CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering

BACHELOR'S THESIS

2011 Dominik Šafránek

CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering

BACHELOR'S THESIS

Information and the Structure of Quantum Theory

Author: Dominik Šafránek
Supervisor: Ing. Petr Jizba, PhD.
Year: 2011

Acknowledgement

I would like to thank to my supervisor Ing. Petr Jizba, PhD., for all consultations, which really gave a lot, and especially for enthusiasm. I really admire that he was able to reply for all my inquisitive questions. I also thank to Prof. Ing. Pavel Štovíček, DrSc. for checking my proof of the Klein's inequality and valuable observations to it.

Prohlášení

Prohlašuji, že jsem svou bakalářskou práci vypracoval samostatně a použil jsem pouze literaturu uvedenou v přiloženém seznamu.

Nemám závažný důvod proti užití školního díla ve smyslu §60 Zákona č 212/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

Declaration

I declare, I wrote my Bachelor's Thesis independently and exclusively with the use of cited bibliography.

I agree with the usage of this thesis in the purport of the §60 Act 121/2000 (Copyright Act).

V Praze dne ....

Název práce:	**Informace a struktura kvantové teorie**
Autor:	Dominik Šafránek
Obor:	Matematické inženýrství
Druh práce:	Bakalářská práce
Vedoucí práce:	Ing. Petr Jizba, PhD. Katedra fyziky, Fakulta jaderná a fyzikálně inženýrská

Abstrakt

Informační teorie a kvantová teorie jsou dnes jedny z nejvíce rozvíjených oblastí. V první části představujeme a porovnáváme tři typy klasických informací a popisujeme některé klasické entropie. V druhé části definujeme qubit a von Neumannovu entropii a dokazujeme Kleinovu nerovnost (tj. nezápornost kvantové relativní entropie), zvláště pak podmínky rovnosti, jejíž důkaz není vždy prezentován korektně, např. v [13]. Ve třetí části zkoumáme Delayed Choice Quantum Eraser experiment a navrhujeme kvantový test svobodné vůle.

Klíčová slova: informace, kvantová teorie informace,
entropie, Kleinova nerovnost, kvantová provázanost,
Delayed Choice Quantum Eraser experiment

Title:	**Information and the Structure of Quantum Theory**
Author:	Dominik Šafránek

Abstract

Information theory and quantum theory are presently fast developing scientific fields. In the first part of this thesis we introduce and compare three types of classical information measures and describes some of classical entropies. In the second part we define qubit and von Neumann entropy and prove Klein's inequality (i.e., non-negativity of the relative quantum entropy), and especially the equality conditions, which is not always presented correctly in literature, see e.g. Ref. [13]. In the third part we scrutinize the Delayed Choice Quantum Eraser experiment and suggest possible free will quantum test.

Keywords: information, quantum theory of information
entropy, Klein's inequality, entanglement,
Delayed Choice Quantum Eraser experiment

Contents

Introduction

Why I chose this theme? Well, it is because I was very interested in quantum theory and especially in extraordinary behavior of the quantum particles. I was too interested in a fashion, how is the quantum mechanics treated, and wanted to understand it better. My supervisor told me that there exist a new trend to regard the quantum theory as the theory of information. My supervisor also did not tell me, that the quantum theory as a theory of information very differs from the original theory of information and quantum theory of information too. If I exaggerate a little, the only thing which is common is the name. But I don't regret it at all. This misunderstanding opened me a beautiful domains of the theory of communication, quantum computation, quantum optics and many others. The other motivation for this work is rapid developing in these fields. For example, this year has been successfully performed quantum teleportation of wave packets of light [11] and also 14-bits entangled state has been created [12].

In the first chapter I will deal with three types of the information and I also introduce Shannon entropy, where I combine different approaches of Shannon and Khinchin, and it's derivations. In the second chapter I, after necessary mathematical preliminary, define and explain the notion of Qubit, compare Qubit with the classical bit and introduce von Neumann entropy, which is in a sense the quantum brother of the Shannon entropy. I also correctly prove Klein's inequality (non-negativity of the relative quantum entropy), whose proof is not presented correctly in [13]. In the third chapter I analyze the entanglement, probably the most astonishing topic in the quantum mechanics. In the last chapter I discuss a double slit experiment, there I highlight the most important concepts, which helps us to understand it properly. I will also scrutinize the Delayed Choice Quantum Eraser experiment adopted from [4] and suggest free will quantum test.

Chapter 1

Information Theory

Information theory is one of the most useful tool nowadays. In era of communication, where over 2 billion people use the Internet, where sending a message over the Atlantic ocean is approximately 17 000 000 000× faster than it was 500[1] years ago, has coding, storing and encrypting information extraordinary significance.

Nevertheless, there is not much space for being concerned with these three applications. We rather aim at basic definitions and their meaning. For better understanding we will combine in the following different approaches of Shannon and Khinchin. Apart from this we also introduce 3 types of post-Shannonian information.

1.1 What is the information?

Each of us have a rough understanding what the information is. It is something necessary in our everyday life, something we cannot live without. Necessity of being surrounded by information results from our nature — people need to communicate.

There are variety of types of information. Some of the information is useless, some is useful and some information, which is useful for one, don't have to be useful for other. A question arises; Does there exist any objective way to measure the information? Well, measuring information by its (subjective) usefulness is not a good idea. For instance, how would we operationally quantify information present in the statement: "Attach hot water line to 90° elbow and route underneath." (*semantic information*)? Semantic information play an important role in everyday life but as yet cannot be successfully

[1] The journey of Christopher Columbus lasted 62 days.

quantified. Now we get to the type of information, which we can quantify. For this reason I'll give here more examples.

- Suppose that all of a sudden we got a taste of banana. We are running to the shop, meanwhile, we remember that we have forgotten a purse. That does not matter, because we have some coins in our pocket. That may suffice! Arriving the shop we fish out of pocket a few coins, but we do not have enough, because only bananas from Martinique arrived and Martinique bananas cost more than those ordinary from Colombia. If we had known!
- Suppose only three people know that next Tuesday will be the new Black Friday[2]. What is this information worth?
- Suppose a meteorite with 5 km in diameter is hurtling to Earth, nevertheless we don't know it yet. What a pity!

These examples look differently, but still have something in common. We can connect each example with its corresponding probability. We are not too surprised that something like banana misfortune happened. This kind of things happens everyday, so importance of this information is quite negligible. Assume that you are stockbroker and one of the three people mentioned in the second example is your very best friend. You believe him so much. And this friend gives you a tip. You sell everything and save billions! Although we believe that knowledge about third event is much more important. [3]

After these considerations we conclude that information can be defined as functional I on an event space X, i. e. $I : X \to \mathbb{R}$ (in accordance with [3]). As we have seen, the importance of an information depends on probability with which the event may happen. We require for two events $x_1, x_2 \in X$ such that $p(x_1) < p(x_2)$, we have $I(p(x_1)) > I(p(x_2))$. We also require intuitive limit conditions $(p(x) \to 1 \Rightarrow I(p(x)) \to 0)$ and $(p(x) \to 0 \Rightarrow I(p(x)) \to \infty)$. An example of such function is the following.

Definition 1.1. *We define an information measure (Hartley's information)* $I : X \to \mathbb{R}$

$$\forall x \in X, \quad I(x) = -\log_a x, \tag{1.1}$$

where $a > 1$ *is a parameter.*

Above definition is ingenious in the sense that an amount of information measured does not depend on actual events (which the information represent) but only the ensuing probability distribution. This allows to compare

[2] Wall Street Crash of 1929

[3] We also believe that some stockbrokers might disagree.

between different events without involving our feelings, knowing only the probabilities. The base of logarithm can be chosen arbitrarily, but for our purposes we choose $a = 2$. The unit of such $I(x)$ is the called *bit*[4]. 1 *bit* of information tells us that event x has exactly 50% chance to happen, in other words, 1 *bit* is an information hidden in an answer to the binary (i.e., YES–NO) question. Further we will write shortly only log instead of $\log_2$.

As we shall se in the next section, the definition of information measure is fundamental for the information theory and allows to find a way of optimal coding, minimum physical source for storing an information, error correction and more (see, e.g., Ref.[14]).

It is necessary to stress that the above definition gives only one specific information measure — the so-called *syntactic* information, which is not the only possible information measure. The other is, for instance, an *algorithmic* information, sometimes called after its founder the Kolmogorov complexity. This information measure is quite different. We will compare them in section 1.4.

1.2 Entropy

Now we will introduce the main function used in Information theory. This function and its derivations appear in an almost all of the theorems. The heart of the theory is called Entropy.

Definition 1.2. *(Shannon entropy) Suppose that* $X = \{x_1, x_2, ...x_n\}$ *is an event source with corresponding probabilities of occurence* $p_1, p_2, ..., p_n$, *where* $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$. *We define entropy* $H(p_1, p_2, ..., p_n)$ *as a function satisfying following properties:*

1. *H is a continuous in the p_i.*

2. *If all the p_i were equal, namely $p_i = \frac{1}{n}$* for all *i, H is monotonously increasing function of n.*

3. *If any occurrence breaks down into two successive possibilities, H should break down into weighted sum of corresponding individual values of H.*

The above definition was obtained from [3], but was primarily introduced by Shannon in his founding paper from 1948 [14]. Well, we see the definition

[4] The unit is added artificially in order to tell us, which parameter do we use. For example if $a = 2.718281828...$ (Euler's number) the logarithm becomes natural and the unit added becomes *nat*

and ask ourselves, “Does there exist function satisfying 1.–3.?” Yes, such function exists and it must be of the form[5]

$$H = -K \sum_{i=1}^{n} p_i \log p_i, \tag{1.2}$$

where K is an arbitrary positive scaling constant. If we choose $K = 1$, (1.2) becomes mean value of information

$$H = \langle I(x) \rangle = -\sum_{x \in X} p(x) \log p(x) = -\sum_{i=1}^{n} p_i \log p_i. \tag{1.3}$$

This is it what entropy really means. It is an average information hidden in an event source X or, in other words, it is an average amount of information what we get if we ask, “Which event will happen?”. It is sometimes said that the entropy is an amount of uncertainty what we have about the source, or what information the source can provide. Entropy with $K = 1$ and logarithm base $a = 2$ is called Shannon entropy. However we can choose K differently. For example if we put $K = \frac{k_B}{\log e}$, where k_B is Boltzmann constant and e Euler's number, (1.2) changes into

$$H = -k_B \sum_{i=1}^{n} p_i \ln p_i \tag{1.4}$$

which is the so-called Gibbs–Boltzmann entropy. Gibbs–Boltzmann entropy is the basic building block in the statistical-thermodynamics[6], but we will not pursue this issue here.

The first two requirements in the definition of the entropy are easily understandable, the last requirement is less obvious[7]. For a better understanding, we can arrive at the very same entropy using different axioms, which were firstly found by Khinchin [9].

Definition 1.3. *A continuous function with respect to all arguments $H(p_1, p_2, ..., p_n)$ is called (Shannon's) entropy if it satisfies following properties:*

1. *For a given n and for $\sum_{i=1}^{n} p_i = 1$, the function takes its largest value for $p_i = \frac{1}{n}$, $(i = 1, \ldots, n)$.*

[5] Proof can be found in [3]

[6] The second law of thermodynamics states that the entropy of an isolated macroscopic system never decreases.

[7] Even though it is well explained in the Shannon's founding paper [14].

2. *$H(X,Y) = H(X) + H(Y|X)$, where X, Y is the joint event source and $H(Y|X) = \sum_{i=1}^{n} p_i H_i = \sum_{i=1}^{n} p(X = x_i) H(Y|X = x_i)$ conditional entropy.*[8]

3. *$H(p_1, p_2, ..., p_n, 0) = H(p_1, p_2, ..., p_n)$ (adding an impossible event does not change the entropy).*

We would like to add that these definitions of entropy are the special cases of the so-called Rényi's entropy. To see it, it is appropriate to rewrite (1.2) as

$$H = \sum_{i=1}^{n} p_i I(p_i). \tag{1.5}$$

This can be viewed as the special case of the *general averaging* which is defined

$$H_f = f^{-1}\left(\sum_{i=1}^{n} p_i f\left(I(p_i)\right)\right). \tag{1.6}$$

The function f is the so-called Kolmogorov–Nagumo function which must be, of course, invertible and if we demand additivity and continuity of information with respect to all arguments of such entropy, only two classes such functions are possible [8]. First is the identity $f(x) = x$, which gives the Shannon entropy and the second possible is exponential, i. e., $f(x) = 2^{(1-\alpha)x}$, which gives

$$H_{[\alpha]} = \frac{1}{(1-\alpha)} \log\left(\sum_{i=1}^{n} p_i^{\alpha}\right). \tag{1.7}$$

Another frequently used entropy is Tsallis entropy, which takes the form

$$\mathfrak{S}_q = \frac{1}{(1-q)}\left(\sum_{k=1}^{n} (p_k)^q - 1\right), \tag{1.8}$$

where q is a positive parameter. In the $q \longrightarrow 1$ limit Tsallis entropy reduces to Shannon entropy. In this case classical additivity of independent information is replaced by so-called pseudoadditivity

$$\mathfrak{S}_q(AB) = \mathfrak{S}_q(A) + \mathfrak{S}_q(B|A) + (1-q)\mathfrak{S}_q(A)\mathfrak{S}_q(B|A), \tag{1.9}$$

where $\mathfrak{S}_q(B|A)$ represents the conditional Tsallis entropy. Tsallis entropy has one more interesting feature. It is a monotonic function of the Shannon entropy and thus they reaches maximum at the same point $p_i = 1/n$ for all i.

For further discussion of both Rényi and Tsallis entropy see, e.g., Ref. [8].

[8] $p(X = x_n)$ is a probability that event x_i happen and $H(Y|X = x_i)$ is an entropy of the event source B provided that event x_i happened.

1.3 Joint and conditional entropies, mutual information, measuring distance

In the previous section we have learned how to define various information entropies but now we are going to discuss more interesting issue, namely we will compare different probability distributions and features of the entropies. We have met the joint and conditional entropy in the alternative definition of the entropy already (see Def. 1.3). Now we introduce these notions correctly.

Definition 1.4. *Let X and Y be event sources with the joint probability distribution $P = \{p(x,y)|x \in X, y \in Y\}$. Then we define joint entropy*

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

conditional entropy[9]

$$\begin{aligned} H(X|Y) \;&=\; -\sum_{x\in X, y\in Y} p(x,y)\log p(x|y) \;=\; H(X,Y) - H(Y) \\ &=\; \sum_{y\in Y} p(y) H(X|Y=y)\,, \end{aligned}$$

and mutual information

$$\begin{aligned} H(X;Y) \;&=\; \sum_{x\in X, y\in Y} p(x,y)\log \frac{p(x,y)}{p(x)p(y)} \;=\; H(X) + H(Y) - H(X,Y) \\ &=\; H(X) - H(X|Y)\,. \end{aligned}$$

We have used the Bayes' theorem to derive the other expressions of each definition. We can see that the definitions are closely related. It is obvious that joint entropy $H(X,Y)$ and mutual information $H(X;Y)$ are symmetric in its arguments X and Y. The second expression of conditional entropy says that conditional entropy is an average amount of information what we can get from joint system X,Y if we have total knowledge about source Y, the third says that conditional entropy is also a mean value of entropy $H(X)$. Mutual information is an amount of information what the sources X and Y have in common.

In order to prove the other properties of entropies we need to introduce following theorem.

[9] $H(X|Y=y) = -\sum_{x\in X} p(x|y)\log p(x|y)$.

Theorem 1.1 (Jensen's inequality).

$$f(\lambda_1 z_1 + \cdots + \lambda_n z_n) \leq \lambda_1 f(z_1) + \cdots + \lambda_n f(z_n),$$

where f is a convex function on some set $\mathbb{C}$ in $\mathbb{R}$, $z_i \in \mathbb{C}$, $\lambda_i > 0, i = 1, ..., n$ and $\lambda_1 + \cdots + \lambda_n = 1$. For strictly convex functions equality holds if and only if $z_1 = \cdots = z_n$.

Theorem 1.2 (Basic properties of Shannon entropy).

1. *$H(X, Y) = H(Y, X)$, $H(X; Y) = H(Y; X)$*
2. *$H(X|Y) \geq 0$ and thus $H(X; Y) \leq H(X)$ with equality if and only if X is a function of Y.*
3. *$H(X; Y) \geq 0$ and thus $H(X, Y) \leq H(X) + H(Y)$ with equality if and only if X, Y are random variables. (subaditivity)*

Proof.

1. Obvious.
2. $$H(X|Y) = -\sum_{x \in X, y \in Y} p(x, y) \log p(x|y) \geq 0\,,$$

   because $(0 \leq p(x|y) \leq 1)$. Let $H(X|Y) = 0$. Then $(\forall x, y \in X, Y)$ $(p(x, y) \log(p(x|y)) = 0)$. So either[10] $p(x, y) = 0$ or $p(x|y) = 1$. That is, if event y happen then either x happen with certainty or do not happen at all, i.e., $X = f(Y)$. Let $X = f(Y)$. Then $p(x|y)$ is equal to either 1 or 0, i.e., $H(X|Y) = 0$.
3. In theorem 1.1 we put $f = -\log$, $\lambda_{x,y} = p(x, y)$, $z_{x,y} = \frac{p(x)p(y)}{p(x,y)}$. Then we have

   $$\begin{aligned} H(X; Y) &= -\sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x)p(y)}{p(x, y)} \\ &\geq -\log \left(\sum_{x \in X, y \in Y} p(x, y) \frac{p(x)p(y)}{p(x, y)} \right) = 0\,, \end{aligned}$$

   with equality if and only if

   $$\frac{p(x)p(y)}{p(x, y)} = \frac{p(\tilde{x})p(\tilde{y})}{p(\tilde{x}, \tilde{y})} = q \text{ for all } x, \tilde{x} \in X, y, \tilde{y} \in Y.$$

   We can multiply the equation by $p(x, y)$ and summarize with respect to all x and y. We get $q = 1$. That means that X and Y are independent.

[10] We define $0 \log 0 = 0$.

□

Now we introduce relative entropy between two probability distributions on the same event set. This is not only a natural distance measure but it will also help us to find an upper bound of the entropy.

Definition 1.5. *Let the P, Q be probability distributions on the same event set X. We define relative entropy (or equivalently Kullback-Leibler divergence) between distributions P, Q as*

$$D(P||Q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} = \left\langle \log \frac{P(x)}{Q(x)} \right\rangle_P = -H(x) - \langle \log Q(x) \rangle_P .$$

As we can see from the definition, the relative entropy cannot be a true metric because it is not symmetric. However it satisfies another important property of a distance metric.

Theorem 1.3. *$D(P||Q) \geq 0$, $D(P||Q) = 0$ if and only if $P = Q$.*

Proof. In theorem 1.1 we put $f = -\log,\ \lambda_x = p(x),\ z_x = \frac{q(x)}{p(x)}$. □

Theorem 1.4. *$H(X) \leq \log N$*

Proof. In theorem 1.3 we put $q(x) = \frac{1}{N}$ for all $x \in X$. □

The previous theorem shows that maximum entropy is achieved only for uniform distribution, if no extra constraints are imposed. It corresponds to a following assumption: If we have absolutely no idea about the event source, the best way how to estimate the distribution is to assign the same probability to each event. In other words, uncertainty about the event source rises if and only if the probability distribution approaches the uniform.

At the end we would like to stress that so far we have been concerned only with finite event sources. Everything can be generalized to an infinite but still discrete sources. We can make also similar theory for continuous event sources but many theorems differ and some do not hold at all, see, e.g., Ref. [3].

1.4 Kolmogorov complexity

One application of the information theory, which I mentioned in Introduction, is optimizing coding efficiency. Consider we want to send the sentence "I like the information theory." What does our computer do? It decompose this sentence in the sequence of 1's and 0's. Because each character in a language has a different frequency of occurrence, sending a message can be optimized. We assign shorter expressions (of 1's and 0's) to frequently used characters and longer for rarely used. In this case, the information theory grasps the sentence like a random sequence, or better, like a random event source. As we can imagine, the language is not completely random, but this approach works pretty well in many practical situations. Nonetheless, this is not the only reasonable approach. We can consider the sentence as a unique sequence of fixed characters and ask whether there exist any program, which could generate it. If the program is shorter than the original sentence, a compression is achieved. The length of the shortest of such programs is called *Kolmogorov complexity* or equivalently *algorithmic information.*

Definition 1.6. *The Kolmogorov complexity $K(x)$ is the shortest size of a program $q(x)$ necessary to generate the sequence x. Such program is a finite set of binary instructions with a length $|q(x)|$ bits.*

For illustration, consider that we have a very silly sense of humor and want to send to a friend an SMS consisting of thousand 01's, i.e.,

$$0101010101010101010101010101010101010101....$$

It would be quite expensive. We may invent a better way — we send him a program, which generates the message automatically. Such program (in a binary code) is surely much shorter and thus cheaper. On the other hand, finding the shortest program is generally a complicated and as yet unsolved problem.

The complexity is, in contrast with syntactic information, an information hidden in the sequence. For instance, an entity with huge Kolmogorov complexity is the DNA. In the DNA is hidden an almost complete information about the person who writes this article.

The definition of Kolmogorov complexity was obtained from [3], where can also be found an implementation of the program generating the sequence by the Turing machine. Comparison of syntactic and algorithmic information with an interesting philosophical reasoning is made in [5].

Chapter 2

Quantum information theory

Quantum information theory paves the way to the modern type of computation and coding. Main motivation for exploring this field are quantum computer, which is supposed to be at least as powerful as the classical one and exists some indications that some types of problems can be solved exponentially faster by a quantum computer. Further important motivations include quantum cryptography and quantum teleportation.

There is not much place for a comprehensive insight into quantum information theory, we rather concentrate onto proper definition and derivations of the main actor — von Neumann entropy.

2.1 Formalism and basic theorems in quantum mechanics

The quantum mechanics and quantum theory in general are expressed in a language of mathematics. The specificity of the quantum theory is that the mathematical formalism is necessary for understanding even the most basic notions and in fact, it forms the only guiding principle of the theory.

The quantum theory is constructed on Hilbert spaces — complete vector spaces with, in our case naturally defined, inner product. Each vector in Hilbert space represents one state and is denoted as $|\psi\rangle$, called *ket.* The inner product of two vectors $|\psi\rangle$, $|\varphi\rangle$ is then denoted as $\langle\psi|\varphi\rangle$, known as *bra(c)ket*[1]. A linear functional $\langle\varphi| : |\psi\rangle \longrightarrow \langle\varphi|\psi\rangle$ is called *bra.* This bracket formalism facilitates the work. I will show only basic definitions and theorems necessary in the following. For a comprehensive presentation see [1].

[1] We define inner product linear in the second argument and antilinear in the first.

Let us suppose that dimension of the Hilbert space is finite. This assumption will allow us to simplify many presented arguments. Many theorems holds even if the dimension is infinite, but some definitions have to be slightly modified and we have to bother with domains of the operators. This presumption is also very restrictive in the quantum theory, but in quantum computation no computer with an infinite number of (qu)bits is possible.

From Riesz representation theorem we know that for each operator A on a Hilbert space $\mathbf{H}$ exists unique operator $A^\dagger$ such that

$$\forall|\psi\rangle,|\varphi\rangle \in \mathbf{H}, \ \ \langle\psi|A\varphi\rangle = \langle A^\dagger\psi|\varphi\rangle.$$

This operator is called adjoint. Some of the operators have special properties that are important for a conceptual development of quantum theory. These are frequently defined through the use of the adjoint operators.

Definition 2.1. *A linear operator is called normal if and only if*

$$AA^\dagger = A^\dagger A,$$

hermitian if

$$A^\dagger = A,$$

unitary if

$$AA^\dagger = I = A^\dagger A,$$

positive if

$$\forall|\psi\rangle \in \mathbf{H}, \ \ \langle\psi|A\psi\rangle \geq 0.$$

We define $|\psi\rangle^\dagger = \langle\psi|$ and $\langle\psi|^\dagger = |\psi\rangle$. Because of an isomorphism between Hilbert space and a space of linear functionals, all properties of the † remain the same. The reason for this notation is that then $(A|\psi\rangle)^\dagger = \langle A\psi| = \langle\psi|A^\dagger$.

Theorem 2.1. *Hermitian and unitary operators are normal, positive operator is hermitian.*

The unitary operators have one important property. In a space of a finite dimension only the unitary operators preserve the inner product. It ensures that the sum of probabilities of all possible outcomes of any event in quantum mechanics is always 1.[2] Therefore, the evolution in physical system should be described by an unitary operator.

Theorem 2.2. *Operator U is unitary if and only if*

$$\forall|\psi\rangle,|\varphi\rangle \in H, \ \ \langle U\psi|Uy\rangle = \langle\psi|y\rangle.$$

[2] It should be stressed that this is true also for anti-unitary operators but these are not considered in this thesis.

Proof. $\Rightarrow$: $\langle U\psi|U\varphi\rangle = \langle\psi|U^\dagger U\varphi\rangle = \langle\psi|\varphi\rangle$

$\Leftarrow$: $\forall|\psi\rangle, |\varphi\rangle \in H$, $\langle\psi|U^\dagger U\varphi - \varphi\rangle = 0$. We put $|\psi\rangle = U^\dagger U|\varphi\rangle - |\varphi\rangle$. Then for all $|\varphi\rangle$ must be $U^\dagger U|\varphi\rangle - |\varphi\rangle = 0$, it means the operator is unitary. □

The subclass of the linear operators — *orthogonal projectors* is closely related to bra-vectors. A projector onto subspace spanned by normalized[3] vector $|i\rangle$ is denoted $|i\rangle\langle i|$ and defined for all $|j\rangle \in \mathbf{H}$ as $|i\rangle\langle i|(|j\rangle) = |i\rangle\langle i|j\rangle$. A linear operator A is diagonalizable if and only if exist λ_i, $|i\rangle\langle i|$ such that $A = \sum_i \lambda_i |i\rangle\langle i|$.

Each linear operator has it's own characteristic called eigenvalues, which occupy a special position in a quantum mechanics. The eigenvalues of the operator are the only values, which you can measure, and after the measurement is done the state of a particle passes to the eigenvector corresponding to the eigenvalue.

Definition 2.2. *Eigenvalue λ of a linear operator A is a value for which exists a non-zero vector such that $A|\psi\rangle = \lambda|\psi\rangle$. The vector $|\psi\rangle$ is then called eigenvector of the eigenvalue λ.*

Each observable in a quantum system is expressed as a hermitian linear operator. The reason why is that the measured values (eigenvalues) are real and thus we know how to interpret them.

Theorem 2.3. *Hermitian operator has real eigenvalues.*

Proof. Let $|\psi\rangle$ be an eigenvector with an eigenvalue λ. Then

$$\lambda\langle\psi|\psi\rangle = \langle\psi|A\psi\rangle = \langle A\psi|\psi\rangle = \overline{\lambda}\langle\psi|\psi\rangle.$$

□

Now I will introduce widely used theorem for normal operators. Proof can be found in [13].

Theorem 2.4 (Spectral Decomposition)**.** *Let $\mathbf{H}$ be a Hilbert space of a finite dimension. Any normal operator A on $\mathbf{H}$ is diagonal with respect to some orthogonal basis for $\mathbf{H}$. That is*

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

$|i\rangle$ are normalized eigenvectors of the operator A with corresponding eigenvalues λ_i. Conversely, any diagonalizable operator is normal.

[3] The state normalization is defined here and throughout as $\||i\rangle\| \equiv \sqrt{\langle i|i\rangle} = 1$.

We use the theorem promptly for a construction of the new operators.

Definition 2.3. *Let the $A = \sum_i \lambda_i |i\rangle\langle i|$ be a spectral decomposition of the operator A, f a continuous function on $\mathbb{C}$. Then we define operator function*

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

The operator function does not depend on spectral decomposition vectors and thus is uniquely defined.

We often want to work with more than one particle. The mathematical procedure for this is to make tensor products of one particle systems. Generally, if we have two Hilbert spaces $\mathbf{H}, \mathbf{G}$, $dim(\mathbf{H}) = m$, $dim(\mathbf{G}) = n$ and $\{|i\rangle,\ i = 1, ..., m\}$ is an orthonormal basis for $\mathbf{H}$, $\{|j\rangle,\ j = 1, ..., n\}$ an orthonormal basis for $\mathbf{G}$, then tensor product of the spaces $\mathbf{H}, \mathbf{G}$ is also Hilbert space $\mathbf{H} \otimes \mathbf{G}$, $dim(\mathbf{H} \otimes \mathbf{G} = m \cdot n)$ and $\{|i\rangle \otimes |j\rangle\}$ is an orthonormal basis for $\mathbf{H} \otimes \mathbf{G}$. Tensor product of vectors satisfies the following basic properties:

Theorem 2.5. $\forall \alpha \in \mathbb{C},\ \ \forall |\psi\rangle \in \mathbf{H},\ \ \forall |\varphi\rangle \in \mathbf{G}$

1. $\alpha(|\psi\rangle \otimes |\varphi\rangle) = (\alpha|\psi\rangle) \otimes |\varphi\rangle = |\psi\rangle \otimes (\alpha|\varphi\rangle)$
2. $(|\psi_1\rangle + |\psi_2\rangle) \otimes |\varphi\rangle = |\psi_1\rangle \otimes |\varphi\rangle + |\psi_2\rangle \otimes |\varphi\rangle$
3. $|\psi\rangle \otimes (|\varphi_1\rangle + |\varphi_2\rangle) = |\psi\rangle \otimes |\varphi_1\rangle + |\psi\rangle \otimes |\varphi_2\rangle$

For simplicity it is often written $|\psi\rangle|\varphi\rangle$ or $|\psi\varphi\rangle$ only instead of $|\psi\rangle \otimes |\varphi\rangle$ and we will use this notation too.

An inner product on $\mathbf{H} \otimes \mathbf{G}$ is defined naturally:

$$\langle \sum_i \alpha_i |\psi_i\rangle \otimes |\varphi_i\rangle | \sum_j \tilde{\alpha}_j |\tilde{\psi_j}\rangle \otimes |\tilde{\varphi}_j\rangle\rangle = \sum_{i,j} \overline{\alpha_i}\tilde{\alpha}_j \langle\psi_i|\tilde{\psi_j}\rangle\langle\varphi_j|\tilde{\varphi}_j\rangle.$$

We can generalize operators on a tensor product of two spaces.

Definition 2.4. *Let A be a linear operator on a space $\mathbf{H}$, B a linear operator on $\mathbf{G}$. We define*

$$\forall |\psi\rangle \otimes |\varphi\rangle \in \mathbf{H} \otimes \mathbf{G},\ \ (A \otimes B)(|\psi\rangle \otimes |\varphi\rangle) = (A|\psi\rangle) \otimes (B|\varphi\rangle),$$

specially

$$A(|\psi\rangle \otimes |\varphi\rangle) = (A|\psi\rangle) \otimes |\varphi\rangle,\ \ B(|\psi\rangle \otimes |\varphi\rangle) = |\psi\rangle \otimes (B|\varphi\rangle).$$

Theorem 2.6. *Tensor product of two unitary operators is unitary, of two hermitian operators is hermitian, of two positive operators is positive.*

The last notion, which we need to introduce, is the trace of an operator A.

Definition 2.5. *Let $\{|i\rangle\}$ be an orthonormal basis of a Hilbert space $\mathbf{H}$, A a linear operator on $\mathbf{H}$. The trace of the operator A is*

$$\mathrm{Tr}(A) = \sum_i \langle i|A|i\rangle.$$

The trace does not depend on a choice of the orthonormal basis and has following properties.

Theorem 2.7. *Let A, B be linear operators, U unitary operator, $\alpha \in \mathbb{C}$. Then*

1. $\mathrm{Tr}(\alpha A + B) \;=\; \alpha\mathrm{Tr}(A) + \mathrm{Tr}(B)$ *(linearity)*
2. $\mathrm{Tr}(AB) \;=\; \mathrm{Tr}(BA)$ *(symmetry)*
3. $\mathrm{Tr}(UAU^\dagger) \;=\; \mathrm{Tr}(AU^\dagger U) = \mathrm{Tr}(A)$ *(conservation in time)*
4. $|\mathrm{Tr}(A^\dagger B)|^2 \;\leq\; \mathrm{Tr}(A^\dagger A)\mathrm{Tr}(B^\dagger B)$ *(Schwarz inequality) The last is Schwarz inequality for an inner product defined on the Hilbert–Schmidt operator space*

   $$\langle A|B\rangle \;=\; \mathrm{Tr}(A^\dagger B) \;=\; \sum_i \langle i|A^\dagger B|i\rangle\,,$$

   where $\{|i\rangle\}$ is an orthonormal basis.

2.2 Qubit

Quantum bit, or for short qubit, is a basic operational unit in the quantum computation and quantum information theory. The classical bit is an information hidden in the event, which has exactly 50% chance to happen. But, as you probably noticed, there exist one different point of view what the bit is. We can apprehend one bit as an element of a set $\{0, 1\}$. Now the bit is not the information hidden in an event, but the event itself. In the computation science, maximal information transfer is achieved, when occurences of events from $\{0, 1\}$ are equal[4], that means both 0 and 1 have 50% chance to happen. This is the point where it corresponds with the former definition. The classical bit can take values of 0 or 1 only, but in quantum mechanics, any linear combination from a set is also the element of the set. In other words, qubit is an element of a linear span of a set $\{|0\rangle, |1\rangle\}$.

[4] Because binary entropy reaches maximum at $p(0) = \frac{1}{2}$, $p(1) = \frac{1}{2}$, see theorem 1.4.

Definition 2.6. *Qubit is the normalized linear combination of the orthogonal states* $\alpha|0\rangle + \beta|1\rangle$*, where* $|\alpha|^2 + |\beta|^2 = 1$*.*

What is the physical representation of a qubit? The simplest is the orientation of spin. Orthogonal states of a spin corresponds to the 2 arbitrary orthogonal directions in space. Qubit differs from classical bit in a very important feature with profound consequences. In general we cannot distinguish between two qubits. If we could, faster-than-light communication is possible.[5] Measuring gives us only the probabilistic predictions. Only if we know that two qubits are orthogonal to each other and in what directions are oriented, we can distinguish between them. This is the only way how to represent a classical bit in a form of the qubit — we assign state $|0\rangle$ to 0 and state $|1\rangle$ to 1. Furthermore, general qubit cannot be copied[6], but can be teleported at expense the original state is destroyed[7].

2.3 von Neumann entropy

How to introduce an entropy in the quantum mechanics? Well, the entropy is still measure of uncertainty about the system, so it should include probabilities in some way. It should also include states, because these are the special constituents of the quantum mechanics. If they were not there, what else should be included in order for we could call the entropy quantum?

Suppose we have a box and exactly defined states $\{|\psi_j\rangle\}$ in it. We fish out one state. What is the probability the state is actually $|\psi_j\rangle$? We define it p_j. The notion which characterizes collective state in the box is a density operator.

Definition 2.7. *Let* $\{|\psi_j\rangle | j = 1, ..., m\}$ *be a set of normalized vectors,* $\{p_j \in \mathbb{R} | j = 1, ..., m\}$ *such that* $\sum_j p_j = 1$*. Then we define density operator*

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|. \tag{2.1}$$

Definition 2.8. *We call physical state pure if and only if dim*$(\mathrm{Ran}(\rho)) = 1$*, mixed if dim*$(\mathrm{Ran}(\rho)) > 1$.[8]

Theorem 2.8 (basic properties of a density operator)**.**

[5] We will talk about it in chapter 3.

[6] This, so-called, non-cloning theorem and is analyzed, e.g., in [13].

[7] Quantum teleportation.

[8] By $\mathrm{Ran}(\rho)$ we mean the range of values of the operator ρ.

1. *Density operator is positive.*

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

3. *The state is pure* $\Leftrightarrow \rho = \rho^2 \Leftrightarrow \mathrm{Tr}(\rho - \rho^2) = 0 \Leftrightarrow \mathrm{Tr}(\rho^2) = 1$.

von Neumann entropy is a measure of uncertainty which we have about the mixed state.

Definition 2.9. *Let ρ be a density operator. Von Neumann entropy of the operator ρ is*

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho). \tag{2.2}$$

Now we derive another formulation of the Von Neumann entropy. Since the density operator is positive, it can be spectral decomposed (theorem 2.4). Let

$$\rho = \sum_i \lambda_i |i\rangle\langle i|$$

be a spectral decomposition of the operator ρ. From the definition 2.3 we know that

$$\rho \log \rho = \sum_i \lambda_i \log \lambda_i |i\rangle\langle i|. \tag{2.3}$$

Then

$$\begin{aligned} S(\rho) &= -\mathrm{Tr}(\rho \log \rho) \\ &= -\sum_{\tilde{i}} \langle \tilde{i}| (\sum_i \lambda_i \log \lambda_i |i\rangle\langle i|) |\tilde{i}\rangle \\ &= -\sum_{i,\tilde{i}} \langle \tilde{i}| (\lambda_i \log \lambda_i) |i\rangle\langle i|\tilde{i}\rangle \\ &= -\sum_i \lambda_i \log \lambda_i \sum_{\tilde{i}} |\langle \tilde{i}|i\rangle|^2 \\ &= -\sum_i \lambda_i \log \lambda_i \|i\|^2 \\ &= -\sum_i \lambda_i \log \lambda_i, \end{aligned} \tag{2.4}$$

where the $\{|\tilde{i}\rangle\}$ is an arbitrary orthonormal basis. On the fourth line we have used Parseval's identity.

Whether states $|\psi_j\rangle$ are orthogonal,[9] $\forall j = 1, ..., m,\ \ p_j = \lambda_j,\ \ |\psi_j\rangle = |j\rangle$ and the Von Neumann entropy is the very same as the Shannon entropy.

[9] For $j > m$ we put $\lambda_j = 0$ and $|j\rangle$ can be chosen arbitrarily.

This is the third case, where the Quantum information theory passes into the classical one, when the states are orthogonal. Just for remind, the first was when orthogonal qubits represent one bit and the second was the only orthogonal states can be copied[10]. The entropy is nonnegative, because all λ_i are nonnegative. If they were not, density operator ρ would not be positive. We will introduce further properties of the entropy in the next section.

2.4 Other entropies, mutual information and basic properties of von Neumann entropy

Other entropies are defined in the same way as the classical theory. However, some properties are different and thus very interesting.

Definition 2.10. *Suppose ρ and σ are density operators. The relative entropy (also known as Kullback-Leibler divergence) is defined by*

$$S(\rho||\sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma) = -\mathrm{Tr}(\rho \log \sigma) - S(\rho).$$

Quantum relative entropy satisfies very similar inequality as the classical one (theorem 1.3).

Theorem 2.9. *(Klein's inequality)*
$S(\rho||\sigma) \geq 0$ *with equality if and only if* $\rho = \sigma$*.*

Proof. The proof can be seen in Appendix A. □

Other entropies which are derived from von Neumann's entropy are defined similarly as in the classical Shannon's information theory.

Definition 2.11. *Joint entropy is*

$$S(X,Y) = -\mathrm{Tr}(\rho_{XY} \log \rho_{XY}),$$

conditional entropy

$$S(X|Y) = S(X,Y) - S(Y),$$

mutual information

$$S(X;Y) = S(X) + S(Y) - S(X,Y) = S(X) - S(X|Y) = S(Y) - S(Y|X),$$

where $\rho_{XY} = \sum_{i,j} p_{ij} |j\rangle|i\rangle\langle i|\langle j|$ is a spectral decomposition of the density operator for the joint system XY, $S(X) = -\mathrm{Tr}(\rho_X \log \rho_X)$, reduced density operator ρ_X is defined as $\rho_X = \sum_{\tilde{j}} \langle \tilde{j}|\rho_{XY}|\tilde{j}\rangle = \sum_i (\sum_j p_{ij})|i\rangle\langle i|$.

[10] Because we can distinguish between them. So we can determine which qubit we have received and then create some more.

Theorem 2.10. *Basic properties of Von Neumann entropy*

1. $S(\rho) \geq 0$. *The entropy is zero if and only if the state is pure.*
2. *In a n-dimensional Hilbert space* $\mathbf{H}$
$$S(\rho) \leq \log n.$$
The equality is achieved if and only if ρ *is maximally mixed state* $\frac{I}{n}$.
3. *Suppose a composite system* XY *is in the pure state. Then* $S(X) = S(Y)$.
4. *Suppose* p_i *are probabilities and the states* ρ_i *are such that for all* $i \neq \tilde{i}$, $\mathrm{Ran}(\rho_i) \cap \mathrm{Ran}(\rho_{\tilde{i}}) = \{|0\rangle\}$.[11] *Then*
$$S\left(\sum_i p_i \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i)\,.$$
5. *Joint entropy theorem*
Suppose $\rho_X = \sum_i p_i |i\rangle\langle i|$ *is a spectral decomposition of the density operator for the system* X, $\{\rho_i\}$ *any set of density operators for another system* Y. *Then*
$$S(\sum_i p_i |i\rangle\langle i| \otimes \rho_i) = S(\rho_X) + \sum_i p_i S(\rho_i)\,,$$
and thus
$$S(\rho \otimes \sigma) = S(\rho) + S(\sigma)\,.$$
6. $S(X,Y) \leq S(X) + S(Y)$ *(subaditivity)*

Proof.

1. $\sum_i p_i = 1$.
$$S(A) \;=\; -\sum_i p_i \log p_i \;=\; 0 \Leftrightarrow p_i \log p_i \;=\; 0 \;\text{ for }\; \forall i$$
The latter can be fulfilled only for $p_k = 1$ or $p_k = 0$. From $\sum_i p_i = 1$ follows that only one $p_i = 1$ and all other are zero. So $S(A) = 0 \Rightarrow$ pure state. The reverse implication is trivial.

[11] In other words, ρ_i have support on orthogonal subspaces.

2. From Klein's inequality $0 \leq S(\rho||I/n)$ with equality if and only if $\rho = I/n$.

Proof of the others can be found in [13]. □

The subaditity says that the mutual information is always non-negative. Notice that the fifth property is very similar to the second in the definition of Shannon entropy 1.3. If the $\rho_i = \rho(Y|X = x_i)$ would be the probability distribution of a system Y provided that event x_i happened[12], the fifth equality turns into

$$S\left(\sum_i p(X = x_i)|i\rangle\langle i| \otimes \rho_{(Y|X=x_i)}\right) = S(\rho_X) + \sum_i p_i S(\rho_{(Y|X=x_i)}).$$

We put $\rho_{(Y|X=x_i)} = \sum_j p(j|X = x_i)|j\rangle\langle j|$, use Bayes' theorem and get

$$S(X,Y) = S\left(\sum_{i,j} p(X = x_i)p(j|X = x_i)|j\rangle|i\rangle\langle i|\langle j|\right) = S(X) + \sum_i p_i S(Y|X = x_i)$$

Which is exactly the same as the conditional entropy in definition 1.3.

Finally we mention how measurements affect the entropy. The following theorems hold (for details see Ref. [13]).

Theorem 2.11. *Projective measurements increase entropy, i.e. ignorance about the measured system.*

Suppose P_i is a complete (i.e., $\sum_i P_i = I$) set of orthogonal projectors and ρ is a density operator. Then

$$S(\rho') = S(\sum_i P_i \rho P_i) \geq S(\rho)$$

with equality if a only if $\rho = \rho'$.

Proof. From Klein's inequality $0 \leq S(\rho||\rho')$. □

Theorem 2.12. *Generalized measurements can decrease entropy*

Suppose $M_1 = |0\rangle\langle 0|$, $M_2 = |0\rangle\langle 1|$. Then

$$S(\rho') = S(M_1 \rho M_1^\dagger + M_2 \rho M_2^\dagger) < S(\rho).$$

[12] In other words, we actually fished out the state x_i from the mixed state ρ_X.

By projective measurement we mean that the density operator ρ is only projected onto operator ρ' in contrast with the generalized measurement, where ρ has the non-zero chance to be projected onto state which is afterward changed into something else.

Not all properties of the Von Neumann entropy are identical to these of Shannon entropy. As we will see in the next chapter, for instance conditional entropy in quantum theory can be negative.

Chapter 3

Entanglement

The entanglement is outstanding element of the quantum theory which demonstrates the most counter-intuitive and fascinating features, especially non-locality. In nutshell it states that if two particles interact once, they are forever bound. These properties are now studied experimentally and till this time the experiments proves that the theory is right.

Entangled state is a joint state, where particles affect each other, independently on the space or time distance. It is necessary to add that this interaction is not causal and thus does not violates any laws of the theory of relativity. However, the classical point of view malfunctions here. We must threw off old prejudices and think in a new way, the way of quantum mechanics.

Consider a joint state of two particles

$$\frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle). \tag{3.1}$$

This so-called Bell state is one of the maximally entangled states. Suppose that we have an access to the first of the two particles in this state and Alice to the second. So that we could imagine what does the measurement do to the state (3.1) we will remind von Neumann's measurement axiom of the quantum mechanics.

After measuring out value a of the observable A on a pure state $|\psi\rangle$ system $|\psi\rangle$ passes onto state $P_{A=a}|\psi\rangle$, where $P_{A=a}$ is a projector onto subspace spanned by eigenvectors of the eigenvalue a. The process is called collapse of the wave function.

If we have measured $|0\rangle$ the (3.1) would pass onto

$$|0\rangle \underbrace{(|0\rangle\langle 0| + |1\rangle\langle 1|)}_{identity}\langle 0|\frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) = \frac{1}{\sqrt{2}}|0\rangle|1\rangle \simeq |0\rangle|1\rangle, \tag{3.2}$$

and Alice must measure $|1\rangle$ on the second particle. This is the correlation between two particles. We, knowing our measurement, automatically know what Alice have measured. If we could influence in some way what physical result we get, then we could communicate with Alice instantly. In fact we could not. From repeated measurements on the same state (3.1) we notice that we get absolutely stochastic results. This can be seen when we compute the reduced density matrix related to our measurement.

$$\begin{aligned}\rho_X &= \mathrm{Tr}_{\text{Alice's state}}(P_X\frac{1}{\sqrt{2}}(|1\rangle|0\rangle + |0\rangle|1\rangle)\frac{1}{\sqrt{2}}(\langle 0|\langle 1| + \langle 1|\langle 0|)P_x) \\ &= \frac{1}{2}\langle 0|(|1\rangle|0\rangle + |0\rangle|1\rangle)(\langle 0|\langle 1| + \langle 1|\langle 0|)|0\rangle \\ &+ \frac{1}{2}\langle 1|(|1\rangle|0\rangle + |0\rangle|1\rangle)(\langle 0|\langle 1| + \langle 1|\langle 0|)|1\rangle \\ &= \frac{1}{2}|1\rangle\langle 1| + \frac{1}{2}|0\rangle\langle 0|.\end{aligned} \tag{3.3}$$

Although, if we compare the list of our results with Alice's, we see that she has the very same, only with zeros changed to ones and vice versa. You could ask, "Isn't it some sort of conservation law only? The law that tells in each measurement your outcome + Alice's outcome gives 1? It other words, sum of the results is conserved?" No, in fact, it is not the conservation law. The true quantum behavior comes out when Alice tries to perform measurement in a little turned basis, say

$$|a\rangle = \frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle, \quad |b\rangle = -\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle. \tag{3.4}$$

Say that we have measured $|0\rangle$. Alice's state after our measurement is

$$|1\rangle = \frac{1}{2}|a\rangle + \frac{\sqrt{3}}{2}|b\rangle \tag{3.5}$$

and thus she has chance $\frac{1}{4}$ to measure value a corresponding to a state $|a\rangle$ and chance $\frac{3}{4}$ to measure value b corresponding to the state $|b\rangle$. If we make more experiments and then compare the results, we observe that for our $|0\rangle$ Alice has more $|b\rangle$ than $|a\rangle$. Although previously the conservation law fits, now, because a and b are different and both measurements a and b occur, no value $0 + (a \text{ or } b)$ is conserved.

The more intriguing fact is that Alice can choose anytime in what basis she wish to measure, for instance, she could change the measuring apparatus just at the moment you have measured. Then any light signal from your measurement could not reach her, and yet, Alice's particle 'knows' what you have measured and behaves in that way. We could say that there is some kind of superluminal signal that tells Alice's particle what you have measured (and still the signal cannot yield any information because the theory of relativity forbids that) or there is no signal at all. I will elucidate why considering superluminal exchange of information is not the proper approach.

Till now, our understanding was based on the causality premiss, i.e.,

1. first: 2 entangled particles scattered. You have access to the first and Alice to the second.

2. second: You measure a particle with a result $|0\rangle$. Then wave function collapses. In other words, a superluminal signal flies from you to Alice and tells the second particle how to behave.

3. third: Alice measure her particle and get a result $|a\rangle$ with $\frac{1}{4}$ probability and $|b\rangle$ with $\frac{3}{4}$ probability.

But our derivation of the possible result does not depend on time at all (we have stationary states). Indeed, as Ref. [4] shows, probability that Alice's outcome will be $|a\rangle$ along with your outcome $|0\rangle$ does not depend on the place or time at which the measurements occur.

> **Consider two entangled particles. We make a measurement on each of them. The result of the experiment then does not depend on a place or time at which the measurements occur.**

It is rather absurd to ask whether the superluminal signal comes from you to Alice or vice versa, because result of the experiment doesn't depend on who measured first. The causal presumption must be comprehend as a mnemonic or as a tool for computing probabilities of the possible results only.

Again, if you measure in basis $\{|0\rangle, |1\rangle\}$ and Alice in $\{|a\rangle, |b\rangle\}$ the only thing we could say is that there will be much more 0-b than 0-a results. This correlation is the true nature of entanglement.

One of the important property of entanglement is that the entangled state cannot be disintegrate on a tensor product of independent parts. For Bell state (3.1) it suggests

$$\begin{aligned}\frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)(|1\rangle - i|0\rangle) = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) + \frac{1}{\sqrt{2}}(-i|0\rangle|0\rangle + i|1\rangle|1\rangle) \\ \neq \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle).\end{aligned} \tag{3.6}$$

In other words, entangled particles are always dependent on each other and thus measurement on one particle changes probability distribution on possible outcomes of the others. The dependency is so strong that if you measure state of one particle from the entangled state, you can predict results of measurements in the same basis on the others, unless at the beginning you could not predict any measurement result with certainty. This is the way how is an entangled state defined. For two particles there are two possible joint states

$$\begin{aligned} &\alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle \\ &\alpha|0\rangle|1\rangle + \beta|1\rangle|0\rangle, \end{aligned} \tag{3.7}$$

where $|\alpha|^2 + |\beta|^2 = 1$, $\alpha \neq 0$, $\beta \neq 0$.

The dependency is closely associated with one notion of the quantum information theory — the conditional entropy.

Theorem 3.1. *Suppose $|XY\rangle$ is a pure state of a composite system. Then $|XY\rangle$ is entangled if and only if $S(Y|X) < 0$.*

Proof.

1. $\Rightarrow$: $|XY\rangle$ is entangled $\Leftrightarrow$ has a form of (3.7). $S(Y|X) = S(XY) - S(X)$. $S(XY) = 0$, because the state XY is pure. For (3.7) $S(X)$ is always positive. Accordingly $S(Y|X) < 0$.

2. $\Leftarrow$: If system $|XY\rangle$ was not entangled, then $S(Y|X) = S(Y)$, that is, measurement on X does not change probability distribution of possible results of the system Y. Again, the state XY is pure and thus $S(Y|X) = -S(X) = -S(Y)$, where the second equality comes from theorem 2.10. $S(Y|X) < 0$ and thus $S(Y) \neq 0$. $S(Y|X) = -S(Y) \neq S(Y)$ so $|XY\rangle$ must be entangled.

□

At the end of this chapter I would like to remark that not only different particles can be entangled. A particle can be entangled with itself too. The entanglement is not only the main actor of the EPR paradox[1] but also is experimentally studied in Quantum eraser experiment we will deal with in the last chapter.

[1] Einstein, Podolsky and Rosen suggested that uncertainty principle violates laws of the Relativity and causality and thus Quantum theory must be incomplete. Later the EPR paradox was reformulated for an entangled state of spin by David Böhm.

Chapter 4

Delayed choice experiment

The delayed choice experiment is one of the best examples of how the quantum behavior differs from our usual comprehension of reality. This experiment has been for a long time only in the realm of theory, but presently several realizations has been performed (e.g. [10, 7]). At first, we will remind double-slit experiment, because almost all delayed choice are based, in one way or another, on a double-slit or equivalently on a beam splitter. Then we introduce slightly modified original delayed choice experiment proposed by Wheeler and finally we will describe quantum eraser experiment proposed in Ref. [4].

4.1 Double-slit experiment

Double-slit experiment is a basic experiment which shows wave-particle duality of quantum particles. Experimental setup is on figure 4.1. On the first picture you can see particle source (e.g. light source or electron source), which produces quantum particles. These particles go through double slit and then interfere, that is, we can see an interference pattern on the screen (represented by a transparent red). We can be seen the interference fringes with the naked eye when a lot of particles are produced. Interestingly enough, the fringes do not disappear when particles are produced one-by-one. The only thing that changes is that we have to wait a longer time. If we record each impact of a particle on a photographic plate, the pattern finally comes out. This is what the Akira Tonomura's team did with electrons [15]. Their record of interference fringes is on the second picture of figure 4.1. We could ask, "How does the particles interfere, when there is only one at a time?" Well, that is because these particles interfere with itself! Richard Feynman in his thesis [2] suggested much stronger statement:

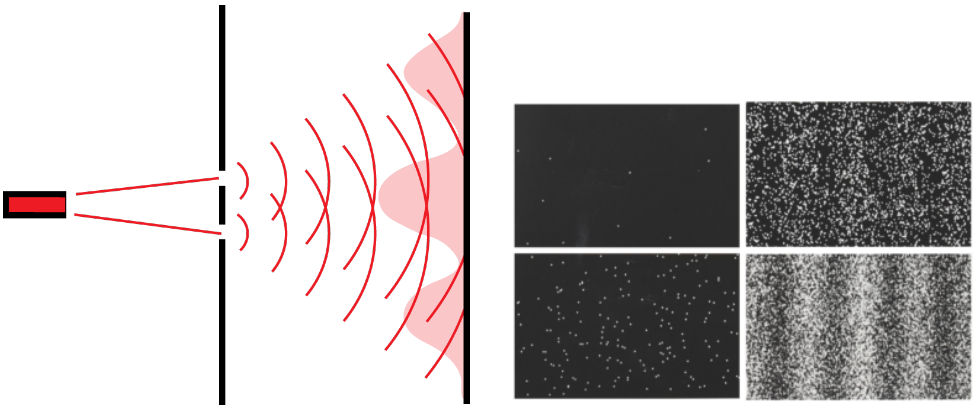

Figure 4.1: Experimental setup of the double slit experiment and results of Tonomura's realization using electrons as quantum particles. Quantum particles are red lighted.

Particles interfere with itself only.

That is the opposite of what we have been taught about interference of other types of waves. For instance, interference of water waves is a consequence of the collective behavior. It is the result of interactions between many particles and resulting wave is a joint movement of these particles.

In order to quantum particle interfere it has to go through both slits at once. This very confusing proposition indicates that the particle cannot be regarded as a point particle[1], but rather as a wave function $|\psi\rangle$ which expand anywhere where can. We are meeting the non-locality again.

Feynman's statement can explain many queries. For example, if we have not known that interference does not disappear when amount of particles dwindle, we could ask, "What happens, when particle density goes to zero?". Nothing happens, interference pattern is always the same. What changes is only the intensity of incoming particles. Good visible pattern is only the statistical consequence of many particles, each interfering with itself. This statement will help us to understand following, much more complicated experiment.

What happens, when we try to look which way (which slit) did the particle go? Suppose that we have measuring device (you can imagine some special camera) on the lower slit. As already said, the particle has to go through both ways at once, but we surely cannot measure only a half of the particle.

[1] How could a point particle go through two different slit at once?

We will discover that the particle either went through the lower way (your device noticed it) or some other way (the device did not notice it). For simplicity we can assume that particle always passes through double slit and the experimental setup is symmetrical, in other words, there is not a preferred slit. In that case, our chance of measuring a particle is exactly 50%. As a result, we will always know which way the particle went. Consequently, the particle always goes through one slit only and thus the interference pattern must be destroyed. The sole information about the way which the particle went destroys the interference pattern. Note that the information do not have to be necessarily possessed. For the destruction of the pattern is sufficient that the information is obtainable.

For example, consider two entangled partners. These partners have generally common origin, that is, they were created (or became entangled) at the same place and their momenta are also bounded by the momentum conservation law. From measuring position or momentum of one of the partners we are able (at least in principle) to determine position or momentum of the second and thus gain second's which-path information too.

As a second example we consider an experimental setup from figure 4.1 with a slight modification. We use the source of polarized light and put correctly oriented half-wave plate behind the upper slit so that the polarization of the photon which goes through the upper slit become perpendicular to the former polarization. Now, the which-path information is in principle obtainable — we can decide to measure polarization and thus determine which way the particle went but we do not have to even do that. Still, experiments show that the interference fringes disappear. So only the possibility of measuring polarization and thus determination whether the particle of light went through an upper or lower slit ensures that there cannot be any interference. We do not have to even measure.

> **Accessible, even in principle, which-path information destroys the interference pattern.**

We can reformulate the previous statement as following.

> **If we are not, even in principle, able to determine which way the particle went, particle interfere.**

We would like to stress that it is not the experimenter's knowledge, but the experimental setup, which destroys the interference pattern.

It is the experimental setup, which destroys the interference pattern.

In some cases, when the which-path information is recorded temporarily and is irrevocably deleted before a particle impacts the screen, the interference pattern is recovered. The deletion must be perfect, i.e., we cannot regain the which-path information even in principle. The experiments which study this recovering of interference pattern are called Quantum Eraser Experiments and we will deal with them in section 4.3.

4.2 Wheeler's Delayed Choice

In the previous section we have talked about a device, which could observe the path along which particle goes. The key point of the delayed choice experiment is that the device can gain which-path information after the particle passes through double slit. We will describe slightly modified version of the famous Wheeler's delayed choice experiment [18]. The experimental setup is on figure 4.2.

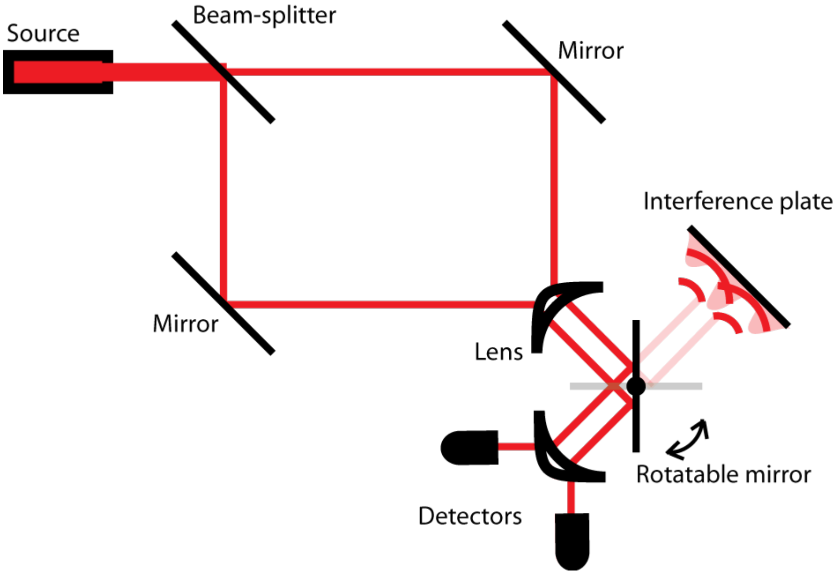

Figure 4.2: Experimental setup of Wheeler's delayed choice experiment.

Path of the particle is red lighted. The particle goes through beam splitter. Consider that the beam-splitter is ideal, i.e., the light beam has 50% chance to be reflected and 50% to pass through. For a single-particle this means that the particle is reflected and passed through at the same time. We should not be surprised after the experience with the double slit. In double slit experiment particle also went through both slits at once, or through one only, that depended upon what kind of measurement we have chosen. If we have tried to measure which slit did the particle goes through, the particle always went through one slit only. If we have not, particle interfere and thus had to go through both slits at once. After passing through beam-splitter, particle is reflected from the mirror and focused by a lens. Now we can choose which measurement do we perform. If the rotatable mirror is vertically positioned, the particle is reflected into detector which observes which way the particle went. If the rotatable mirror is horizontally positioned, the particle is reflected towards interference plate and interfere.

Remember that choosing the measurement predetermines whether the particle goes one way only, or two ways at once. The interesting thing is that we can turn the mirror anytime, for example just after the particle passed beam-splitter and before it hit the rotatable mirror, and thus choose the measurement. In other words, you can choose whether the particle goes one way only or both ways at once *after* the particle went through beam-splitter!

In a classical point of view, you influence what happened after it happened, you influence the past. It leads to an idea that the classical point of view is not right in this case. Of course, it is not right, because determining which way the particle went is only our thought construct which helps us to build a mental picture of what happens. Until the measurement we cannot say, not even in principle, which way the particle goes.

4.3 Delayed Choice Quantum Eraser Experiment

The delayed choice quantum eraser experiment which we will now present was adopted from Ref. [4]. As we have said in Section 4.1, the eraser experiment is a type of an experiment where the which-path information is irreversibly lost, erased. This experiment also combine the delayed choice. Two outcomes are possible. Either an interference pattern is gained *before* the which-path information is irreversibly erased or interference pattern is destroyed *before* the which-path information is revealed. The experimental setup is on figure 4.3.

A photon goes through double slit and hit the BBO[2] crystal. The BBO crystal transforms the photon into an entangled pair via Spontaneous para-

[2] beta barium borate

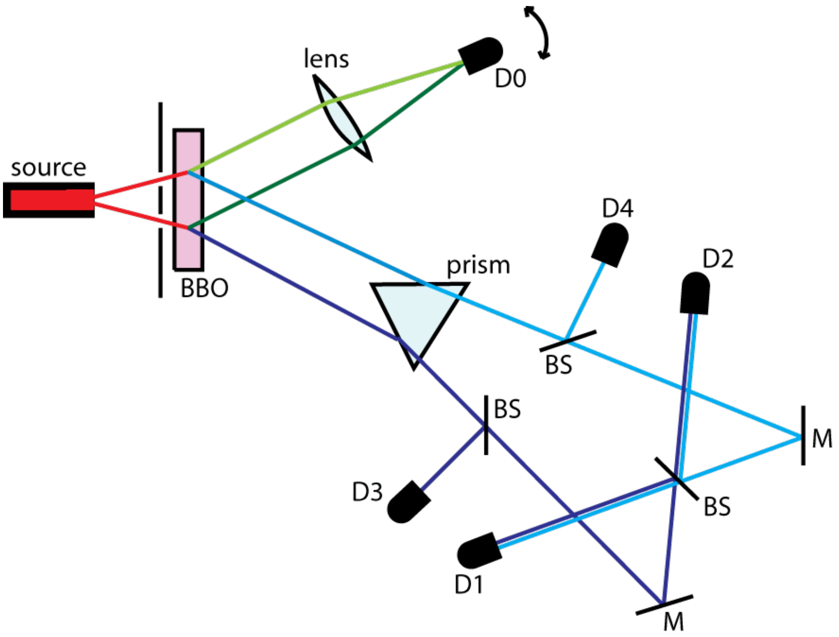

Figure 4.3: Experimental setup of delayed choice quantum eraser experiment.

metric down-conversion. One from the pair (signal photon; green path) goes up and the second (idler photon; blue path) goes down. Notice that the creation of pair took place on the two different points simultaneously and thus each of the pair come along two paths at once too. If the subsequent measurement reveals that the idler went through lower path (dark blue), then, because of common origin, signal have to go through lower path too (dark green). Similarly with the upper path. Idler then meets some beam-splitters (BS) and mirrors (M) and is detected by one of the four detectors (D1, D2, D3, D4). Detector D0 is movable and we will use it to explore possible interference of signal photon.

If the idler hits detector D3 or D4, which-path information is revealed and signal does not interfere. Since the last beam-splitter irreversibly erases the which-path information, if the idler is received by detector D1 or D2, signal photon interfere. Note that if both paths of the signal photon from BBO crystal to detector D0 is the same, paths of the idler from BBO to the last beam-splitter must be also the same. Otherwise, knowing when the signal photon has arrived, we could determine the path just from time difference between arrivals.

The interesting thing is that we can make way of the signal photon much shorter than the way of idler so the signal photon is always detected *before* the idler. As we have said in the previous chapter 3 (see also Ref. [4]), the experiment result does not depend on a place or time at which the measurements occur. The measurement made on one of the pair only changes the probability distribution of possible outcomes of the other, but probability distribution of the whole, i.e., outcome of the measurement on signal will be a along with outcome of the measurement on idler will be b,[3] is always the same and does not depend on which measurement occurs first. Loosely speaking, if the idler hits detector D3 or D4, we have bigger chance to find the signal at potential peak of the interference pattern, if the idler hits detector D1 or D2, the signal will not fill the potential interference pattern and vice versa. **If the signal hits peak of the potential interference pattern, we have the bigger chance that the idler will be received by detector D1 or D2 than D3 or D4.**[4]

There is one important question that could be asked. Consider the previous experiment arrangement with one additional detector D5 in the back, without D3 and D4 and a person ready to quickly put the detector D5 in the idler way (figure 4.4).

Suppose that detector D0 receive the signal before the idler reaches the prism. If the person put the detector in the way of the idler right after the signal is detected does it affect the signal photon interference? Well, the experiment with a person standing there is an absolutely different experiment. Now, the person and additional detector are, together with signal and idler photons, parts of a big collective wave function and thus person there standing, with an ability to determine the path, could affect the interference pattern. Furthermore, the big collective wave function depends on time now so the measurements depend on time too. How much the person affects the measurement on signal photon? We do not know exactly. Analyzing such a complicated system is far beyond our present capability. Nevertheless, we can say something about it.

Using the heuristic argument: "The experiment will run so that the result will be what we expect" will help us to understand what happens. All we know is that the result should not lead to paradoxes, i.e., if the signal photon interferes,[5] the idler should interfere too no matter what the person does. Loosely said, if the person put the detector D5 in the assumed way of the idler, no photon will be detected, because the idler hits somewhere else, filling

[3] Suppose that we measure observable $\hat{A}$ with a possible result a and $\hat{B}$ with a possible result b.

[4] Consider that the interference pattern is not visible as a whole, but is step by step filled up by isolated hits of signal particles.

[5] Assume the signal hits the peak of potential interference pattern and this place is very improbable for photons heading from only one slit.

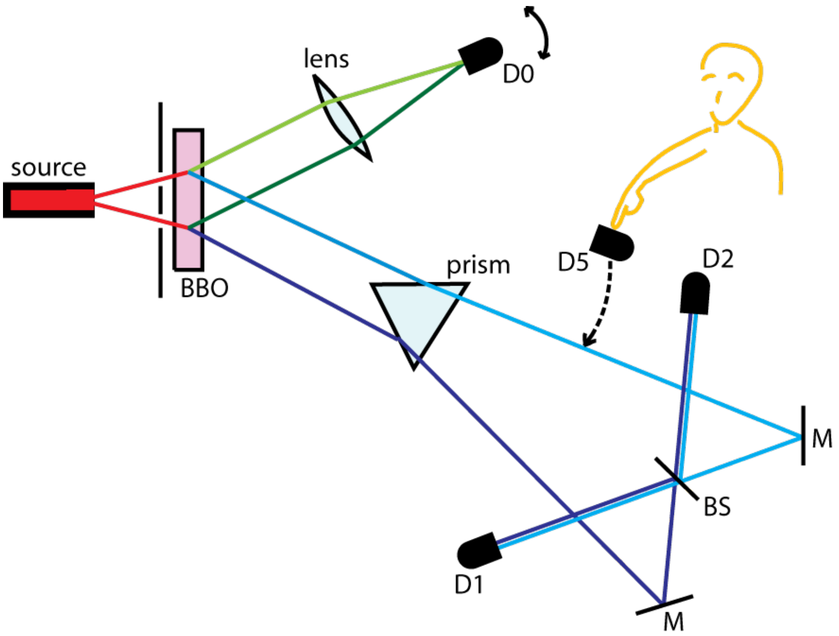

Figure 4.4: Experimental setup of delayed choice quantum eraser experiment with a very quick person and additional detector.

up his own interference pattern. If the signal photon does not interfere,[6] the idler should not interfere neither. Roughly speaking, the person putting the detector in the way receives the photon.

We would like to stress that one logical trap we could fall into is not considering photons whose entangled friend is not detected. As we have shown, we could put the detector D5 in the way in order to obtain the which-path information of the idler, but this information can not be revealed while the signal has interfered. For elucidating this, consider experimental setup on the figure 4.5.

In this case, the which path information of the idler is always revealed and thus the signal should not never interfere. But is it really so? If the signal interferes we do probably not detect the idler and, retroactively, the signal can interfere. In this kind of delayed choice experiments we do not know exactly whether it is the experimenter or the nature who chooses whether the which path information will be obtained. In this kind of experiments the nature could choose by some unknown algorithm whether it allows us to reveal the information. This problem should be studied experimentally and

[6] Signal hits the place which is very probably for photons heading from the upper (or lower) slit and lies in the minimum of the potential interference pattern.

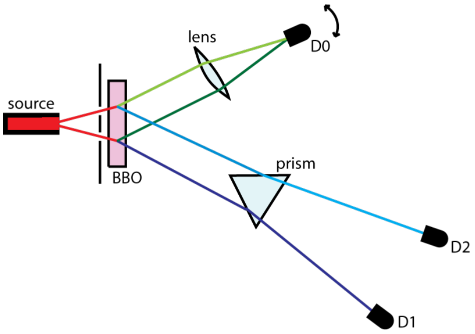

Figure 4.5: Experiment where the which-path information is always obtained.

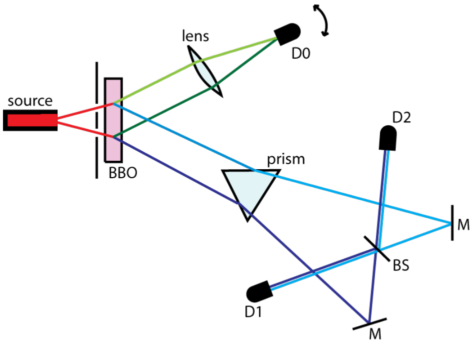

Figure 4.6: Experiment where the which-path information is always erased.

is crucial for understanding the logical structure of quantum mechanics.

4.4 Constructing the free will quantum test

At the very beginning we would like to emphasize that the following considerations depend on many experimental results and if some of them emerge not to be true, constructing such a free will quantum test is not possible.

The first experiment result we need to is that mentioned in the previous section. We need to know whether the results of the experiments shown on figures 4.5 and 4.6 are always the same, i.e., in the former signal photon never interfere and in the latter signal photon always interfere.[7] If not, we can not continue.[8] If so, we make another experiment shown on figure 4.7.

We use the timer which pushes automatically the detectors in the way of the idler in the right time — after the signal photon is detected and before the idler reaches the prism. The signal photon 'should know' what is preparing and behave accordingly, i.e., since the signal photon is, together with the idler and detectors and whole pushing mechanism, part of one big wave function, it has to 'know' that the detectors D3 and D4 is going to reveal the which-path information and therefore does not interfere. This must be studied experimentally too. If our guess is right, we can step to the free will quantum test.

The experimental setup of a test, which verify whether we have the free will or not, is on figure 4.8. The only changed thing is that the mechanism pushing the detectors is not automatical, but is controlled by a person whom

[7] Remember that the signal photon is always detected before the idler.

[8] Unless we change the experimental setup somehow.

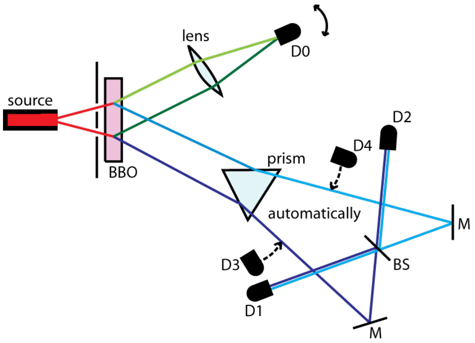

Figure 4.7: Experiment with a mechanism which pushes detectors D3 and D4 into the assumed way of the idler photon automatically.

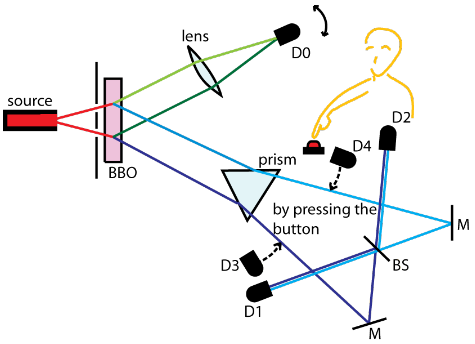

Figure 4.8: Experiment where the person decides whether push the detectors or not.

we test. Now it is the person who decides whether reveal the which-path information or not. This decision can be made *after* the signal photon is detected. But the signal had to 'know' what is the person going to do in order to 'know', how to behave — whether interfere or not.

If the signal photon always interfere when the person decides not to push the button and does not interfere when the person decides to push the button, the person does not have the free will, because the signal photon has always known, what is the person going to do. Person's decision was only a consequence of former events, influencing and inborn personality, everything what happened in past. Thus the decision is an illusion, because it was determined what is the person going to do *before* the decision is made.

If the person has the free will, something unexpected should happen or some of the presumptions used to the construction of the free will test are not correct.

Now we see how important the first presumption is — if the result of the experiment is in some way stochastic, decision of the person does not affect the experimental result at all and we cannot say anything about it's free will.

4.5 How is the delayed choice related to information theory?

In their papers [16, 17] Časlav Brukner and Anton Zeilinger state that a whole quantum theory is in fact the theory of information. The information is there introduced as a most fundamental notion and they states that the whole quantum theory can be derived from a fact which they called the finiteness of information. Their construct is also based on propositions, statements which can the experimenter say about nature. They also define a new measure of information which is in fact the renormalized Tsallis entropy 1.8 with a parameter $q = 2$. Measurements in their interpretation are represented as a instantaneous changes in a so-called information vector.

If their look on the quantum theory is correct, then also the delayed choice can be explained using this approach. Nevertheless, their approach is still in progress and at present it is conceptually difficult to explain the delayed choice experiment with Brukner *et all.* approach. The research in this direction is presently under intense investigation.

Conclusion

In the first chapter we introduced the concept of information measurement, compared several types of information and we also axiomatically introduced Shannon entropy by to different approaches — namely Shannon's and Khinchin's axiomatics in order to better understand the meaning of the classical information measure. We also discussed the fact that Shannon's entropy is not the only possible logically viable information measure and briefly referred also to other existent ones. Then we defined entropies derived from Shannon's entropy and specified some of their properties. In the second chapter we reminded necessary mathematical formalism, defined a quantum bit and compared it with the classical one and introduced a quantum measure of average information — von Neumann entropy. We also provide a mathematically exact proof of Klein's inequality (i.e., non-negativity of the quantum relative entropy), especially the equality conditions, because in the literature it is often mistreated. In the third chapter we wanted to highlight how much the notion of entanglement is differs from what we know from everyday life. In the last chapter we reminded the double slit experiment in order to better understand the Delayed Choice Quantum Eraser experiment. We also suggested the free will quantum test.

Appendix A

Proof of Klein's inequality

Theorem A.1. *(Klein's inequality)*
$S(\rho||\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.

Proof. In [13] the proof is not done correctly. Assume that $P_{ij} \in [0,1]$, $\sum_j P_{ij} = 1$. The authors [13] have written:

...Because $\log(\cdot)$ is a strictly concave function it follows that $\sum_j P_{ij} \log q_j \leq \log r_i$, where $r_i \equiv \sum_j P_{ij} q_j$, with equality if and only if there exist a value of j for which $P_{ij} = 1$.

The equality condition is not true. Consider (q_1, q_2, q_3), $q_1 = \frac{1}{4}$, $q_2 = \frac{1}{4}$, $q_3 = \frac{1}{2}$. If we choose $P_{i1} = \lambda \in (0,1)$, $P_{i1} = 1 - \lambda, P_{i3} = 0$, then we have

$$P_{i1} \log \frac{1}{4} + P_{i2} \log \frac{1}{4} + P_{i3} \log \frac{1}{2} = \log \left(P_{i1} \frac{1}{4} + P_{i2} \frac{1}{4} + P_{i3} \frac{1}{2} \right) \tag{A.1}$$

$$\lambda \log \frac{1}{4} + (1-\lambda) \log \frac{1}{4} + 0 \log \frac{1}{2} = \log \left(\lambda \frac{1}{4} + (1-\lambda) \frac{1}{4} + 0 \frac{1}{2} \right). \tag{A.2}$$

The equality still holds even thought such j does not exist. It is the consequence of the $q_1 = q_2$. The statement should be corrected (for clearer notation we change P_{ij} to V_{ij}):

Suppose that all q_i are different, $V_{ij} \in [0,1]$, $\sum_j V_{ij} = 1$. Then for strictly concave function $\log$ follows that $\sum_j V_{ij} \log q_j \leq \log r_i$, where $r_i \equiv \sum_j V_{ij} q_j$, with equality if and only if there exist a value of j for which $V_{ij} = 1$.

Proof is the following. Let $\rho = \sum_i p_i P_i$, $\sigma = \sum_j q_j Q_j$, where p_i, q_i are nonnegative and different, be an orthonormal decomposition for ρ and σ, where $P_i = \sum_k |\psi_{i,k}\rangle\langle\psi_{i,k}|$ is the projector onto the linear subspace spanned by orthonormal eigenvectors $\{|\psi_{i,k}\rangle | 1 \leq k \leq dim(\mathrm{Ran}(P_i)) = R_i\}$ of the eigenvalue p_i.[1] $Q_j = \sum_l |\varphi_{j,l}\rangle\langle\varphi_{j,l}|$ similarly. Let $\{|k\rangle\}$ be an orthonormal basis. Since trace of the operator does not depend on a choice of the orthonormal basis so we can assume that $\{|k\rangle\}$ is made from eigenvectors of the density operator σ, that is $\{|k\rangle\} = \{|\varphi_{j,l}\rangle\}$. Consequently we easily deduce that $V_{ij} \in [0,1]$ in (A.5) and $r_i \in [0,1]$ in (A.7). Suppose that the kernel of σ has trivial intersection with the support of ρ. In the case when it has not we define $S(\rho||\sigma) = +\infty$.[2]

1. inequality

$$
\begin{aligned}
S(\rho||\sigma) &= \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma) = \\
&= \sum_k \langle k| \left(\sum_i p_i \log p_i P_i \right) |k\rangle - \sum_k \langle k| \left(\sum_i p_i P_i \sum_j \log q_j Q_j \right) |k\rangle = \\
&= \sum_i p_i \log p_i \underbrace{\left(\sum_k \langle k|P_i|k\rangle \right)}_{R_i} - \sum_{i,j} p_i \log q_j \sum_k \langle k|P_i Q_j|k\rangle = \\
&= \sum_i (R_i p_i \log(R_i p_i) - R_i p_i \Big(\sum_j \underbrace{\left(\sum_k \frac{\langle k|P_i Q_j|k\rangle}{R_i} \right)}_{V_{ij}} \log q_j + \log(R_i) \Big)) \overset{(1)}{\geq} \\
&\overset{(1)}{\geq} \sum_i (R_i p_i \log(R_i p_i) - R_i p_i (\log \left(\sum_j V_{ij} q_j \right) + \log R_i)) = \\
&= \sum_i (R_i p_i \log(R_i p_i) - R_i p_i \log \underbrace{\left(\sum_{j,k} \langle k|P_i Q_j|k\rangle q_j \right)}_{r_i}) \overset{(2)}{\geq} 0,
\end{aligned}
\tag{A.3}
$$

where

$$R_i = \mathrm{Tr}(P_i) = dim(\mathrm{Ran}(P_i)), \tag{A.4}$$

$$\sum_j V_{ij} = \sum_k \frac{\langle k|P_i \sum_j Q_j|k\rangle}{R_i} = \frac{R_i}{R_i} = 1 \tag{A.5}$$

and since projectors are positive $V_{ij} \in [0,1]$. The inequality (1) comes

[1] By $\mathrm{Ran}(P_i)$ we mean the range of values of the projector P_i.

[2] A problem with infinity can be seen on the fourth line of the following derivation.

from the corrected statement above

$$\forall i,\quad \sum_j V_{ij}\log q_j \le \log\left(\sum_j V_{ij}q_j\right) \tag{A.6}$$

with equality if and only if $\exists j_i$ such that $V_{ij_i} = 1$.

$$r_i \in [0,1],\quad \sum_i r_i = \sum_{i,j,k}\langle k|P_iQ_j|k\rangle q_j = \mathrm{Tr}(\sigma) = 1 \tag{A.7}$$

$$R_ip_i \in [0,1],\quad \sum_i R_ip_i = \mathrm{Tr}(\rho) = 1 \tag{A.8}$$

and thus the second inequality (2) comes from theorem 1.3 with equality if and only if

$$\forall i,\quad R_ip_i = \sum_{j,k}\langle k|P_iQ_j|k\rangle q_j \Leftrightarrow p_i = \sum_j V_{ij}q_j. \tag{A.9}$$

2. equality conditions

   (a) $\Leftarrow$: Trivial.

   (b) $\Rightarrow$: From the equality conditions (A.6),(A.9) we have

   $$\forall i,\quad \exists j_i,\quad p_i = q_{j_i}. \tag{A.10}$$

   Function $f : i \longrightarrow j_i$ is injective. For reductio ad absurdum suppose that both $p_i^{(1)}$ and $p_i^{(2)}$ has the same output q_{j_i}. Then $p_i^{(1)} = q_{j_i} = p_i^{(2)} \neq p_i^{(1)}$ (A.10). Now we will prove that for all i $P_i = Q_{j_i}$. From the equality conditions we also have

   $$V_{ij_i} = \sum_k \frac{\langle k|P_iQ_{j_i}|k\rangle}{R_i} = 1, \tag{A.11}$$

   $$V_{ij\neq j_i} = \sum_k \frac{\langle k|P_iQ_j|k\rangle}{R_i} = 0. \tag{A.12}$$

   Since $R_i = \mathrm{Tr}(P_i)$, we can rewrite (A.11) as $\mathrm{Tr}(P_i) = \mathrm{Tr}(P_iQ_{j_i})$. Now, using general properties of projectors ($P_i = P_i^\dagger = P_i^2$), ($Q_j = Q_j^\dagger = Q_j^2$) and Schwarz inequality from theorem 2.7, we get

   $$\mathrm{Tr}(P_i)^2 = |\mathrm{Tr}(P_iQ_{j_i})|^2 \le \mathrm{Tr}(P_i)\mathrm{Tr}(Q_{j_i}). \tag{A.13}$$

We can divide, because $\mathrm{Tr}(P_i) \neq 0$, and we get $\mathrm{Tr}(P_i) \leq \mathrm{Tr}(Q_{j_i})$. Now we sum all these inequalities:

$$dim(\mathbf{H}) = \sum_i \mathrm{Tr}(P_i) \leq \sum_i \mathrm{Tr}(Q_{j_i}) \leq dim(\mathbf{H}) \tag{A.14}$$

and thus $f : i \longrightarrow j_i$ is surjective and

$$\forall i,\ \ dim(\mathrm{Ran}(P_i)) = \mathrm{Tr}(P_i) = \mathrm{Tr}(Q_{j_i}) = dim(\mathrm{Ran}(Q_{j_i})). \tag{A.15}$$

Now we use (A.12) and derive that $P_i = Q_{j_i}$. We also use special basis made from eigenvectors of the density operator σ defined at the beginning $\{|\varphi_{j,l}\rangle\}$.

For $j \neq j_i$ we have

$$\begin{aligned} 0 = \mathrm{Tr}(P_i Q_j) &= \sum_m \sum_{l=1}^{dim(\mathrm{Ran}(Q_m))} \langle\varphi_{m,l}|P_i \underbrace{Q_j|\varphi_{m,l}\rangle}_{\delta_{jm}|\varphi_{m,l}\rangle} = \\ &= \sum_{l=1}^{dim(\mathrm{Ran}(Q_j))} \underbrace{\langle\varphi_{j,l}|P_i|\varphi_{j,l}\rangle}_{\geq 0} \end{aligned} \tag{A.16}$$

and thus

$$\forall l = 1, 2, ..., dim(\mathrm{Ran}(Q_j)),\ \ 0 = \langle\varphi_{j,l}|P_i|\varphi_{j,l}\rangle = \|P_i|\varphi_{j,l}\rangle\|^2. \tag{A.17}$$

That is

$$\forall j \neq j_i,\ \ P_i Q_j = 0. \tag{A.18}$$

In other words

$$\forall j \neq j_i,\ \ \mathrm{Ran}Q_j \subset \mathrm{Ker}P_i = (\mathrm{Ran}P_i)^{\perp}, \tag{A.19}$$

$$\mathrm{Ran}Q_j \subset \bigcap_{\forall i, j_i \neq j} (\mathrm{Ran}P_i)^{\perp} = \mathrm{Ran}P_{i,j=j_i}. \tag{A.20}$$

Since dimensions of the subspaces are equal (A.15)

$$\forall i,\ \ P_i = Q_{j_i}. \tag{A.21}$$

At last we have

$$(\forall i,\ \ p_i = q_{j_i},\ \ P_i = Q_{j_i}) \Leftrightarrow \rho = \sigma. \tag{A.22}$$

□

Bibliography

[1] J. Blank, P. Exner, and M. Havlíček. *Hilbert Space Operators in Quantum Physics*. Springer, 2008.

[2] L. M. Brown and R. P. Feynman. *Feynman's Thesis: A New Approach to Quantum Theory*. World Scientific Publishing Company, 2005.

[3] E. Desurvire. *Classical and Quantum Information Theory: An Introduction for the Telecom Scientist*. Cambridge University Press, 2009.

[4] B. Gaasbeek. Demystifying the Delayed Choice Experiments. 2010. [quant-ph/1007.3977v1].

[5] P. D. Grünwald and P. M. B. Vitányi. Kolmogorov Complexity and Information Theory with an interpretation in terms of questions and answers. *Journal of Logic, Language and Information*, 12:497–529, 2003.

[6] T. J. Herzog, P. G. Kwiat, H. Weinfurter, and A. Zeilinger. Complementarity and the Quantum Eraser. *Phys. Rev. Lett.*, 75:3034, 1995.

[7] V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, and J. Roch. Experimental Realization of Wheeler's Delayed-Choice Gedanken Experiment. In *Conference on Coherence and Quantum Optics*, page CWB4. Optical Society of America, 2007.

[8] P. Jizba and T. Arimitsu. The world according to Rényi: Thermodynamics of multifractal systems. *Ann. Phys.*, 312:17, 2004.

[9] A. I. Khinchin. *Mathematical Foundations of Information Theory*. Dover Publications, 1957.

[10] Y. Kim, R. Yu, S. P. Kulik, Y. Shih, and M. O. Scully. Delayed Choice Quantum Eraser. *Phys. Rev. Lett.*, 84:1–5, 2000.

[11] N. Lee, H. Benichi, Y. Takeno, S. Takeda, J. Webb, E. Huntington, and A. Furusawa. Teleportation of Nonclassical Wave Packets of Light. *Science*, 332:330–333, 2011.

[12] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt. 14-Qubit Entanglement: Creation and Coherence. *Phys. Rev. Lett.*, 106:130506, 2011.

[13] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences).* Cambridge University Press, 2004.

[14] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.

[15] A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki, and H. Ezawa. Demonstration of single-electron buildup of an interference pattern. *Am. J. Phys.*, 57:117–120, 1989.

[16] Č. Brukner and A. Zeilinger. Operationally Invariant Information in Quantum Measurements. *Phys. Rev. Lett.*, 83:3354, 1999.

[17] Č. Brukner and A. Zeilinger. Information and fundamental elements of the structure of quantum theory, 2003. [quant-ph/0212084v1].

[18] J. A. Wheeler. *Mathematica Foundations of Quantum Theory*, chapter The Past and Delayed Choice Double Slit Experiment. Academic Press, 1978.