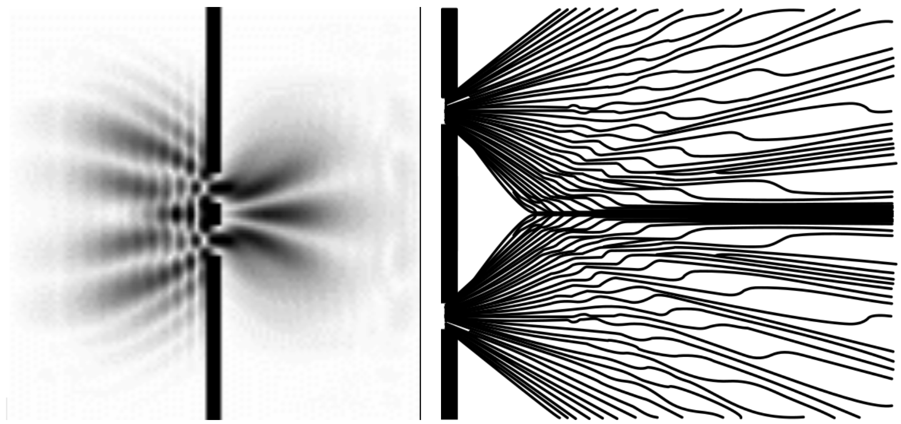
Quantum theory is one of the most important theories for describing the phenomena in Nature. Like the theory of relativity, it pictures the structure of our universe, but on a microscopic scale. However we currently debate the interpretations of quantum theory. These debates result in a number of coexisting theories; the most famous being the De Broglie-Bohm Theory and the Copenhagen Interpretation[1]. Why do such debates not exist for the theory of relativity?

Zeilinger suggests that this is because both the special and the general theory of Relativity are based on firm foundational principles, like the Principle of Equivalences[1]. This theory is built upon these principles and even counterintuitive consequences can be easily accepted. Using only the Principle of Equivalences, time dilation, Lorentz contraction or the fact that the speed of light must be a universal constant can be quickly derived and suddenly seem like the most natural and logical phenomena.

Yet quantum mechanics lacks such broadly accepted conceptual principles. When trying to interpret quite elementary experiments, such as the double slit experiment, a lot of conceptual questions arise. For example, one might want to study the trajectory of the particle. You can only measure the position on the measurement device; it is the purpose of the theory to give insights about what happens between the source and the detector. According to the De-Broglie Bohm Theory, the particle has a distinct trajectory which could theoretically be determined by solving a differential equation with the appropriate boundary conditions. On the other hand, the Copenhagen Interpretation states that the position for all times is merely given by a density of probabilities and if you looked for its distinct trajectory you would destroy the actual experiment. Interestingly, there is no way to distinguish these interpretations because they both predict exactly the same

[1]A few more examples are the Many Worlds Interpretation, the Transactional Interpretation, and Mermin's Ithaca Interpretation[1]. For a list with many different interpretations and a rough comparison see [2].

measurement results but they differ on a fundamental level.



(a) A solution at a time t for the double slit experiment using the Copenhagen Interpretation is shown. The black clouds indicate the value of the probability density of the wave function.[3]

(b) A simulation of a few Bohmian trajectories after the double slit are shown[4].

Figure 1: This Figure indicates the highly different fundamental concepts for the same experiment of two coexisting theories. On the one hand, you have a probability density and on the other, you have distinct trajectories for the propagation of the particle. In the end, both concepts predict the same result for the outcome of the experiment.

As you can see, an interpretation gives answers to elementary questions about Nature and this is why the debates over these interpretations exist. A possibility to solve these debates is to introduce a fundamental principle like in the theory of relativity and to reduce the known phenomena to it. In Part I an idea based on information will be expressed. Section 2 introduces the foundational principle and a lot of conclusions and derivations about experiments in quantum physics are shown. A reconstruction of quantum theory based on this principle will be attempted. Part II deals with the link between the concept of entropy and the results derived so far. We will conclude that entropy in Quantum Information builds a mathematical framework for the topics discussed.

2 The Elementary System

The idea of reconstructing quantum theory has already been discussed by several authors [1, 5, 6, 7, 10]. In an attempt to find a foundational principle, a certain view might be needed. Like Clifton, Bub and Halvorson appropriately state:

> *We suggest substituting for the conceptually problematic mechanical perspective on quantum theory an information-theoretic perspective. That is, we are suggesting that quantum theory be viewed, not as first and foremost a mechanical theory of waves and particles but as a theory about the possibilities and impossibilities of information transfer.[5]*

While Clifton et al. then continue to introduce their theory based on a strict mathematical approach, Zeilinger's more conceptual approach[1] will be shown here, but the idea of using concepts of information theory stays the same.

When trying to describe a physical object, essentially a set of true propositions is necessary, e.g. *"Is the object spherical? Does the object have a specific momentum of the value* p*? Is the object located at* x_0*? ..."* The answer would either be *true* or *false*, which is actually one bit of information. Evidently, for any macroscopic object a lot of propositions have to be made if you want to give a complete and accurate description. Also note that when you have full knowledge, all the propositions you need to ask must have a truth-value.

Now imagine splitting the system into smaller and smaller parts. What is left is an object which can be described by only a single proposition. It is called an *elementary system*, which is defined by the following:

> *An elementary system represents the truth value of one proposition.*

Or, in other words:

> *An elementary system carries 1 bit of information.*

In essence, this can be seen as quantization of information.

It is important to notice that we only consider binary experiments where the probability that the measurement outcome yields *true* is p and the probability for the outcome *false* is $1-p$ respectively. To extend the notion of this principle, let us consider an n-fold alternative to the binary one. Obviously, if $n = 2^N$, it is possible to reduce the system to binary ones. In a more general case if $n \neq 2^N$, such a decomposition is not trivial[10].

3 The Inconsistency of Hidden Variables

As a first consequence, imagine measuring the spin of a particle along the z-axis. You have prepared the state, such that the measurement yields the result $\uparrow$. Now, the system represents a truth value to the proposition: "A spin measurement along the z-axis will give the result $\uparrow$." But if you chose any other basis, you would lose that certainty and you would gain an element of randomness. An important property of this randomness is, that it is irreducible, i.e. it cannot be reduced to hidden variables in the system[10]. This is a consequence of the feature that

the information content of a system is finite.

Therefore, the loss of information for a specific measurement direction in a spin measurement depends on the angle θ. In the most extreme case, with $\theta = \frac{\pi}{2}$, i.e. measuring in x-basis after preparing the state in z-direction you will have no information whatsoever about the outcome. Zeilinger and Brukner call it *mutually exclusive propositions or variables*[10]. Another well-known example for mutually exclusive variables are the momentum p_0 and the location x_0 of a particle.

To emphasize the meaning of irreducible randomness and hidden properties, consider partial beam splitters, which reflect an incoming photon with probability p or transmit the photon with probability $1-p$. The beam splitters are aligned as shown in Figure 2, such that a detection event gives information about the path of the photon. Which path the photon takes seems to be probabilistic but could it be that there are some boundary conditions which determine the path of the photon deterministically? According to the principle introduced earlier, this is not an option. The photon would be able to carry 1 bit of information for every beam splitter it passes. You could even extend this experiment by repeatedly adding more and more partial beam splitters. In fact, the elementary system would be required to carry an almost infinite amount of information for each probabilistic event that occurs, which not only contradicts the principles suggested earlier but also raises a lot of new conceptual problems. The concept of so-called hidden variables is canceled out in this picture.

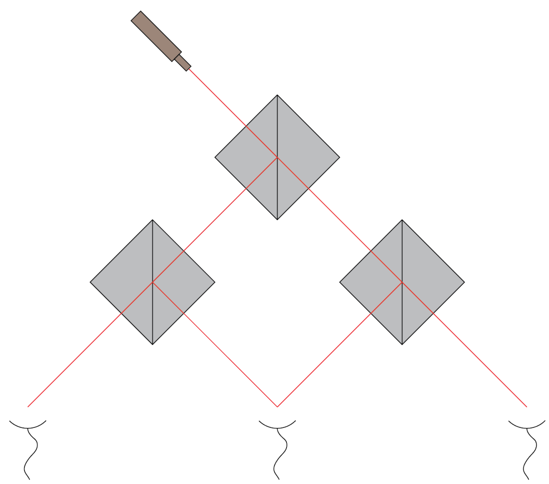


Figure 2: To illustrate the concept of hidden variables, consider an experiment with partial beam splitters. If its path was deterministic before the experiment is being made, the particle would have to carry 1 bit of information each time the particle passes a beam splitter. By adding more beam splitters, this idea is in contradiction to the statement that the information content of a system is finite.

4 Catalogue of Knowledge

In situations where it is not possible to assign simultaneously definite truth values to mutually exclusive propositions, like the velocity v_0 and the position x_0, or two orthogonal spin directions, we can assert measures of information about their truth values. Consider the situation for an experimenter who wants to predict a series of future experimental outcomes. Classically, he might make a list of deterministic predictions, while in quantum mechanics he can only guess and make probabilistic predictions: "In a series of N measurements, what is the number of occurrences of the outcome i, with probability p?"

Let the experimenters uncertainty before a single experiment be denoted by U_1, with

$$U_1(p) = p(1-p)N \equiv \sigma^2. \tag{1}$$

Because each trial contributes the same amount of information, the uncertainty will decrease with each individual performance for a single future experimental outcome such that after N trials

$$U_N(p) = \frac{\sigma^2}{N}. \tag{2}$$

So far, we only considered an experiment with two outcomes with the probabilities p and $1-p$. If there are n outcomes, denoted by a probability vector $\vec{p} \equiv (p_1, p_2, ..., p_n)$, satisfying $\sum_j p_j = 1$, equation (2) needs to have the more general form:

$$U(\vec{p}) = \sum_{j=1}^{n} U(p_j) = \sum_{j=1}^{n} p_j(1-p_j) = 1 - \sum_{j=1}^{n} p_j^2. \tag{3}$$

You can see that the uncertainty is minimal if a single $p_j = 1$ and is maximal if all probabilities are equal $p_j = \frac{1}{N}\ \forall j$.

Now that we know the uncertainties we should move on to construct a mathematical description for the information, denoted by $I(\vec{p})$. Naturally, the information is connected to the complement of $U(\vec{p})$. It is tempting to think that $I(\vec{p}) = 1 - U(\vec{p}) = \sum_{j=1}^{n} p_j^2$, which is the length of the probability vector $\vec{p}$. The problem is that not all vectors are possible. Since $\sum_{j=1}^{N} p_j = 1$, there exists a minimum vector length when all p_j are equal. Brukner and Zeilinger therefore suggest that $I(\vec{p})$ should be normalized with a normalization $\mathcal{N}$ [10]:

$$I(\vec{p}) = \mathcal{N} \sum_{j=1}^{n} \left(p_i - \frac{1}{n} \right)^2. \tag{4}$$

For example, an experimenter who performs an experiment with two outcomes would have an information content of

$$I(p_1, p_2) = (p_1 - p_2)^2. \tag{5}$$

While we now have a description for the information content of the system, it is not a complete description because it is invariant under the permutation of the set of possible outcomes. For the previously mentioned example, $I(\vec{p})$

would have the same value if you interchanged p_1 and p_2. In this particular case, another variable can be introduced to solve this ambiguity:

$$i = p_1 - p_2. \tag{6}$$

In a Stern-Gerlach experiment (see Figure 3) the variable i as introduced in equation (6) is a valid description for the information content for the measurement in one direction. Combining the three mutually exclusive directions is best accomplished by creating a vector

$$\vec{i} = (i_1, i_2, i_3) = (p_x^+ - p_x^-, p_y^+ - p_y^-, p_z^+ - p_z^-). \tag{7}$$

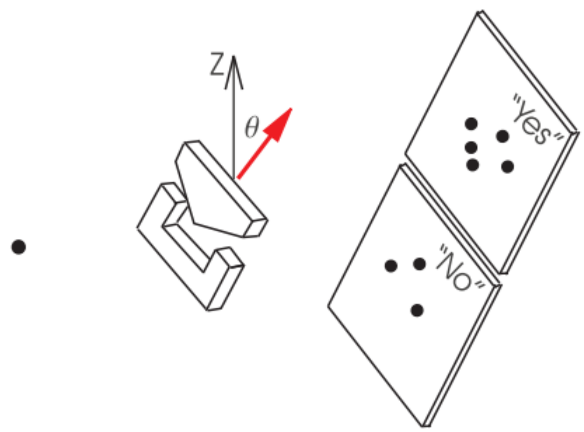


Figure 3: [10] The Stern-Gerlach Experiment consists of a magnet which creates an inhomogeneous magnetic field. It can be rotated around an angle θ. The spin of the atoms couples with this field and a separation of atoms which have spin $\uparrow$ or $\downarrow$ is detected. The first experiment was done with silver atoms and showed the quantization of the spin variable.

The variable $\vec{i}$ from equation (7) represents all the information an experimenter can have in a Stern-Gerlach experiment.n Thus we have derived a representation for the state vector $|\psi\rangle$ of the system which is in essence a *catalog of knowledge* and is able to give full information about the probability of the results of future experiments [13].

5 Measure of Information

As a first example to see that $\vec{i}$ is indeed the description of a physical object, we wish to specify a mapping of θ onto $\vec{i}(\theta)$ for rotations of the apparatus. Two assumptions have to be made in order to continue with the calculation.

The first assumption is, that the total information content of the system I_{total} is invariant under the change of the representation of the catalog of knowledge $\vec{i}$. To make this assumption more clear, we need to define the total information content of a system. The elementary system is only able to carry one bit as a maximum boundary and it is clear that it depends on the three mutually exclusive directions how much information an experimenter can actually extract. In this sense, the total information content is related to these three spatial directions

$$I_{total} = I_x + I_y + I_z = 1. \tag{8}$$

In more general situations for n-dimensional Hilbert spaces, $n^2 - 1$ parameters are needed to completely specify the density matrix ρ. However, like in the Stern-Gerlach experiment, measurements in a particular basis can yield only $n - 1$ independent probabilities which means that $n + 1$ distinct basis sets are needed to provide the $n^2 - 1$ parameters. This is why it seems natural to define I_{total} as a sum over all the individual measures of information over a complete set of $n + 1$ mutually complementary measurements,

$$I_{total} = \sum_{i=1}^{n+1} I_i(\vec{p}^{\,i}) = \mathcal{N} \sum_{i=1}^{n+1} \sum_{j=1}^{n} \left(p_j^i - \frac{1}{n} \right)^2 , \tag{9}$$

with $\vec{p}^{\,i} = (p_1^i, ..., p_n^i)$ representing the probabilities for the outcomes in the i-th measurement.

Going back to the rotation of the Stern-Gerlach experiment, we have three mutually exclusive directions:

$$I_{total} = I_1(\theta) + I_2(\theta) + I_3(\theta) = i_1^2(\theta) + i_2^2(\theta) + i_3^2(\theta). \tag{10}$$

6 Malus' Law

The other assumption is that the mapping we are going to specify is homogeneous, in a sense that a change of the parameter θ can be done continuously and is reversible. This assumption is reasonable because of the cosmological principle. It states that space is homogeneous and isotropic [14]. Additionally, if you change θ well enough, you end up with the same state you have had earlier: $\vec{i}(\theta_0) = \vec{i}(\theta_0+b)$. For the Stern-Gerlach Apparatus we know the value of b and it is equal to a full rotation, $b = 2\pi$.

Considering the first assumption, the mapping can be seen as a rotation R of our catalog of knowledge i in the vector space of information, which conserves the length of the vector

$$\vec{i}(\theta) = R(\theta_1 - \theta_0, \theta_0)\vec{i}(\theta_0). \tag{11}$$

Using the ansatz we got from our first assumption, the second assumption implies that the rotation does not depend on the specific value θ_0 but rather on the difference between the initial and the final values,

$$\begin{aligned} R(\theta - \theta_0, \theta_0)\vec{i}(\theta_0) &= R(\theta - \theta_0, \theta_0 + b)\vec{i}(\theta_0 + b) \\ \Rightarrow R(\theta - \theta_0, \theta_0) &= R(\theta - \theta_0, \theta_0 + b). \end{aligned} \tag{12}$$

As a consequence, a rotation of the apparatus for two consecutive times must not be different from a continuous rotation:

$$R(\theta_2 - \theta_1)R(\theta_1 - \theta_0) = R(\theta_2 - \theta_0). \tag{13}$$

For two arbitrary vectors x and y the following condition must hold:

$$\langle Rx|Ry\rangle = \langle x|R^T Ry\rangle = \langle x|y\rangle \Rightarrow R^T R = \mathbb{1}. \tag{14}$$

Every rotation Matrix (which conserves length) is orthogonal. It follows that $R^{-1} = R^T$.

In the 3-dimensional Hilbert space, for rotations along an axis, R has known forms. For example,

$$R(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & f(\theta) & -g(\theta) \\ 0 & g(\theta) & f(\theta) \end{pmatrix}, \tag{15}$$

with $\theta_0 = 0$ and analytical functions $f(\theta)$ and $g(\theta)$ satisfying

$$\begin{aligned} f(\theta)^2 + g(\theta)^2 &= 1 \\ f(0) = 1, g(0) &= 0. \end{aligned} \tag{16}$$

Yet another condition can be obtained by looking at infinitely small rotations $d\theta$ in equation (13),

$$R(\theta + d\theta) = R(\theta)R(d\theta). \tag{17}$$

Multiplying the two matrices yields the following condition:

$$f(\theta + d\theta) = f(\theta)f(d\theta) - g(\theta)g(d\theta). \tag{18}$$

When using the conditions (16) we get a first-order differential equation with boundary values,

$$f'(\theta) = -n\sqrt{1 - f^2(\theta)}. \tag{19}$$

The solution is $f(\theta) = \cos n\theta$, where $n = -g'(0)$ which is a constant. We have now derived an expression for the rotation matrix,

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos n\theta & -\sin n\theta \\ 0 & \sin n\theta & \cos n\theta \end{pmatrix}. \tag{20}$$

This result gives the usual result for the probability in quantum mechanics,

$$p = \cos^2 \frac{n\theta}{2}. \tag{21}$$

Naturally, for general rotations more matrices have to be found, each rotating the vector around an orthogonal set of vectors $\{x, y, z\}$,

$$R(\alpha, \beta, \gamma) = R_x(\alpha)R_y(\beta)R_z(\gamma). \tag{22}$$

7 Time Evolution of the Information Vector

In the following Section we use the same approach to obtain the time evolution of a system. We want to specify a map for the parameter t on $\vec{i}(t)$. Again we assume that the total information content of the system remains invariant under the time evolution

$$I_{total}(t) = \sum_{n=1}^{3} i_n^2(t) = \sum_{n=1}^{3} i_n^2(t_0) = I_{total}(t_0). \tag{23}$$

This assumption is valid if we additionally assume that the system is not in contact with an environment. Hence the time evolution can also be seen as a rotation in the vector space of information. In equation (14) we have proved that $R^{-1} = R^T$. It is therefore possible to write $\vec{i}(t)$ as

$$\vec{i}(t) = R^T(t, t_0)\vec{i}(t_0). \tag{24}$$

When defining $K(t, t_0) \equiv \frac{dR(t,t_0)}{dt} R^T(t, t_0)$, the time evolution reduces to

$$\frac{d\vec{i}(t)}{dt} = K(t, t_0)\vec{i}(t). \tag{25}$$

A short analysis of the matrix $K(t, t_0)$ reveals that it is antisymmetric:

$$\begin{aligned} K^T(t) = R(t)\frac{dR^T(t)}{dt} = R \lim_{\Delta t \to 0} \frac{R^T(t+\Delta t) - R^T}{\Delta t} = \\ R(t) \lim_{\Delta t \to 0} R^T(t) \frac{R(t) - R(t+\Delta t)}{\Delta t} R^T(t+\Delta t) = \\ \lim_{\Delta t \to 0} \frac{R(t) - R(t+\Delta t)}{\Delta t} R^T(t) = -K(t). \end{aligned} \tag{26}$$

Every antisymmetric matrix K in three dimensions must have a distinct form and there exists a unique vector $\vec{k}$ which satisfies

$$K\vec{y} = \begin{pmatrix} 0 & -k_3 & k_2 \\ k_3 & 0 & -k_1 \\ -k_2 & k_1 & 0 \end{pmatrix} \cdot \vec{y} = \vec{k} \times \vec{y}, \quad \forall \vec{y}. \tag{27}$$

The time evolution is therefore given by

$$\frac{d\vec{i}(t)}{dt} = \vec{k}(t, t_0) \times \vec{i}(t). \tag{28}$$

Notice that the axis $\vec{k}(t, t_0)$, the information vector $\vec{i}(t)$ is rotated around, may also change in time.

8 The von Neumann Equation

In the previous section, a mathematical description of physical experiments without any knowledge about quantum mechanics was deduced. In this section, we will further discuss the time evolution we have derived in equation (28) and we will show that it is the same as the time evolution given by the von Neumann equation.

In quantum mechanics, a physical system can be described by its density matrix. A general mathematical formalism uses the Hilbert-Schmidt basis. It consists of the unity operator and the Pauli-matrices:

$$\mathbb{1}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{29}$$

The density operator can be written as

$$\rho(t) = \frac{1}{2}\left(\mathbb{1}_2 + \sum_{j=1}^{3} i_j(t)\sigma_j\right). \tag{30}$$

Like in the previous section, we take the time derivative of the catalog of knowledge,

$$i\hbar\frac{d\rho(t)}{dt} = \frac{1}{2}\sum_{j=1}^{3}\frac{i_j(t)}{dt}\sigma_j. \tag{31}$$

Now equation (28) is plugged in, using the antisymmetric ε-*tensor* to calculate the cross product,

$$i\hbar\frac{d\rho(t)}{dt} = \frac{1}{2}\sum_{i,j,k=1}^{3}\varepsilon_{ijk}k_i(t)i_j(t)\sigma_k. \tag{32}$$

The commutator of pauli matrices can be calculated as $[\sigma_i, \sigma_j] = 2i\sum_{k=1}^{3}\varepsilon_{ijk}\sigma_k$ and therefore

$$i\hbar\frac{d\rho(t)}{dt} = \frac{1}{4}\sum_{i,j=1}^{3}k_i(t)i_j(t)(\sigma_i\sigma_j - \sigma_j\sigma_i). \tag{33}$$

When defining $k_i(t) \equiv Tr[H(t)\sigma_i]$ we end up with the von Neumann equation in quantum mechanics,

$$i\hbar\frac{d\rho(t)}{dt} = [H(t), \rho(t)]. \tag{34}$$

The principle introduced eventually led to a time evolution of a quantum mechanical system which is consistent with methods that are already experimentally tested. For pure states equation (34) is reduced to the Schrödinger equation.

9 Superposition

Because the state of the system is only determined by our information about it, superposition can be more easily understood. Consider the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle). \tag{35}$$

The system itself is not physically realized in the two states at the same time but rather it is our knowledge before the experiment which causes a seemingly irritating view of a particle being in two states at once. It is merely a mathematical description.

10 Entanglement

So far, only a single elementary system has been inspected. What about more complex systems, specifically systems made up on N elementary systems? If the systems are separated, they have nothing to do with each other. Clearly, the information content of the system will just be the sum of the individual sub-systems [1].

> N elementary systems represent the truth values of N propositions.
>
> N elementary systems carry N bits of information.

Now let the systems interact with each other and then separate them again. Upon interaction the systems may distribute their information content. Nevertheless, it is suggestive to assume that the information is conserved. After the interaction, each elementary system might not carry the same information as before, as the system as a whole could now be represented by joint properties. When no individual system carries any information on its own and all the information is found in joint properties, the system is in an *entangled* state.

Entangled States

To describe the situation in our formalism we can ask some propositions. Considering two particles the most simple case is if we ask each system individually about its spin e.g. along the z-axis: "The spin of particle 1 along the z-axis is up" and "The spin of particle 2 along the z-axis is up". The resulting possibilities for the states are called product states:

$$\begin{aligned} |\psi\rangle_1 &= |z+\rangle_1 |z+\rangle_2 \quad & |\psi\rangle_2 &= |z+\rangle_1 |z-\rangle_2 \\ |\psi\rangle_3 &= |z-\rangle_1 |z+\rangle_2 \quad & |\psi\rangle_4 &= |z-\rangle_1 |z-\rangle_2 \end{aligned} \tag{36}$$

For convenience an abbreviated mathematical description is used, $|a\rangle_1 |b\rangle_2 \equiv |a\rangle_1 \otimes |b\rangle_1$, for the respective Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2$.

This proposition is not the only possibility. Another method is making propositions about joint properties of the system. For two particles a suitable proposition would be: "The two spins along the z-axis are the same." The second proposition could be a proposition about one of the two systems and the mathematical description needed would again be the product states. As an alternative, the second proposition could be one that describes joint properties again: "The two spins are equal along the x-axis." How can these propositions both be true for the same particles? They can when introducing a maximally entangled state, a Bell state given by

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|z+\rangle_1 |z+\rangle_2 + |z-\rangle_1 |z-\rangle_2) \\ &= \frac{1}{\sqrt{2}}(|x+\rangle_1 |x+\rangle_2) + |x-\rangle_1 |x-\rangle_2). \end{aligned} \tag{37}$$

Such a state contains no information about the individual subparts because the two bits of information are used up by defining the state.

This notion can be extended to any number of entangled particles. For three particles suitable elementary propositions would be
"The spins of particle 1 and 2 are equal along the z-axis."
"The spins of particle 1 and 3 are equal along the z-axis."
"The spins of all three particles are an even number of permutations of being in the up-direction along the x-axis."
The first two propositions yield four pairs of superposition states, each with a different relative phase of either $\varphi = 0$ or $\varphi = \pi$ resulting in a different sign. The last proposition we need to construct is necessary for the eight distinct states to be distinguished. Greenberger, Horne and Zeilinger introduced such entangled states consisting of three particles, so-called GHZ states [16]. The state for which these three propositions are true is

$$\begin{aligned} |\psi\rangle &= \frac{1}{\sqrt{2}}(|z+\rangle_1 |z+\rangle_2 |z+\rangle_3 + |z-\rangle_1 |z-\rangle_2 |z-\rangle_3) \\ &= \tfrac{1}{8}(|x+\rangle_1 |x+\rangle_2 |x+\rangle_3 + |x+\rangle_1 |x-\rangle_2 |x-\rangle_3 + \\ &\quad |x-\rangle_1 |x+\rangle_2 |x-\rangle_3 + |x-\rangle_1 |x-\rangle_2 |x+\rangle_3). \end{aligned} \tag{38}$$

We can understand entanglement as a consequence of our principles.

Total Information Content

In many ways, entanglement reveals properties that cannot be explained classically. For instance, we have calculated the total information content of the binary experiment in equation (5) as $I = (p_1 - p_2)^2$. Now we want to calculate the total information content of two particles [15]. The following propositions determine the state of the system uniquely: "The spin of particle 1 along x and the spin of particle 2 along z are the same" and "The spin of particle 1 along x and the spin of particle 2 along z are different". In this case, we will not worry about mutually exclusive complementary propositions for the composite system but we consider the subset of them which concerns joint properties. We denote the possibility of the measurement outcome by p^+_{xz} and p^-_{xz} respectively so that equation (5) can be used and we get

$$I_{xz} = (p^+_{xz} - p^-_{xz})^2. \tag{39}$$

These two propositions lead to the product states as shown in (36). Only one proposition with definite truth value can be made about joint properties because the other proposition is used to define one particle individually.

Thus the correlations are fully represented by the correlations between spin measurements along an axis, e.g. the x-axis on both elementary systems:

$$I_{xx} = 1. \tag{40}$$

Because the choice of axis for measurement is arbitrary, the total information contained should not be dependent on the axis used in the measurement. Therefore, the total information contained is defined as the sum of partial measures of information contained in the set of complementary observations within the $x-z$-planes,

$$I_{corr} = I_{xx} + I_{xz} + I_{zx} + I_{zz}. \tag{41}$$

The maximal value of I_{corr} can be obtained by an optimization over all possible two dimensional planes of measurements on both sides. When considering the Bell state as denoted in equation (37), all the information is stored in joint properties. This means that $I_{xx} = I_{zz} = 1$ and $I_{xz} = I_{zx} = 0$, and consequently

$$I_{corr}^{Bell} = 2. \tag{42}$$

In contrast to product states, Brukner, Zukowski and Zeilinger suggest entanglement of two elementary systems to be defined in general so that more than one bit (of the two available ones) is used to define joint properties [15]:

$$I_{corr}^{entangled} > 1, \quad I_{corr}^{not\ entangled} \leq 1. \tag{43}$$

11 A General Bell Inequality

Bell's inequality is a historically as well as a fundamentally important relation describing the differences between classical and quantum physics. In their famous paper in 1935, Einstein, Podolsky and Rosen challenged quantum theory by arguing that if it was a theory of *local realism* and *physical reality*, the notion of entanglement would lead to a contradiction by violating Heisenberg's uncertainty principle. They suggested that the theory is incomplete because one would need to introduce hidden variables [11].

In 1964, John Bell introduced an inequality which was suitable for testing the argument of Einstein, Podolsky and Rosen quantitatively. The inequality holds for all local hidden variable theories [12]. However, experiments showed that the inequality is violated. In this section, a Bell inequality is dervied from our foundational principle [15].

Consider N observers and let each choose two dichotomic variables which are determined by two parameters $\vec{n}_1$ and $\vec{n}_2$, e.g. the spin of a particle with its respective directions. Then $A_j(\vec{n}_1)$ and $A_j(\vec{n}_2)$ are two numbers which take the values $+1$ or -1 after the measurement. To represent the correlation between all N observations it is applicable to use the product $\prod_{j=1}^{N} A_j(\vec{n}_{k_j})$, with $k_j = 1, 2$. The average over many runs of the experiment is the correlation function

$$E(k_1, ..., k_N) = \left\langle \prod_{j=1}^{N} A_j(\vec{n}_{k_j}) \right\rangle . \tag{44}$$

For each observer j one has the identities $|A_j(\vec{n}_1) + A_j(\vec{n}_2)| = 0$ and $|A_j(\vec{n}_1) - A_j(\vec{n}_2)| = 2$ (or the other way around). Take an arbitrary function $S(s_1, ..., s_N) = \pm 1$ with $s_1, ..., s_N \in \{-1, +1\}$, then the following algebraic identity must hold (for predetermined results):

$$\sum_{s1,...,s_N=\pm 1} S(s_1, ..., s_N) \prod_{j=1}^{N} [A_j(\vec{n}_1) + s_j A_j(\vec{n}_2)] = \pm 2^N . \tag{45}$$

For all sign sequences of $s_1, ..., s_N$ the product $\prod_{j=1}^{N} [A_j(\vec{n}_1) + s_j A_j(\vec{n}_2)]$ vanishes except for just one sign sequence, which gives $\pm 2^N$.

Taking the absolute value in (45) and using (44) give a set of Bell inequalities [8],

$$\left| \sum_{s_1,...,s_N=\pm 1} S(s_1, ..., s_N) \sum_{k_1,...,k_N=1,2} s_1^{k_1-1} \cdots s_N^{k_N-1} E(k_1, ..., k_N) \right| \leq 2^N . \tag{46}$$

For $N = 2$ the CHSH-inequality [9] is obtained:

$$|E(1, 1) + E(1, 2) + E(2, 1) - E(2, 2)| \leq 2, \tag{47}$$

for $N = 3$ the resulting inequality is

$$|E(1, 2, 2) + E(2, 1, 2) + E(2, 2, 1) - E(1, 1, 1)| \leq 2. \tag{48}$$

12 The Non-Dichotomic Case

So far, we have restricted the elementary system to a dichotomic variable. What about systems which need statements that are not binary? Obviously, any $n = 2^N$ fold can be decomposed into N binary ones. For $N \neq 2^N$ one needs to consider the total information content (9). Somehow, this information content must be stored in the system's density matrix ρ because it represents the state of the system. As discussed in Section 5, to define the density matrix ρ in an n-dimensional Hilbert space, one needs $n^2 - 1$ parameters. A measurement within a single basis set can only yield $n-1$ parameters, since the sum over all possible probabilities in an experiment is equal to one. This means that $n + 1$ uncorrelated basis sets are needed to fully describe the system. Wootters and Fields showed that these sets can be found if n is the power of a prime number [17]. For that reason, Brukner and Zeilinger suggest that this is the reason why the notion of the elementary system can be extended to all prime number dimensional systems [10].

13 Measurement

Another counterintuitive consequence of quantum mechanics is the act of measurement. When viewing the collapse of the wave function as a real physical process, conceptual difficulties arise. Using the information-theoretic approach, we see that the wave function is essentially a mathematical description for our knowledge of the properties of the system. The wave function having a non zero part at some point in space does not automatically mean that a particle is physically there. It only means that our knowledge allows the particle the possibility of being present at that point. When a measurement is performed, our knowledge of the system changes, and therefore its representation, the quantum state, also changes. In the measurement, the state therefore must appear to be changed in accordance with the new information. Also, complementary information is unavoidably and irrecoverably lost. The difference between a measurement in classical mechanics is, that a quantum mechanical measurement does not add some knowledge but changes our knowledge because the information content of a system is finite [10].

Part II
Entropy

As we have seen in Part I, using an information theoretic approach to quantum physics can be beneficial for a deeper understanding. In Part II we continue with our approach and the concepts of entropy will be introduced. It is demonstrated how these concepts can be used to also derive an information theory on quantum entanglement and measurement.

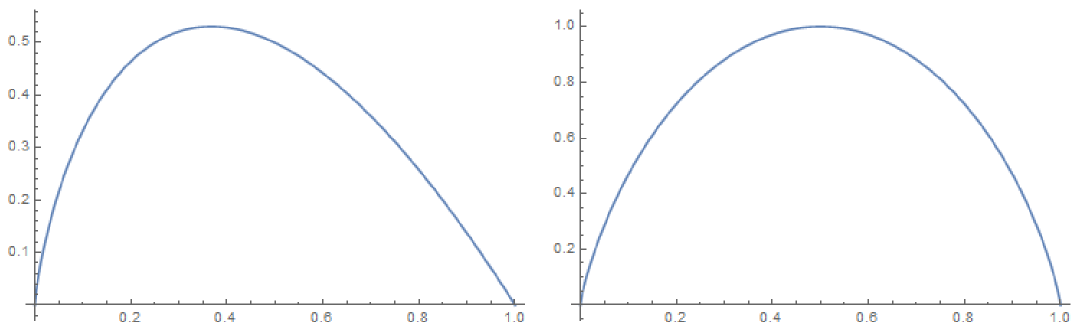
14 Shannon Entropy

First, the classical concept for entropy in information theory is discussed [18]. The Shannon entropy plays a central role in this model. There are two equivalent views on how to interpret Shannon entropy: On the one hand it is a measure for the information we gain, on average, when learning about a random variable X. On the other hand it can be seen as the amount of uncertainty about X before we learn its value. Thus it is in some sense a measure for the lack of knowledge. For a random variable X which can take n values or symbols with respective probabilities $p_x,\ x = \{1, ..., n\}$ it is defined as

$$H(X) \equiv H(p_1, ..., p_n) = -\sum_x p_x \log_2 p_x. \tag{49}$$

The most simple example is the binary experiment. $H(X)$ will have a maximum if $p_1 = p_2 = \frac{1}{2}$ and vanishes if p_1 or p_2 are equal to 1 or 0 respectively (see Figure 4).

The usefulness of the definition of the Shannon entropy is not obvious at first. However it becomes clear when looking at an example. Thinking about a source which produces four symbols $\{A, B, C, D\}$ with the respective probabilities $\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right\}$, one would use two bits to store each of the symbols, in a naive approach. However, we can make use of the bias to compress the source. Encode A as the bit string 0, B as 10, C as 110 and D as 111. The average length of the bit string is $\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 \cdot 2 = \frac{7}{4}$ bits which remarkably corresponds to its Shannon entropy $H(X) = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = \frac{7}{4}$ bits! Surprisingly, it is less than the



(a) **(b)**

Figure 4: The functional behavior of two functions connected to the Shannon entropy is demonstrated. In (a), you see how the function $y(x) = -x \log_2 x$ behaves. Notice that $\lim_{x\to 0} y(x) = 0$. We define $0 \cdot \log_2 0 \equiv 0$, which makes sense because outcomes for the random variable X which are impossible should not be taken into account for the entropy. In (b) $H(X)$ for the binary experiment is shown. Notice that the entropy is maximal when our knowledge of the experimental outcome is minimal.

naive approach would suggest. The result $H(X)$ gives quantifies the optimal compression that may be achieved, or in other words, it is used to quantify the minimum resources needed to store information.

Conditional Entropy and Mutual Information

Suppose X and Y are two random variables. To describe how the information content of X is related to the information content of Y the *conditional entropy* and *mutual information* can be helpful concepts.

First, the joint entropy will be introduced. It is defined in a straightforward manner:

$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y) \tag{50}$$

and it may in the same way be extended for any number of random variables. Like before, it can be seen as the uncertainty about the pair (X, Y).

If the value of Y is known we have acquired $H(Y)$ bits of information about the pair (X, Y). The remaining uncertainty about (X, Y) is associated with the lack of knowledge for the value of X. The entropy of X *conditional* on knowing Y is then defined as

$$H(X|Y) \equiv H(X,Y) - H(Y) = -\sum_{x,y} p_{x,y} \log_2 p_{x|y}, \tag{51}$$

with $p_{x,y}$ being the joint probability and $p_{x|y}$ being the conditional probability [20]. When X and Y are functionally dependent, then $H(X|Y) = 0$ and when X and Y are stochastically independent, then $H(X|Y) = H(X)$. Because of that, the following relation must hold:

$$0 \leq H(X|Y) \leq H(X). \tag{52}$$

The mutual information is a measure of information which is common in X and Y. An ansatz is to add the information content of X, namely $H(X)$ to the information content of Y, but information which is common in X and Y would be counted twice in this sum and the information which is not common would be counted only once. Since we know the joint information of the pair, $H(X,Y)$ we can subtract this quantity again and end up with

$$H(X:Y) = H(X) + H(Y) - H(X,Y). \tag{53}$$

Combining equations (51) and (53) the connection between mutual information and conditional entropy is built:

$$H(X) = H(X|Y) + H(X:Y). \tag{54}$$

15 Von Neumann Entropy

Shannon entropy describes the entropy in a classical way. Quantum states given by the density matrix ρ have the ability to be in a superposition and another framework is needed. John von Neumann defined the entropy as

$$S(\rho) = -\mathrm{Tr}[\rho \log_2 \rho]. \tag{55}$$

If λ_x are the eigenvalues of ρ, the entropy can be re-written as

$$S(\rho) = -\sum_x \lambda_x \log_2 \lambda_x, \tag{56}$$

which somehow resembles the Shannon entropy [19].

Von Neumann Conditional Entropy and Mutual Information

Using the same approach as for the Shannon conditonal entropy (51), the conditional von Neumann entropy is defined [20] as

$$S(X|Y) = -\mathrm{Tr}_{XY}[\rho_{XY} \log_2 \rho_{X|Y}], \tag{57}$$

with $\rho_{X|Y}$, the so-called *amplitude matrix*, being analogue to the conditional probability $p_{i|j}$,

$$\rho_{X|Y} = \exp\left\{\log_2 \rho_{XY} - \log_2 \left[\mathbb{1}_X \otimes \rho_Y\right]\right\} = \lim_{n\to\infty} \left[\rho_{XY}^{1/n} \left(\mathbb{1}_X \otimes \rho_Y\right)^{-1/n}\right]^n. \tag{58}$$

Notice that the reduced density matrix $\rho_Y = \mathrm{Tr}_X[\rho_{XY}]$ is also analogous to the marginal probability $p_j = \sum_i p_{ij}$. The reason it is called amplitude matrix is because it retains the quantum phases relating X and Y in contrast with a conditional probability based on squared amplitudes. In a classical limit, ρ_{XY} is diagonal with entries p_{ij} and so is $\rho_{X|Y}$ with diagonal entries $p_{i|j}$.

The relation (51) still holds for the von Neumann entropy:

$$S(X|Y) = S(X,Y) - S(Y). \tag{59}$$

One can verify this by plugging in the definition of the amplitude matrix (58) into the conditional von Neumann entropy (57):

$$S(X|Y) = -\mathrm{Tr}_{XY}\left[\rho_{XY} \log_2 \rho_{XY}\right] + \mathrm{Tr}_{XY}\left[\rho_{XY} \log_2 \left(\mathbb{1}_X \otimes \rho_Y\right)\right]. \tag{60}$$

The first term on the right side in (60) clearly is $S(X,Y)$ and the second term is equal to $-S(Y)$, which can be shown by using the relation

$$\begin{aligned} \mathrm{Tr}_{XY}\left[\rho_{XY} \log_2 \left(\mathbb{1}_X \otimes \rho_Y\right)\right] = \mathrm{Tr}_{XY}\left[\rho_{XY} \left(\mathbb{1}_X \otimes \log_2 \rho_Y\right)\right] = \\ \mathrm{Tr}_Y\left\{\mathrm{Tr}_X[\rho_{XY}] \log_2 \rho_Y\right\} = \mathrm{Tr}_Y[\rho_Y \log_2 \rho_Y]. \end{aligned} \tag{61}$$

While classically $p_{i|j}$ satisfies $0 \leq p_{i|j} \leq 1$, the amplitude matrix ρ_{XY} does not. It can have eigenvalues larger than one and therefore the conditional von Neumann entropy can be negative. As a consequence, in quantum information theory it is acceptable that the entropy of the entire System can be smaller than the entropy of one of its subparts, $S(X,Y) < S(Y)$, which is the case in quantum entanglement.

An expression for the mutual von Neumann entropy can also be analogously defined,

$$S(X:Y) = -\mathrm{Tr}_{XY}[\rho_{XY} \log_2 \rho_{X:Y}]. \tag{62}$$

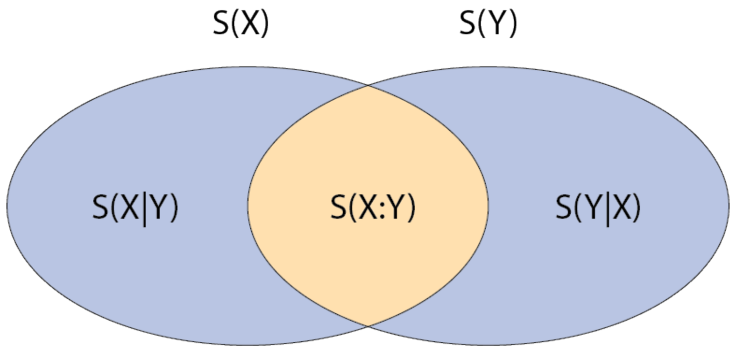


Figure 5: To illustrate the concept of mutual information and conditional entropy, a Venn diagram is shown. On the top, the respective entropies are separated from each other. The relative complements are the conditional entropies and the intersection is the information which is common in X and Y, which is the mutual information.

The mutual amplitude matrix is accordingly constructed as

$$\rho_{X:Y} = \exp\left\{\log_2\left(\rho_X \otimes \rho_Y\right) - \log_2 \rho_{XY}\right\} = \lim_{n\to\infty}\left[\left(\rho_X \otimes \rho_Y\right)^{1/n} \rho_{XY}^{-1/n}\right]^n. \tag{63}$$

As above, the relation (53) still holds for the von Neumann entropy:

$$S(X:Y) = S(X) - S(X|Y) = S(X) + S(Y) - S(X,Y). \tag{64}$$

16 The Entropy of Entangled Systems

The constructed formalism above allows a description of quantum phenomena such as entanglement [20]. Consider one of the Bell states which followed intuitively from Zeilinger's foundational principle in Section 10,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right). \tag{65}$$

A notation to shorten the mathematical formalism is used where $|z+\rangle \equiv |0\rangle$, $|z-\rangle \equiv |1\rangle$ and the tensor product between the two Hilbert spaces is denoted by $|\psi\rangle_X \otimes |\phi\rangle_Y \equiv |\psi\phi\rangle$.

$|\Phi\rangle^+$ is a pure state in $\mathcal{H}_X \otimes \mathcal{H}_Y$, the density matrix is idempotent, $\rho_{XY} = \rho_{XY}^2$, and thus its total entropy correspondingly vanishes $S(X,Y) = 0$. Calculating the respective density matrices leads to

$$\rho_{XY} = |\Phi^+\rangle \langle \Phi^+| = \frac{1}{2} \cdot \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \tag{66}$$

$$\rho_X = \mathrm{Tr}_Y[\rho_{XY}] = \rho_Y = \mathrm{Tr}_X[\rho_{XY}] = \frac{1}{2} \cdot \mathbb{1}_2. \tag{67}$$

The conditional density matrix (58) can now be calculated:

$$\rho_{X|Y} = \rho_{XY} \left(\mathbb{1}_X \otimes \rho_Y\right)^{-1} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{68}$$

Plugging in the calculated results above for the density matrix of the composite system (66) and the conditional density matrix (68) into the definition for the von Neumann conditional entropy (57) and using the relations (55) and (59) we get

$$S(X,Y) = S(X) + S(X|Y) = 1 - 1 = 0, \tag{69}$$

which confirms the definition. This is only possible because $\rho_{X|Y}$ has an eigenvalue greater than 1 and in this way violates classical information theory by making negative entropy possible.

This concept can be extended to multipartite systems. For three systems, the conditional entropies, i.e. $S(X|Y,Z)$, $S(Y|X,Z)$, $S(Z|X,Y)$ are given as a simple generalization of equation (59), e.g.

$$S(X|Y,Z) = S(X,Y,Z) - S(Y,Z), \tag{70}$$

and likewise for the conditional mutual entropies,

$$S(X:Y|Z) = S(X|Z) - S(X|Y,Z). \tag{71}$$

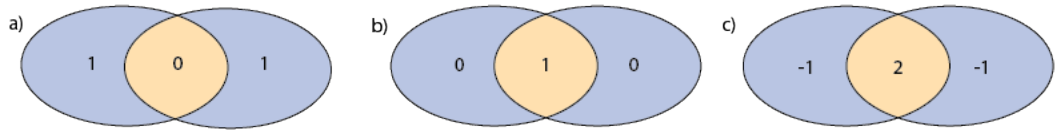


Figure 6: In this Figure, we see the values for the mutual information and conditional entropy for selected situations in analogy to Figure 5. In case a) and b) the two systems are classically correlated. Case c) shows the values when we have entanglement

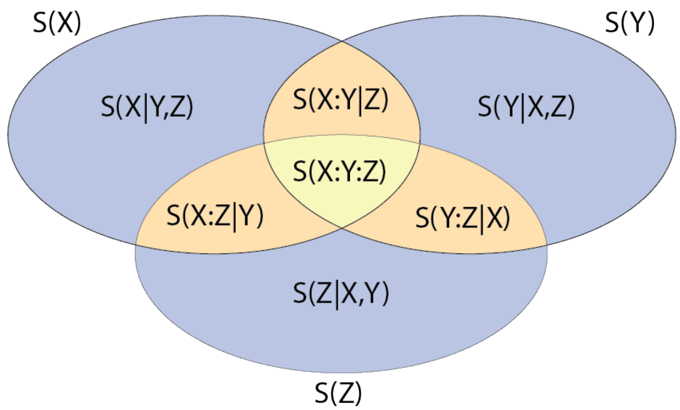


Figure 7: A Venn diagram for the entropy relations for three systems is illustrated.

Plugging in (59) and (70) into (71) we can rewrite this equation as

$$S(X:Y|Z) = S(X,Z) + S(Y,Z) - S(Z) - S(X,Y,Z). \tag{72}$$

The Venn Diagram is shown in Figure 7.

Looking at a GHZ-state which resulted in section 10,

$$|\psi_{XYZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \tag{73}$$

we know that it is maximally entangled and therefore $S(X,Y,Z) = 0$. When looking at the subsystem XY, we find that

$$\rho_{XY} = \mathrm{Tr}_Z[\rho_{XYZ}] = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) = \frac{1}{2} \cdot \mathbb{1}_4. \tag{74}$$

ρ_{XY} is physically indistinguishable from a statistical ensemble prepared with an equal number of $|00\rangle$ and $|11\rangle$ states, hence X and Y are correlated in Shannon theory, when ignoring Z. The state is completely mixed, but not entangled. Looking at Figure 8 it can be seen that by performing the partial trace over Z, the part with negative entropy is removed and classical correlations emerge. (X, Y) has a positive entropy $S(X, Y)$ of 1 bit but the entropy of Z conditional on (X, Y), $S(Z|X, Y)$ is again negative and cancels out $S(X, Y)$:

$$S(X, Y, Z) = S(X, Y) + S(Z|X, Y) = 1 - 1 = 0. \tag{75}$$

This further confirms the mathematical structure for this formulation and the concept of negative entropy in entangled states. Obviously, this process can be extended to multipartite systems.

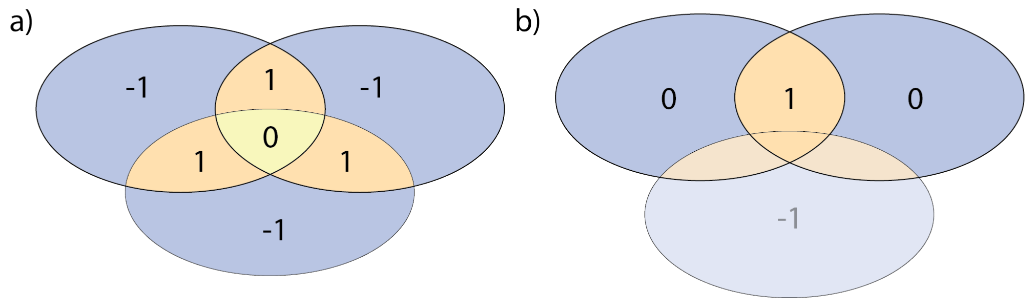


Figure 8: a) This is the Venn diagram for an entangled state of three particles. b) By tracing over Z and thereby ignoring it, the other two systems appear as being classically correlated

17 Entropic Bell Inequalities

In Section 11 a general Bell inequality has been derived. This time a different approach is used [21].

Consider three dichotomic random variables X, Y and Z, representing properties of the system. After a measurement a result x, y or z is obtained which can only take on the values $+1$ or -1. The outcomes must obey

$$xy + xz - yz \leq 1. \tag{76}$$

This inequality still holds for cyclic permutations of x, y and z. Taking the average of equation (76) and its cyclic permutations we get three basic Bell inequalities

$$\begin{aligned} \langle xy \rangle + \langle xz \rangle - \langle yz \rangle &\leq 1, \\ \langle xy \rangle - \langle xz \rangle + \langle yz \rangle &\leq 1, \\ -\langle xy \rangle + \langle xz \rangle + \langle yz \rangle &\leq 1. \end{aligned} \tag{77}$$

These equations can be consolidated into a single inequality,

$$|\langle xy \rangle - \langle xz \rangle| + \langle bc \rangle \leq 1. \tag{78}$$

This inequality can be violated when viewing the system with a joint density matrix ρ_{XYZ} which forces corresponding entropies to be negative.

The relation (54) can be straightforwardly extended:

$$H(X:Y) = H(X:Y|Z) + H(X:Y:Z). \tag{79}$$

The first term on the right hand side is that piece of information that is not shared by the third variable. The other term is the mutual information of X, Y and Z and can be written as

$$\begin{aligned} H(X) + H(Y) + H(Z) - H(X,Y) - H(X,Z) - H(Y,Z) + H(X,Y,Z) = \\ H(X:Y:Z). \end{aligned} \tag{80}$$

In the diagram in Figure 7 we have experimental access to the marginal statistics of any single variable and any pair of two systems, providing six constraints, but we do not have enough constraints to fill the diagram. $H(X:Y:Z)$ is a part which is inaccessible. Nonetheless, we can combine our constraints independent of $H(X:Y:Z)$ and find that

$$\begin{aligned} H(X|Y,Z) - H(Y:Z|X) &= H(X) + H(Y:Z) - H(X:Y) - H(X:Z), \\ H(Y|X,Z) - H(X:Z|Y) &= H(Y) + H(X:Z) - H(X:Y) - H(Y:Z), \\ H(Z|X,Y) - H(X:Y|Z) &= H(Z) + H(X:Y) - H(X:Z) - H(Y:Z). \end{aligned} \tag{81}$$

Classically, all the entropies except $H(X:Y:Z)$ are required to be non-negative. It follows that

$$\begin{aligned} H(X:Y)+H(X:Z)-H(Y,Z) &\leq H(A), \\ H(X:Y)-H(X:Z)+H(Y,Z) &\leq H(B), \\ -H(X:Y)+H(X:Z)+H(Y,Z) &\leq H(C). \end{aligned} \tag{82}$$

These expressions already are similar to what we got in (77). Using the same method as before we can summarize these equations:

$$|H(X:Y)-H(X:Z)|+H(Y:Z)\leq 1. \tag{83}$$

This inequality only holds if the entropies can be described with Shannon's information theory and therefore are a type of a Bell inequality [21].

18 Entropy in Quantum Measurement

In von Neumann's formalism for quantum mechanics the measurement process involves an interaction between the system Q and a quantum measurement device (the ancilla) A.

Let $|a_i\rangle$ be the eigenvectors of an arbitrary observable we are measuring. The system is initially in the state

$$|Q\rangle = \sum_i \alpha_i\,|a_i\rangle\,. \tag{84}$$

The von Neumann measurement is represented by a unitary transformation that advances the original state $|Q0\rangle$ into

$$|QA\rangle = \sum_i \alpha_i\,|a_i, i\rangle\,, \tag{85}$$

with $\{|i\rangle\}$ being the eigenstates of the ancilla A. Von Neumann suggested that the measurement induced correlations between the System Q and the ancilla A. The measurement device can be imagined as a pointer, which shows the measurement result. Before the measurement the pointer shows "0" while after the measurment it points to one of its eigenstates $|i\rangle$.
If it is initially not the case that one $\alpha_i = 1$ and all the other vanish, then the system Q is in a state of superposition. Therefore the system is not in an eigenstate of the observable and the measurement device points at a superposition of its eigenvectors $|i\rangle$. The result is no longer deterministic. Von

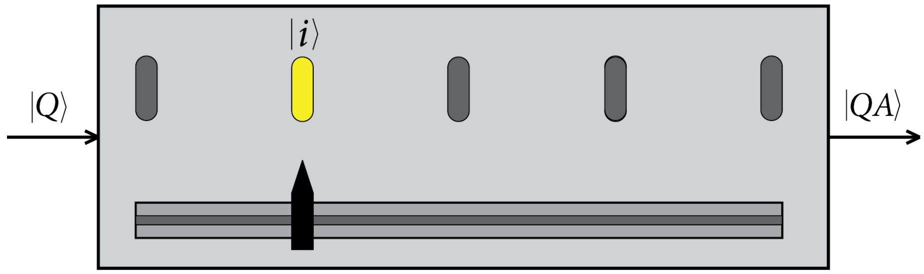


Figure 9: This Figure illustrates the second stage of the measurement according to John von Neumann. The wave function collapses and the measurement device points to one of its eigenstates. The situation is not as apparent when measuring a superposition of its eigenstates, because the pointer cannot point to a position between those states.

Neumann interpreted the measurement itself as the first stage, the second stage being the observation which is the collapse of the wave function and what you see on the measuring device is indeed one of the eigenstates $|i\rangle$.

The state (85) in fact describes an entanglement between the system and the ancilla. The two systems are not just correlated in classical formalisms. The entanglement is also the reason why it is not possible to clone an arbitrary quantum system, except for those which are in an eigenstate of the ancilla [22].

A basic problem occurs when Q is initially not in an eigenstate of the measurement device, i.e. in a superposition. The measurement pointer in Figure 9 would have to point between two of its eigenstates, but this is not possible. A way to overcome the difficulty of a superposition as a measurement result is to iterate the von Neumann measurement with another ancilla A' [20]. For simplicity the eigenstates A' are also denoted by $|i\rangle$. During the measurement, the initial state $|Q00\rangle$ evolves to

$$|QAA'\rangle = \sum_i \alpha_i \, |a_i, i, i\rangle \, . \tag{86}$$

The resulting state is now a pure state because it has undergone two unitary transformations and can be thought of a GHZ-state. Again, for pure states the entropy vanishes $S(Q, A, A') = 0$. Examining the ancillae AA' by performing the partial trace over Q, the reduced density matrix describes the mixed state

$$\rho_{AA'} = \mathrm{Tr}_Q[\rho_{QAA'}] = \sum_i |\alpha_i|^2 \left|i,i\right\rangle \left\langle i,i\right| . \tag{87}$$

It follows that the correlation is maximal, i.e.

$$S(A:A') = S(A) = S(A') = S(A,A'). \tag{88}$$

This correlation is fundamentally classical.

The second stage of the measurement consists of observing this correlation. Evidently, the system Q itself must be ignored in this process. This circumstance can be seen when looking at the mutual entropy,

$$S(Q:A:A') = S(A:A') - S(A:A'|Q) = 0. \tag{89}$$

After the measurement quantum and classical entropy no longer need to be distinguished, and therefore

$$S(A,A') = H(p_i), \quad p_i = |\alpha_i|^2. \tag{90}$$

$H(p_i)$ is the classical entropy associated with the probability distribution $p_i = |\alpha_i|^2$ which is predicted by quantum mechanics. This is because the system Q is initially in a pure state. Nevertheless, if you take the system Q into account conditional on A, A',

$$S(A,A') + S(Q|A,A') = S(Q,A,A') = 0, \tag{91}$$

the entropy is negative.

As an example consider the Bell state $|\Phi^+\rangle$ denoted as

$$|Q_1Q_2\rangle \equiv |\Phi^+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle). \tag{92}$$

The measurement process of each system entangles the system with the respective ancilla:

$$|Q_1Q_2A_1A_2\rangle = \frac{1}{\sqrt{2}}(|\uparrow\uparrow 11\rangle + |\downarrow\downarrow 00\rangle . \tag{93}$$

The ancillae A_1 and A_2 can be described by tracing over the systems Q_1 and Q_2:

$$\rho_{A_1,A_2} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|). \tag{94}$$

It is apparent that this is a completely mixed and classically correlated state as expected when considering GHZ-states. The ancillae are entangled individually with the respective systems Q_1 and Q_2.

Part III
Summary and Conclusion

As we have seen in Part I, various known phenomena in quantum mechanics can logically be reduced as a consequence of Zeilinger's foundational principle. It is possible to intuitively understand Heisenberg's Uncertainty relation because of mutually exclusive variables. Malus' Law for quantum mechanics has been deduced. Entanglement followed logically and we derived the time evolution of a quantum system using a formulation for the state vector of a system. Clearly, the results seem more natural when we are able to have a foundational principle in mind.

Part II deals with a different mathematical description of the information theoretic understanding of quantum physics. We briefly introduced the notion of Shannon entropy in classical information theory as well as the quantum physical von Neumann entropy and extended the formalism to understand quantum phenomena in terms of entropy. The entropy of entanglement in multipartite systems has been considered and entropic Bell inequalities have been obtained. Finally, the measurement process, especially when measuring a state in superposition was explained in terms of entropy.

Although one is tempted to think of quantum physics as a mechanical theory, it is more intuitive to use an information theoretic approach and it was shown that it is possible to make use of a mathematical formalism that gives the same results as expected and proves to be a simple and elegant way to understand quantum theory.

References

[1] Anton Zeilinger: *A foundational principle for quantum mechanics*, Foundations of Physics, Vol . 29, No. 4, 1999.

[2] `http://en.wikipedia.org/wiki/Interpretations_of_quantum_mechanics`, accessed: 04/17/2015.

[3] `http://bugman123.com/Physics/Schrodinger/DoubleSlit.gif`, accessed: 04/15/2015.

[4] `http://upload.wikimedia.org/wikipedia/commons/0/02/Doppelspalt.svg`, accessed: 04/15/2015.

[5] Rob Clifton, Jesffrey Bub, Hans Halvorson: *Characterizing quantum theory in terms of information-theoretic constraints*, Foundations of Physics 33, 1561-1591, 2003.

[6] Lucien Hardy, *Quantum Theory From Five Reasonable Axioms*, quant-ph/0101012, 2001.

[7] Robert Spekkens, *In defense of the epistemic view of quantum states: a toy theory*, quantph/0401052, 2004.

[8] Marek Zukowsi, Caslav Brukner: *Bell's theorem for general N-qubit states*, arXiv:quant-ph/0102039, 2002.

[9] John Clauser, Michael Horne, Abner Shimony, Richard Holt: *Proposed Experiment to Test Local Hidden-Variable Theories*, Phys. Rev. Lett. 23, 880, 1969.

[10] Caslav Brukner, Anton Zeilinger: *Information and Fundamental Elements of the Structure of Quantum Theory*, in "Time, Quantum, Information", edited by L. Castell and O. Ischebek, Springer, 2003.

[11] Albert Einstein, Boris Podolsky, Nathan Rosen, *Can Quantum-Mechanical Description of Physical Reality be Considered Complete*, Phys. Rev. 47, 777, 1935.

[12] John Bell, *On the einstein podalsky rosen paradox*, Physics, 1(3):195-200, 1964.

[13] Erwin Schrödinger: *Die gegenwärtige Situation in der Quantenmechanik*, Naturwissenschaften 23, 807 ff., 1935.

[14] Andrew Liddle: *An Introduction to Modern Cosmology*, 2nd. ed., John Wiley & Sons, p.2, 2003.

[15] Caslav Brukner, Marek Zukowski, Anton Zeilinger: *The essence of entanglement*, arXiv:quant-ph/0106119v1, 2001.

[16] Dani Greenberger, Michael Horne, Anton Zeilinger: *Going beyond Bell's theorem*, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, by M. Kafatos, ed. Kluwer Academic, Dordrecht, p. 69, 1989.

[17] William Wootters, Brian Fields, *A gap for the maximum number of mutually unbiased bases*, Ann. Phys. 191, 363, 1989.

[18] Michael Nielsen, Isaac Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, first published 2000, 10th anniversary edition, 2010.

[19] John von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, 1927.

[20] Nicholas Cerf, Christoph Adami, *Information theory of quantum entanglement and measurement*, Physica D 120, 62-81, 1998.

[21] Nicholas Cerf, Christoph Adami: *Entropic Bell Inequalities*, arXiv:quant-ph/9608047v2, 1997.

[22] William Wooters, Wojciech Zurek: *A single quantum cannot be cloned*, Nature 299, 802 - 803, 1982.