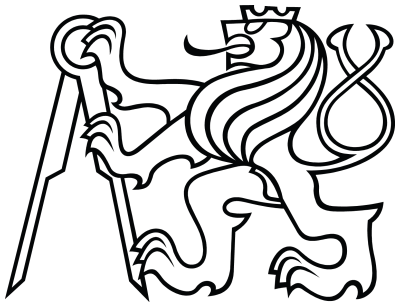
CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering

RESEARCH PROJECT

2012 Dominik Šafránek

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



RESEARCH PROJECT

Delayed Choice Experiments and Particle Entanglement

Author: Dominik Šafránek
Supervisor: Ing. Petr Jizba, PhD.
Consultants: Dr. Almut Beige and Dr. Jacob Dunningham
Year: 2012

Acknowledgement

I would like to thank to Dr. Almut Beige I have been working with on Hidden Quantum Markov Models for a great leadership and financial support, to Dr. Jacob Dunningham for extraordinary thoughts and opinions about the Free Will experiment and entanglement, "You inspired me a lot, Jacob!", and to Ing. Petr Jizba, Ph.D without whom nothing would be possible.

Prohlášení

Prohlašuji, že jsem svůj výzkumný úkol 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 Research Project 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: **"Delayed choice" experimenty a částicová provázanost**
Autor: Dominik Šafránek
Obor: Matematické inženýrství
Druh práce: Výzkumný úkol
Vedoucí práce: Ing. Petr Jizba, PhD.
Konzultanti: Dr. Almut Beige and Dr. Jacob Dunningham
Katedra fyziky, Fakulta jaderná a fyzikálně inženýrská

Abstrakt

Skryté Markovovy modely jsou používány déle než 60 let a prezentujeme zde jejich kvantové zobecnění se všemi základními statistickými veličinami nutnými pro jejich popis a dokazujeme, že mohou modelovat skutečné fyzikální experimenty. V druhé části navrhujeme Kvantový test svobodné vůle založený na Kvantové gumě se zpožděným výběrem, který by teoreticky mohl určit, zda testovaná osoba má či nemá svobodnou vůli. Ukazujeme však, že žádné experimenty podobného typu nemohou fungovat, a efektivně dokazujeme, že použitím pouze provázaných stavů nemůžeme poslat žádnou informaci.

Klíčová slova: Skryté Kvantové Markovovy modely,
kvantová provázanost, svobodná vůle
Kvantová guma se zpožděným výběrem

Title: **Delayed Choice Experiments and Particle Entanglement**
Author: Dominik Šafránek

Abstract

The Hidden Markov models are used for a more than a sixty years and here we present the quantum generalization of these, with all of basic statistical quantities used for describing them, and prove they can model some real physical experiments. In the second part we propose the Free Will experiment, the experiment which could possibly determine whether the tested person has or has not the free will based on the delayed choice quantum eraser. We also show that none of the similar Free Will test can be done and prove effectively that any information cannot be send through the entanglement alone.

Keywords: Hidden Quantum Markov models
entanglement, free will
Delayed Choice Quantum Eraser

Contents

Introduction

Most of this work has been created during my 3-month-long internship at the University of Leeds, where I have been working on Hidden Quantum Markov models with Dr. Almut Beige, Tom Barlow and Ben O'Neill in the Quantum Physics group. Meanwhile, I have been visiting Dr. Jacob Dunningham to consult my former idea which I have been working on with ing. Petr Jizba Ph.D. From both of these sessions many thoughts came out and I had written most of them here. Those two areas were a little different, although both of them are parts of the Quantum Information physics, so the structure of this work is also divided into two main sections. The Hidden Quantum Markov models and Entanglement.

In the first part I present mainly basic ideas leading to a consistent theory of Hidden Quantum Markov models, I introduce a main definition based on the physical background and prove that such Markov models could model some real physical systems and that the name is also well-describing, since the Hidden Quantum Markov models generalize the classical Hidden Markov Models. I briefly introduced both Markov models and Hidden Markov models in the appendix, so if you are getting lost, feel free to take a look. I did also enclosed some examples of Quantum Markov models there. Still, this work is mainly about mathematical consistence of the idea, since only 2 papers has been published [2], [7], and is continuing of the first. But some application may be found in future. For example Hidden Markov models are widely used for the analysis of the sequences of the DNA cite. But the DNA is so small that the quantum effects may matter and thus the Hidden Quantum Markov models could come to stage. Concrete applications are not known nowadays and have to be the objects of the further research.

In the second part I briefly introduce the Delayed Choice and Quantum Eraser experiments both with one example and consequences their analysis bring to the quantum mechanics. Then I talk about the Free Will experiment, which is the delayed choice quantum eraser experiment, and could possibly test the person's free will. We will show that such free will experiments are predetermined not to work and also a very effective proof why entanglement

alone cannot be used for sending any information.

At the very end we will be engaged with the Wigner's friend, which is the expansion of the Schrödinger's cat and derive that this approach to a measurement predicts there is not an absolute reality.

Chapter 1

Quantum Markov Models

Introduction

1.1 Definition of Quantum Markov Model, Theorems

Definition 1.1.1. *Let* $\mathbf{H}$ *be* n*-dimensional Hilbert space,* $\mathbf{K} = \{K_1, K_2, ..., K_m\}$ *set of linear operators on* $\mathbf{H}$*. We call couple* $(\mathbf{H}, \mathbf{K})$ n*-dimensional quantum Markov model with* m *possible outputs if and only if the following conditions are satisfied:*

1. $K_1^\dagger K_1 + K_2^\dagger K_2 + \cdots + K_m^\dagger K_m = I$
2. *Vectors from* $\mathbf{H}$ *form a growing chain in the following sense:*
   (a) *Each segment of the chain is represented by normalized*[1] *vector(≡ state) from* $\mathbf{H}$*.*
   (b) *In each segment a value from* $\{1, 2, ..., m\}$ *is produced.*
   (c) *If the segment≡state is* $|\psi\rangle$*, then probability of producing value* $i \in \{1, 2, ..., m\}$ *from the previous clause is* $P(i) = \langle\psi|K_i^\dagger K_i|\psi\rangle$*.*
   (d) *If value* i *is produced, next segment of the chain is* $\frac{K_i|\psi\rangle}{\sqrt{\langle\psi|K_i^\dagger K_i|\psi\rangle}}$*.*

Probability from item c) is correctly defined, since

$$0 \leq \langle\psi|K_i^\dagger K_i|\psi\rangle \leq 1 \ ,$$

[1] $|\psi\rangle \in \mathbf{H}$ such that $\langle\psi|\psi\rangle = 1$. Futhermore, we will always talk about normalized and thus non-zero vectors.

$$\sum_{i=1}^{m} P(i) = \sum_{i=1}^{m} \langle\psi|K_i^\dagger K_i|\psi\rangle = \langle\psi|I|\psi\rangle = 1 \ .$$

Moreover, the quantum Markov model form a Markov model in the classical sense, because the next state in the chain depends only on the previous. The first condition says that the set $\mathbf{K} = \{K_1, K_2, ..., K_m\}$ is a set of so-called trace-preserving Kraus operators [1].

The question arises. Can be such Markov model physically implemented? Yes, it can. In fact, any Quantum Markov model can be in principle implemented. The precise algorithm together with some examples will be presented in 1.3.2.

The special case of the n-dimensional quantum Markov model with m possible outputs is 1-qubit (2-dimensional) quantum Markov model with two possible outputs $\{0, 1\}$. Later, we will look closer at them.

1.1.1 Probability distribution of possible states after n iterations

We have defined how such model is evolving, but what if we want to predict the evolution? We know that the next state depends only on the previous, but it differs if another value is produced and we cannot predict which value will be produced. If we know the previous state and the output, we are able to calculate the actual state, but if we want to predict what the next state will be, we need to use density operator.

Suppose that we know the initial state $|\psi\rangle$. The initial state will be then described by density operator

$$\rho_0 = |\psi\rangle\langle\psi| \ . \tag{1.1}$$

If the outputted value is i, the next state will be characterized by density operator

$$\rho(i) = \frac{K_i|\psi\rangle\langle\psi|K_i^\dagger}{\langle\psi|K_i^\dagger K_i|\psi\rangle} \ . \tag{1.2}$$

But we do not know the outputted value, so the best thing to do is to characterize the next state by the weighted average of all possible density operators.

$$\rho_1 = \sum_i p(i)\rho(i) = \sum_i \langle\psi|K_i^\dagger K_i|\psi\rangle \frac{K_i|\psi\rangle\langle\psi|K_i^\dagger}{\langle\psi|K_i^\dagger K_i|\psi\rangle} = \sum_i K_i\rho_0 K_i^\dagger \ . \tag{1.3}$$

For writing the above in a compact form we define the linear operator (usually called superoperator) K as

$$\forall\rho, \quad \mathsf{K}(\rho) = \sum_i K_i\rho K_i^\dagger \ . \tag{1.4}$$

The probability distribution after n iterations (i. e. n steps after initial state ψ) will be then described by density operator

$$\rho_n = \mathsf{K}^n(\rho_0) \ . \tag{1.5}$$

1.1.2 Stationary distribution

Definition 1.1.2. *We define stationary distribution (or incorrectly stationary state) of the quantum Markov model* $(\mathbf{H}, \mathbf{K})$ *as a density operator* ρ_s *satisfying*

$$\rho_s = \mathsf{K}(\rho_s) \ .$$

The stationary distribution is distribution on the set of states from $\mathbf{H}$ which does not change with iterations. In other words, if one element of the Markov chain is characterized by a stationary distribution ρ_s, then the next one will also be characterized by the same distribution ρ_s.

We have already mentioned the distribution after n iterations ρ_n. What if the limit of such distributions exist? The limit would well characterize the distribution on the states after the machine has been running for a long time. Moreover, such limit is a stationary distribution and thus does not change anymore.

Theorem 1.1.1. *Suppose that the limit* $\lim_{n\to\infty}\rho_n = \lim_{n\to\infty}\mathsf{K}^n(\rho_0)$ *exist. Then the limit is a stationary distribution.*

Proof.

$$\mathsf{K}(\lim_{n\to\infty}\mathsf{K}^n(\rho_0)) - \lim_{n\to\infty}\mathsf{K}^n(\rho_0) = \lim_{n\to\infty}\mathsf{K}^{n+1}(\rho_0) - \lim_{n\to\infty}\mathsf{K}^n(\rho_0) = 0$$

□

1.1.3 Word probabilities

What is the word in any language? Some sequence of characters which the language use. Still, in any language there are sequences which does not make sense, but for us the word will be just arbitrary sequence of such characters.

Definition 1.1.3. *The word of the length k is any ordered sequence of values from the set of possible characters $\{1, 2, ..., m\}$. For example,* 1925483 *is the word of length* 7 *from* $\{1, 2, ..., 9\}$.

Suppose we know the initial state $|\psi\rangle$ and set of possible values is $\{1, 2, ..., 9\}$. What is the chance of obtaining the word 43? It is just multiple of some other probabilities.

$$P(43, |\psi\rangle) = P(4, |\psi\rangle) P\left(3, \frac{K_4|\psi\rangle}{\sqrt{\langle\psi|K_4^\dagger K_4|\psi\rangle}}\right) = \langle\psi|K_4^\dagger K_4|\psi\rangle \frac{\langle\psi|K_4^\dagger K_3^\dagger K_3 K_4|\psi\rangle}{\langle\psi|K_4^\dagger K_4|\psi\rangle}.$$

Theorem 1.1.2. *Suppose we know the initial state $|\psi\rangle$. The probability of obtaining word $\boldsymbol{v} = (v_1 v_2 ... v_k)$, where $v_i \in \{1, 2, ..., m\}$, is*

$$P(\boldsymbol{v}, |\psi\rangle) = \langle\psi|K_{v_1}^\dagger K_{v_2}^\dagger ... K_{v_k}^\dagger K_{v_k} ... K_{v_2} K_{v_1}|\psi\rangle \ .$$

If we do not know the initial state but rather density operator ρ describing the actual state (with special case of stationary distribution), derivations are very similar. The chance of measuring 4 is

$$P(4, \rho) = \mathrm{Tr}(K_4 \rho K_4^\dagger) \ .$$

After that, density operator describing the actual state passes on density operator

$$\rho = \frac{K_4 \rho K_4^\dagger}{\mathrm{Tr}(K_4 \rho K_4^\dagger)} \tag{1.6}$$

describing the next state.

$$P(43, \rho) = \mathrm{Tr}(K_4 \rho K_4^\dagger) \frac{\mathrm{Tr}(K_3 K_4 \rho K_4^\dagger K_3^\dagger)}{\mathrm{Tr}(K_4 \rho K_4^\dagger)} \ .$$

Theorem 1.1.3. *Suppose we have density operator ρ describing the actual state. The probability of obtaining word $\boldsymbol{v} = (v_1 v_2 ... v_k)$, where $v_i \in \{1, 2, ..., m\}$, is*

$$P(\boldsymbol{v}, \rho) = \mathrm{Tr}(K_{v_k} K_{v_{k-1}} ... K_{v_1} \rho K_{v_1}^\dagger ... K_{v_{k-1}}^\dagger K_{v_k}^\dagger) \ .$$

Using theorems 1.1.2, 1.1.3 and definition (1.6) we can obtain any other thinkable word probability. For example, suppose that we want to obtain probability of obtaining word 43XX2 in the fourth step with initial state $|\psi\rangle$ — initial density operator $\rho_0 = |\psi\rangle\langle\psi|$, where X is an arbitrary letter. Such probability will be

$$P(XXX43XX2) =$$
$$P\big(4, \mathsf{K}^3(\rho_0)\big)\, P\left(3, \frac{K_4\mathsf{K}^3(\rho_0)K_4^\dagger}{\mathrm{Tr}(K_4\mathsf{K}^3(\rho_0)K_4^\dagger)}\right) P\left(2, \mathsf{K}^2\left(\frac{K_3K_4\mathsf{K}^3(\rho_0)K_4^\dagger K_3^\dagger}{\mathrm{Tr}(K_3K_4\mathsf{K}^3(\rho_0)K_4^\dagger K_3^\dagger)}\right)\right) =$$
$$= \mathrm{Tr}\left(K_2\mathsf{K}^2\left(K_3K_4\mathsf{K}^3(\rho_0)K_4^\dagger K_3^\dagger\right)K_2^\dagger\right).$$

As you can see, passing into the next state without knowing the outcome value is represented by acting of superoperator K on the actual density operator ρ, by contrast, passing into the next state with known outcome is represented by reduction of the density operator in accordance with definition (1.6). The final probability is then a multiple of probabilities of outcomes in different steps.

1.1.4 Parametrization of the quantum Markov model

The only free parameters are hidden in the operators $\{K_1, K_n..., K_m\}$ so if we parametrize these operators, we have parametrized the whole model. Suppose that we have some orthonormal (ortogonal and normalized) basis. Matrices of operators $K_k \in \mathbb{C}^{n\times n}$ will be in such basis parametrized as

$$\forall k = 1, 2, ..., m, \quad K_k = \begin{pmatrix} K_{11}^{(k)} & K_{12}^{(k)} & \dots & K_{1n}^{(k)} \\ K_{21}^{(k)} & K_{22}^{(k)} & \dots & K_{2n}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1}^{(k)} & K_{n2}^{(k)} & \dots & K_{nn}^{(k)} \end{pmatrix}. \tag{1.7}$$

From $K_1^\dagger K_1 + K_2^\dagger K_2 + ... + K_m^\dagger K_m = I$ we get n real-number constraints

$$\forall j = 1, 2, ..., n, \quad \sum_{k=1}^{m}\sum_{i=1}^{n} |K_{ij}^{(k)}|^2 = 1 \tag{1.8}$$

and $\frac{(n-1)(n)}{2}$ constraints with both real and imaginary part

$$\forall i = 1, 2, ..., n, \quad \forall j = i+1, ..., n, \quad \sum_{k=1}^{m}\sum_{l=1}^{n} \overline{K_{il}^{(k)}} K_{lj}^{(k)} = 0\ . \tag{1.9}$$

Eventually, we have $n + 2\frac{(n-1)(n)}{2} = n^2$ real-number constraints and thus $2mn^2 - n^2 = n^2(2m-1)$ free real parameters.

Theorem 1.1.4. *n-dimensional quantum Markov model with m possible outputs can be parametrized by $n^2(2m-1)$ real parameters. The parametrization can be computed from constraints (1.8) and (1.9).*

Conditions (1.8) and (1.9) are in general hard to solve, but for the 1-qubit quantum Markov model with 2 possible outputs we have found explicit solution. We used parametrization of hypersphere (in this case 4-sphere) to solve (1.8) and (1.9) led to 2 additional constraints.

$$K_0 = e^{i\xi}\begin{pmatrix} \cos\phi_1 & \cos\varphi_1 e^{ix} \\ \sin\phi_1\cos\phi_2 e^{iy} & \sin\varphi_1\cos\varphi_2 e^{i(x+y+z)} \end{pmatrix}, \tag{1.10}$$

$$K_1 = e^{i\sigma}\begin{pmatrix} \sin\phi_1\sin\phi_2\cos\phi_3 & \sin\varphi_1\sin\varphi_2\cos\varphi_3 e^{iv} \\ \sin\phi_1\sin\phi_2\sin\phi_3 e^{iw} & \sin\varphi_1\sin\varphi_2\sin\varphi_3 e^{i(v+w+u)} \end{pmatrix}, \tag{1.11}$$

where

$$\phi_2 = -\arctan\left(\frac{\sin(x+z)-\tan(x)\cos(x+z)}{X\tan\varphi_2}\right),$$

$$\phi_1 = -\arctan\left(\frac{\cos(x)}{Y\tan\varphi_1}\right),$$

$$\begin{aligned} X &= \cos\phi_3\cos\varphi_3(\sin(v)-\tan(x)\cos(v))+ \\ &\quad +\sin\phi_3\sin\varphi_3(\sin(v+u)-\tan(x)\cos(v+u)), \\ Y &= \cos\phi_2\cos\varphi_2\cos(x+z)+ \\ &\quad +\sin\phi_2\sin\varphi_2(\cos\phi_3\cos\varphi_3\cos(v)+\sin\phi_3\sin\varphi_3\cos(v+u)), \end{aligned}$$

where $\varphi_1, \varphi_2 \in [0,\pi]$, $\phi_3, \varphi_3, x, y, z, u, v, w, \xi, \sigma \in [0, 2\pi]$.

This parametrization has 12 real parameters, exactly how the theorem 1.1.4 predicted. Still, parameters ξ and σ, which characterize overall phase shift, have no physical meaning. They do not influence any probabilities and cannot be observed. So the effective number of substantial real parameters is $n^2(2m-1)-m$.

1.2 Implementation Markov model via Quantum Markov model, comparison of the classical and quantum

The crucial question is: "Are quantum Markov models more powerful than classical ones?" No, they are not. Still, we show that some Markov models can be implemented by quantum Markov models and some cannot. "So

are the classical ones more general? Neither they are. These models are just different. For example, quantum models possess the great property of superposition, which classical models pray to possess. Moreover, the classical Markov models do not have memory, since this Markovian property. By contrast, quantum models do have a memory — exactly one qu-n-digit, information, which can be stored in n orthogonal states. Thus the quantum Markov models are not by definition the pure Markov models, because the next state, in fact, does depend on the former states, although only vicariously. In the next chapter we will see that the extended definition of Quantum Markov model, Hidden Quantum Markov models, which possess also some kind of classical uncertainty, generalize both Markov and Hidden Markov models.

So how to implement classical Markov model via quantum? What exactly is the implementation? The implementation must have the very same properties as the classical Markov model, namely we require two characteristics:

1. The following state must be determined unambiguously only by the outputted value.

2. Probabilities of passing from the one state to another must remain the same as in its classical version.

Since in the classical version outputted value are identified with states themselves it follows that $n \equiv m$. Let the $\{|i\rangle\}_{i=1}^{n}$ be an orthonormal base the Hilbert space $\mathbf{H}$. We must identify each state s_i from the classical model with one basis vector $|i\rangle$. If we would not, some information could be stored between passings and the first characteristic would be broken. Classical Markov models do not allow any superposition and according to the characteristic 1) Kraus operators $\{K_i\}$ must take the form

$$\forall i = 1, 2, ..., n, \quad K_i = |i\rangle\langle\psi_i|. \tag{1.12}$$

where $|\psi_i\rangle$ is partially determined by the characteristic 2). All probabilities must remain the same, so $|\psi_i\rangle$ must be of the form

$$|\psi_i\rangle = e^{-i\varphi_{i1}}\sqrt{\mathbf{T}_{i1}}|1\rangle + e^{-i\varphi_{i2}}\sqrt{\mathbf{T}_{i2}}|2\rangle + \cdots + e^{-i\varphi_{in}}\sqrt{\mathbf{T}_{in}}|n\rangle \ , \tag{1.13}$$

where $\mathbf{T}_{ij}$ are the elements of the transition matrix, $\varphi_{i1}, ..., \varphi_{in}$ are some real parameters, which will be determined later.

In matrix representation

$$K_1 = \begin{pmatrix} e^{i\varphi_{11}}\sqrt{\mathbf{T}_{11}} & e^{i\varphi_{12}}\sqrt{\mathbf{T}_{12}} & \dots & e^{i\varphi_{1n}}\sqrt{\mathbf{T}_{1n}} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \tag{1.14}$$

$$\dots,$$

$$K_n = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ e^{i\varphi_{n1}}\sqrt{\mathbf{T}_{n1}} & e^{i\varphi_{n2}}\sqrt{\mathbf{T}_{n2}} & \dots & e^{i\varphi_{nn}}\sqrt{\mathbf{T}_{nn}} \end{pmatrix} \tag{1.15}$$

.

Still, the operators $K_1, ..., K_n$ must be trace-preserving, i. e. must satisfy the relation $K_1^\dagger K_1 + K_2^\dagger K_2 + ... + K_n^\dagger K_n = I$. This gives us additional constraints on the real parameters φ_{ij}. It follows that only Markov models, for which the solution of

$$\sum_{k=1}^{n} \sqrt{\mathbf{T}_{ki}}\sqrt{\mathbf{T}_{kj}} e^{i(\varphi_{kj}-\varphi_{ki})} = \delta_{ij} \tag{1.16}$$

exists, can be implemented. This solution may or may not exist, as you will see in the following example.

Suppose 1-qubit quantum Markov model with 2 possible outputs. The condition for $i = j$ is always satisfied due to theorem A.1.1. Conditions for $i \neq j$ then look like

$$\sqrt{\mathbf{T}_{12}}\sqrt{\mathbf{T}_{11}} e^{i(\varphi_{11}-\varphi_{12})} + \sqrt{\mathbf{T}_{22}}\sqrt{\mathbf{T}_{21}} e^{i(\varphi_{21}-\varphi_{22})} = 0 \tag{1.17}$$

$$\sqrt{\mathbf{T}_{11}}\sqrt{\mathbf{T}_{12}} e^{i(\varphi_{12}-\varphi_{11})} + \sqrt{\mathbf{T}_{21}}\sqrt{\mathbf{T}_{22}} e^{i(\varphi_{22}-\varphi_{21})} = 0 \tag{1.18}$$

From the first condition it follows

$$\sqrt{\mathbf{T}_{12}}\sqrt{\mathbf{T}_{11}} e^{i(\varphi_{11}-\varphi_{12})} = -\sqrt{\mathbf{T}_{22}}\sqrt{\mathbf{T}_{21}} e^{i(\varphi_{21}-\varphi_{22})} \quad \Rightarrow \tag{1.19}$$

$$\Rightarrow \quad \mathbf{T}_{12}\mathbf{T}_{11} = \mathbf{T}_{22}\mathbf{T}_{21}. \tag{1.20}$$

If we put $\varphi_{11} - \varphi_{12} = 0$ and $\varphi_{21} - \varphi_{22} = \pi$ the conditions (1.17) and (1.18) are satisfied. From equation (1.20) and theorem A.1.1 we can easily derive that only Markov models of the form

$$\mathbf{T} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix} \tag{1.21}$$

can be implemented. For example, Markov model with transition matrix

$$\mathbf{T} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} \end{pmatrix} \tag{1.22}$$

cannot be implemented.

We have seen that there is a difference between quantum and classical Markov models, that one cannot be in general implemented by the other,

nevertheless, many properties and their description remain analogous. The next table show the comparison.

classical	**quantum**
Every segment of the chain is represented by a state from a finite set of states.	Every segment of the chain is represented by a state from finite-dimensional Hilbert space.
Producing value and passing onto next state is arranged by a set of conditional probabilities.	Producing value and passing onto next state is arranged by a set of Kraus operators.
After measuring the value i the state s_j passes onto state s_i.	After measuring value i state $\|\psi\rangle$ passes onto state $\frac{K_i\|\psi\rangle}{\|K_i\|\psi\rangle\|}$.
Prediction is described by probability vector $\mathbf{p_n}$ and acting of the transition matrix $\mathbf{T}$.	Prediciton is described by density operator ρ_n and acting of the superoperator K.
Word probabilities are calculated using multiplication of the probabilities and acting of the transition matrix $\mathbf{T}$.	Word probabilities are calculated using Kraus operators and acting of the superoperator K.

1.3 Physical implementation of the Quantum Markov model

Are quantum Markov models physical? Or are they only some mathematical construct which have a little connection to reality? No, they are not only a mathematical construct. In the next two sections we show that there exist a lot of physical system which leads to the quantum Markov models. Moreover, in the second we show that every thinkable quantum Markov model is in principle realizable.

1.3.1 Physical example which leads to a quantum Markov model

The first example which leads to quantum Markov model is very simple. Suppose that instead of general operators we use the projective measurements. Suppose that we have observable

$$\mathbf{A} = \sum_{i=1}^{n} i|i\rangle\langle i|, \tag{1.23}$$

initial state $|\psi\rangle$ and we measure such observable repeatedly on the system. Kraus operators are nothing but the projectors themselves:

$$K_1 = |1\rangle\langle 1|,\ K_2 = |2\rangle\langle 2|, ..., K_n = |n\rangle\langle n|\ , \tag{1.24}$$

but this system is very simple. That is because after we measure value i on the state ψ, the state passes onto

$$\frac{|i\rangle\langle i|\psi\rangle}{\||i\rangle\langle i|\psi\rangle\|} \tag{1.25}$$

and iterative measurement will always lead to the value i. The only possible output string will be of the form

$$iiiiiiiiiiiiiiii... \tag{1.26}$$

For that reason we present more complicated example.

Suppose we have auxiliary m-dimensional Hilbert space $\mathbf{H}_A$ with orthonormal base $\{|i\rangle_A\}_{i=1}^m$. We take one of the basis states, for example $|1\rangle_A$, and create composite system $|1\rangle_A|\psi\rangle \in \mathbf{H}_A \otimes \mathbf{H}$. We entangle the system $|1\rangle_A|\psi\rangle$ using some unitary transformation $\mathbf{U}$ on the tensor product $\mathbf{H}_A \otimes \mathbf{H}$, getting

$$\mathbf{U}|1\rangle_A|\psi\rangle\ . \tag{1.27}$$

Now we measure observable

$$\mathbf{A}_A = \sum_{i=1}^{m} i|i\rangle_{AA}\langle i| \tag{1.28}$$

on the auxiliary system. After measuring the value i the composite system passes onto

$$|i\rangle_{AA}\langle i|\mathbf{U}|1\rangle_A|\psi\rangle\ . \tag{1.29}$$

After the measurement the composite system is no longer entangled (the result is written as the tensor product of the subsystems) and thus we can reset the auxiliary system $|i\rangle_A$ to the default setting $|1\rangle_A$, obtaining

$$|1\rangle_{AA}\langle i|\mathbf{U}|1\rangle_A|\psi\rangle\ . \tag{1.30}$$

The ${}_A\langle i|\mathbf{U}|1\rangle_A$ can be now comprehend as the linear operator on the system $\mathbf{H}$, in fact, we put

$$\mathbf{K}_i = {}_A\langle i|\mathbf{U}|1\rangle_A \ . \tag{1.31}$$

These operators satisfy the requested relation $K_1^\dagger K_1 + K_2^\dagger K_2 + ... + K_m^\dagger K_m = I_{\mathbf{H}}$:

$$\sum_{i=1}^{m} {}_A\langle 1|\mathbf{U}^\dagger|i\rangle_{AA}\langle i|\mathbf{U}|1\rangle_A = {}_A\langle 1|\mathbf{U}^\dagger \left(\sum_{i=1}^{m} |i\rangle_{AA}\langle i|\right)\mathbf{U}|1\rangle_A = {}_A\langle 1|1\rangle_A = I_{\mathbf{H}},$$

where we wrote projectors in this simplified form $|i\rangle_{AA}\langle i| \otimes I_{\mathbf{H}} \equiv |i\rangle_{AA}\langle i|$.

Now we know that after measuring value i state $|\psi\rangle$ passes onto state $K_i|\psi\rangle$. If we require normalization, resultant state is

$$\frac{K_i|\psi\rangle}{\|K_i|\psi\rangle\|} = \frac{K_i|\psi\rangle}{\sqrt{\langle\psi|K_i^\dagger K_i|\psi\rangle}} \ . \tag{1.32}$$

All we need to do to be our image complete is to verify that the probabilities of the obtaining value i remain the same as in definition 1.1.1. To do that we need to introduce density operator of the whole system before the measurement.

$$\rho = \mathbf{U}|1\rangle_A|\psi\rangle\langle\psi|_A\langle 1|\mathbf{U}^\dagger \tag{1.33}$$

Probability of measuring value i on the subsystem $\mathbf{H}_A$ is

$$\begin{aligned} P(i) &= \mathrm{Tr}_{\mathbf{H}}({}_A\langle i|\mathbf{U}|1\rangle_A|\psi\rangle\langle\psi|_A\langle 1|\mathbf{U}^\dagger|i\rangle_A) = \\ &= \mathrm{Tr}(K_i|\psi\rangle\langle\psi|K_i^\dagger) = \|K_i|\psi\rangle\|^2 = \langle\psi|K_i^\dagger K_i|\psi\rangle \ , \end{aligned} \tag{1.34}$$

exactly as the definition required. If we want to generate the whole Markov chain, we just keep putting the state from Hilbert space $\mathbf{H}$ to the machine again and again. In the next section we show that every thinkable quantum Markov model can be physically implemented using this scheme.

1.3.2 Constructing the physical model for a given quantum Markov model

Suppose we have given a quantum Markov model and we want to know how to implement it physically. One way and only way we show here is to use scheme sketched in the previous section. We will use a bigger Hilbert space, to be specific, a tensor product of the original space and auxiliary m-dimensional Hilbert space $\mathbf{H}_A \otimes \mathbf{H}$. We will construct the unitary operator on this bigger Hilbert space such that its reduction lead to Kraus operators. Concretely we require

$$\forall i = 1, 2, ..., m, \quad {}_A\langle i|\mathbf{U}|1\rangle_A = \mathbf{K}_i \ . \tag{1.35}$$

So the unitary operator will be in the basis $\{|i\rangle \otimes |j\rangle\}_{i=1,j=1}^{m,n}$ represented as $m \cdot n \times m \cdot n$ complex matrix

$$U = \begin{pmatrix} K_1 & u_{11}^{(12)} & u_{12}^{(12)} & \dots & \dots & u_{11}^{(1m)} & u_{12}^{(1m)} & \dots \\ & & \dots & & & & \dots & \\ K_2 & u_{11}^{(22)} & u_{12}^{(22)} & \dots & \dots & u_{11}^{(2m)} & u_{12}^{(2m)} & \dots \\ & & \dots & & & & \dots & \\ \vdots & & \vdots & & \ddots & & \vdots & \\ K_m & u_{11}^{(m2)} & u_{12}^{(m2)} & \dots & \dots & u_{11}^{(mm)} & u_{12}^{(mm)} & \dots \\ & & \dots & & & & \dots & \end{pmatrix} . \tag{1.36}$$

All we need is to find complex numbers $u_{ij}^{(kl)}$ such that $\mathbf{U}$ is unitary, i. e.,

$$\mathbf{U}^\dagger \mathbf{U} = \mathbf{I} . \tag{1.37}$$

From the condition $K_1^\dagger K_1 + K_2^\dagger K_2 + ... + K_m^\dagger K_m = I$ we have many equations from the (1.37) already fulfilled and we can always find solutions of the rest. Though, the solutions are not determined unambiguously. Let us compute the number of free real parameters. To do that we need to identify number of real parameters and number of constraints given by equation (1.37) and then subtract these.

Number of the real parameters is $2n^2m(m-1)$. The following table shows the number of constraints.

orthogonality of the first n columns to the others
$2n^2(m-1)$
orthogonality of the others among themselves
$n(m-1)(n(m-1)-1)$
normalization of the others
$n(m-1)$

Number of the free real parameters is then $2n^2m(m-1) - 2n^2(m-1) - n(m-1)(n(m-1)-1) - n(m-1) = n^2(m-1)^2$.

The concrete solution of the (1.37) is very hard to find, but at least we know that it is possible. If we manage to find the solution we are facing other problem — how to construct such unitary matrix? Nonetheless, such question goes far beyond the scope of this work.

Chapter 2

Hidden Quantum Markov Models

Definition 2.0.1. *Let* $\mathbf{H}$ *be n-dimensional Hilbert space,*

$$\mathbf{K} = \{K_1^{(1)}, K_1^{(2)}, ..., K_1^{(d_1)}, K_2^{(1)}, K_2^{(2)}, ..., K_2^{(d_2)}, ..., K_m^{(1)}, K_m^{(2)}, ..., K_m^{(d_m)}\}$$

set of linear operators on $\mathbf{H}$*. We call couple* $(\mathbf{H}, \mathbf{K})$ *n-dimensional Hidden Quantum Markov Model with m possible outputs and degenerations* $d_1, d_2, ..., d_m$ *if and only if the following conditions are satisfied:*

1. $\sum_{i=1}^{m} \sum_{p=1}^{d_i} {K_i^{(p)}}^\dagger K_i^{(p)} = I$
2. *Vectors from* $\mathbf{H}$ *form a growing chain in the following sense:*
   (a) *Each segment of the chain is represented by normalized vector(≡ state) from* $\mathbf{H}$*.*
   (b) *If the segment is* $|\psi\rangle$*, the next segment of the chain will be* $\frac{K_i^{(p)}|\psi\rangle}{\|K_i^{(p)}|\psi\rangle\|}$ *with probability*

$$P = \|K_i^{(p)}|\psi\rangle\|^2 = \langle\psi|{K_i^{(p)}}^\dagger K_i^{(p)}|\psi\rangle$$

   while value i is produced. ($i \in \{1, 2, ..., m\}$*)*

As you can see, the definition is very similar to the first one 1.1.1. The only change is that now we do have more Kraus operators than outputted values. The change is than now some new classical uncertainty comes to the stage. In the non-hidden Quantum Markov models, knowing the initial state and outputted values we could have compute every following state. It is not possible, because the outputted value still gives us some uncertainty about the state. For instance, if the outputted value is number 2 with degeneration

$d_2 = 3$ and the initial state is $|\psi\rangle$, there are still three possibilities of the next state

$$\frac{K_2^{(1)}|\psi\rangle}{\|K_2^{(1)}|\psi\rangle\|}, \quad \frac{K_2^{(2)}|\psi\rangle}{\|K_2^{(2)}|\psi\rangle\|}, \quad \frac{K_2^{(3)}|\psi\rangle}{\|K_2^{(3)}|\psi\rangle\|} \ .$$

These states remain hidden, the only thing we see now are the outputted values, what makes the Hidden Quantum Markov Models totaly analogous to the classical ones (see appendix B). This uncertainty about the next states gives us no other choice than using the density operators for describing them, which is analogous to the probability vector[1] in Hidden Markov Models.

2.1 Evolution and the word probabilities

In order to obtain very compact and efficient expression of the word probabilities and density operator evolution we need to define new linear superoperators. Let ρ be a density operator. We define

- Evolution superoperator for the unknown outputted value

$$\mathsf{K}(\rho) = \sum_{i=1}^{m} \sum_{p=1}^{d_i} K_i^{(p)} \rho K_i^{(p)\dagger} \tag{2.1}$$

- Evolution superoperator for the outputted value i

$$\mathsf{K}_i(\rho) = \sum_{p=1}^{d_i} K_i^{(p)} \rho K_i^{(p)\dagger} \tag{2.2}$$

If one of the k-th segment≡state of the chain is described by a density operator ρ_k and we do not know the outputted value or we want to predict the future, the next state will be described by density operator

$$\rho_{k+1} = \mathsf{K}(\rho_k) \ . \tag{2.3}$$

If the outputted value is i, the next state will be described by density operator

$$\rho_{k+1} = \frac{\mathsf{K}_i(\rho_k)}{\mathrm{Tr}(\mathsf{K}_i(\rho_k))} \ . \tag{2.4}$$

Let ρ_0 be an initial state. Probability of the word $123XX43X2$, where X refers to an arbitrary character, is

$$P(123XX43X2) = \mathrm{Tr}(\mathsf{K}_2\mathsf{K}\mathsf{K}_3\mathsf{K}_4\mathsf{K}\mathsf{K}\mathsf{K}_3\mathsf{K}_2\mathsf{K}_1(\rho_0)) \tag{2.5}$$

[1] describing the probability distribution over the set of the hidden states

and the density operator describing the following state is

$$\rho_9 = \rho_{|123XX43X2} = \frac{\mathsf{K}_2\mathsf{K}\mathsf{K}_3\mathsf{K}_4\mathsf{K}\mathsf{K}\mathsf{K}_3\mathsf{K}_2\mathsf{K}_1(\rho_0)}{\mathrm{Tr}(\cdot)} \ . \tag{2.6}$$

2.2 Implementation of the Hidden Markov models and Markov models via Hidden Quantum Markov models

The idea of implementation of the classical Hidden Markov models is quite easy. For each hidden state we just assign the quantum state which is orthogonal to the others and the density operator will take the role of the probability vector. The relatively hard part is finding such set of Kraus operators that represent a given Hidden Markov model accurately, i. e., the statistical behavior like word probabilities and probability distribution over the set of hidden states have to be the very same. Since number of the hidden states in Hidden Markov model is n we will use the n-dimensional Hilbert space with orthonormal basis $\{|p\rangle\}_{p=1}^{n}$.

Theorem 2.2.1. *Let $P(Y = y, X = s|\tilde{X} = \tilde{s}) \equiv \mathbf{P}_{ys\tilde{s}}$ be the transition probabilities of the Mealy Hidden Markov model (see appendix B). The set of Kraus operators which represent the given Hidden Markov model accurately is*

$$i = 1, ..., m, \quad p, q = 1, ..., n, \quad K_i^{(p,q)} = \sqrt{\mathbf{P}_{ipq}}|p\rangle\langle q| \tag{2.7}$$

Proof. First, we need to verify that definition (2.7) is a set of trace-preserving Kraus operators. Indeed it is, since

$$\sum_{i=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n} {K_i^{(p,q)}}^\dagger K_i^{(p,q)} = \sum_{i=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n} \sqrt{\mathbf{P}_{ipq}}|q\rangle\langle p|\sqrt{\mathbf{P}_{ipq}}|p\rangle\langle q| =$$
$$= \sum_{i=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n} \mathbf{P}_{ipq}|q\rangle\langle p|p\rangle\langle q| = \sum_{q=1}^{n}|q\rangle\langle q|\sum_{i=1}^{m}\sum_{p=1}^{n}\mathbf{P}_{ipq} = I \ .$$

We may suppose that the initial state is one of the basis states. If not, we just wait for one iteration and we can be sure that the state will be basis vector, since all of $K_i^{(p,q)}$ project to some.[2] If the initial state is $|\tilde{p}\rangle$, the probability of passing onto basis state $|p\rangle$ while outputting the value i is

$$\sum_{q=1}^{n}\mathrm{Tr}(K_i^{(p,q)}|\tilde{p}\rangle\langle\tilde{p}|{K_i^{(p,q)}}^\dagger) = \sum_{q=1}^{n}\sum_{\tilde{q}=1}^{n}\langle\tilde{q}|K_i^{(p,q)}|\tilde{p}\rangle\langle\tilde{p}|{K_i^{(p,q)}}^\dagger|\tilde{q}\rangle = \mathbf{P}_{ip\tilde{p}} \ .^{3}$$

[2] In fact, the first iteration is the only one where HQMM may differ from HMM.

[3] Actually, this is the probability of obtaining eigenvalues $i, p, \tilde{q}$ in the measurement.

So the implementation via Hidden Quantum Markov model behaves in the very same way as the original, thus all of the statistical behaviour must be the same. The density operator is now diagonal in the orthonormal basis $\{|p\rangle\}$ and its diagonal elements are now the very same as in the probability vector $\mathbf{p}$ in classical version. □

Implementing of the non-hidden Markov model via HQMM can be done using set

$$i = 1, ..., m, \quad p, q = 1, ..., n, \quad K_i^{(p)} = \sqrt{\mathbf{T}_{ip}}|i\rangle\langle p| \ ,$$

so after measuring i the states passes onto state $|i\rangle$ which is exactly what we expect from the Markov model.

Now, we can easily compute the number of Kraus operators needed for implementing hidden and non-hidden Markov models. It is $m \cdot n^2$ for the Hidden Markov models and n^2 for the non-hidden. Such huge amount of resources seem to be practically impossible to obtain, so the previous theorem may be simply a fundamental statement without any practical usage.

2.3 Physical Realization of the Hidden Quantum Markov Model

Once we achieve to physically realize any Quantum Markov Model, we can easily construct any Hidden Quantum Markov Model. In the non-hidden case we assign one Kraus operator K_i to each outputted value i. We can think of the hidden case in the very same way. We assign one Kraus operator $K_i^{(p)}$ to each couple of outputted values i, p but the only difference is that we do not see the hidden value p. That makes from the Quantum Markov Model the Hidden Markov Model. It is just omitting some information classically.

Practically, if we have the couple of outputted values i, p on our monitor, we just do not look at the value p and from the Quantum Markov model we have immediately gained the Hidden QMM. Another example would be the case, where the measurement device knows about both hidden and non-hidden outputted value and is able to project the previous state $|\psi\rangle$ to another $\frac{K_i^{(p)}|\psi\rangle}{\|\cdot\|}$, but for some reason it shows only the value i. Mathematically

Since we are in *Hidden* Quantum Markov model, we throw information about values p, q out, so we do not know what is the consequent state, but that does not mean that the Machine does not know. This is the reason why we can calculate the probability in this way. From the form of the Kraus operator $K_i^{(p,\tilde{q})}$ it is obvious that after measuring "hidden" eigenvalues $p, \tilde{q}$ and non-hidden i the consequent state will be $K_i^{(p,\tilde{q})}|\tilde{p}\rangle$. We sum over q's since for all q $K_i^{(p,q)}$ lead to eigenvector $|p\rangle$.

speaking, we are using a projective function

$$f : (i, p) \longrightarrow i \tag{2.8}$$

and we can formally write

$$\text{HQMM} = \text{QMM} \oplus f \ . \tag{2.9}$$

2.4 Comparing Hidden Markov Models and Hidden Quantum Markov Models

As we have seen, the Hidden Quantum Markov models are the generalization of the non-quantum. But are they more effective? What makes them better? And what are the other differences? Does exist an effective way of comparing them? As we said earlier, the superpositions make them more flexible and the statistical behavior has to be more complex. In the classical version we have fixed and finite amount of the transition probabilities. By contrast, we have infinite, even uncountable set of possible states in the quantum version and thus potentially infinite number of transition probabilities. Also the definition 2.0.1 gives us more freedom in the number of Kraus operators (In the classical version for the given number of states and outputs we had fixed amount of transitions), which may be a big problem in comparing these (what amount of Kraus operators should we use?). This problem may be solved using the Stinespring theorem [3] about superoperators and their decomposition onto set of Kraus operators, but we do not know now. That and whether other differences are statistically important or not would be the task of the further research.

Chapter 3

Entanglement

Entanglement is one of the mightiest mystery in the quantum mechanics. It shows us how the quantum world differs from the classical. Suppose you have two particles which once interacted to each other. Now they are entangled. What does it mean? It means that by measuring one at the arbitrary time we can gain some information about the other. You could argue that it is the same in the classical mechanics: If you measure the momentum of the first particle and know the total momentum, you can easily calculate momentum of the second, so you gained information about the other. But these correlations are in quantum mechanics much stronger than in classical. The above classical example was a simple law of conservation and that is all what the classical case can offer. By contrast, in quantum mechanics we can predict the probability of the measurement outcome of the second particle knowing something about the first, but it is not just a simple law of conservation, because actually nothing is conserved actually[1] To prove that, suppose a simple yes-no experiment, denote the first particle A and the second B. In quantum mechanics it is possible to obtain multiple possible outcomes with the same initial state and measuring device setting

$$A\ yes\ B\ yes,\ \ A\ no\ B\ yes,\ \ A\ yes\ B\ no,\ \ A\ no\ B\ no, \tag{3.1}$$

It cannot be the law of conservation, because the conserved value has to be the same for system before the measurement and after that. This is the main behavior of the quantum mechanics. We have the very same beginning and still we can end in different results. There has been done a lot of work about studying the upper problem. It is known as EPR paradox and we will not deal with it here.

[1] To be honest, some people think that the quantum information hidden in the system is conserved, which is pretty new topic and nowadays under strong investigation. See [11].

Now a little of mathematical formalism. Let $\{|0\rangle, |1\rangle\}$ be an orthonormal basis. One of the possible entangled state (which is called Bell state – maximally entangled state) of two particles can be

$$\frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle) \ . \tag{3.2}$$

If the Alice measure the 0 on the first particle we know that the Bob will measure the 1 and vice versa, if the Alice measure the 1 on the first particle, Bob will measure 0 on the second. This is exactly what we have been talking about in the beginning. Making a measurement on the first particle provide us some information about the other. If we choose to measure in another orthonormal base (for example $\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$), we can get every possible result sketched in the equation (3.1).

The one of the possible study of the entanglement with all of its consequences are the Delayed Choice and the Quantum Eraser experiments of which we will talk in the next chapter.

At the end of this very brief introduction we would like to point out that entangled states are very usual in the quantum mechanics - every interacting system of many particles will became entangled through the interaction. On the other hand, every system can be in principle disentangled either of measuring one part of the complex system or by using the inverse unitary of that which entangled the system.[2]

[2]Closed system always evolve through some unitary operator U and for every unitary there exist the inverse U^{-1}, applying the U^{-1} to the entangled system will push it backwards and disentangle it. The question is if every inverse can be physically implemented. Specially, for large systems with interacting particles it seems quite impossible to make them separated again.

Chapter 4

Delayed Choice and Quantum Eraser

The Delayed Choice and Quantum Eraser experiment exploit how the quantum world differs from our everyday experience. It shows the non-locality of quantum particles, their wave-like behavior and how closely attached the observer is to the observed system. It also demonstrate how irrelevant the system is when no-one is looking.

Let us briefly explain the main ideas of the Quantum Eraser and Delayed choice. Both of these study the interference or not interference patterns depending on the system setting.

The Quantum Eraser experiments are based on the simple idea that distinguishable (even in principle) histories (space-time trajectories) of one particle cannot interfere. But if we make these histories again indistinguishable, they will interfere. We *erase* the information about the histories and we regain the interference. The simple quantum eraser experiment is shown on figure 4.1.

On the other hand, the Delayed Choice experiments are trying to expose difficulties with the classical interpretation and understanding physics in a way "One particle is here a the second there." It shows that asking questions as "Where exactly the particle is?" or "What is happening right now?" is irrelevant, unless it refers to a specific measurement. The Delayed Choice experiments shows that unambiguous answers to such questions would lead to logical paradoxes. The original (although with some improvements) Wheeler's Delayed Choice experiment is shown on the figure 4.2

In the classical point of view when the particle of light (photon) hits the beam-splitter it has 50 percent chance to be reflected and 50 percent to go through. Then it is reflected from a mirror and go through a lens. After that it hits the mirror and is reflected to the one of two possible measurement

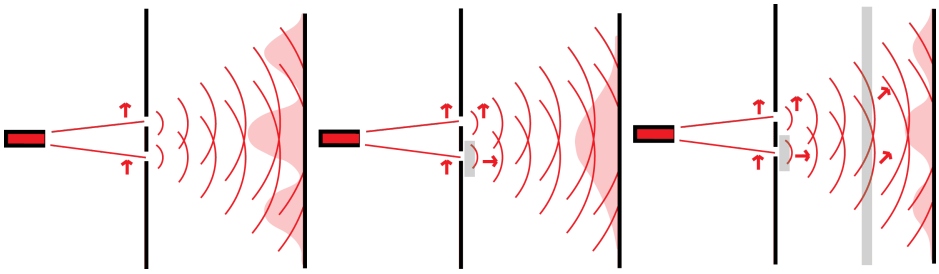


Figure 4.1: Quantum Eraser. On the first picture we have the classical double-slit experiment and we do see the interference pattern. But if we put the half-wave plate in front of one slit, it turns the polarization of the light and stop interfere, because we could in principle distinguish the path by a polarization measurement. Still, we can erase this information by putting the polarization filter and regain interference.

apparatus. Which one of the measuring devices we choose to use is decided by the position of rotatable mirror. If the mirror is positioned vertically we measure whether the particle went through upper part of the experiment (has gone through the beam-splitter) or lower part (has been reflected from the beam-splitter), if the mirror is positioned horizontally we measure the interference pattern. The point is we can rotate the mirror anytime, for instance right after the photon went through the beam-splitter. This is called a delayed choice.

If we have the rotatable mirror positioned vertically we will find out whether the particle went through upper or lower part. So there is not a paradox with the classical point of view. But we can also set the mirror horizontally and we will see the interference pattern. The question whether the particle went through upper or lower part now seems to be irrelevant, since it must have gone through both paths simultaneously, in order to interfere.

The problem is with the wrong deduction: if the photon has been detected in the detector $D1$, it must have gone through the lower part. But the measurement says nothing about what happened *before* the measurement was done. Whether the particle will hit detector $D1$ or detector $D2$ is not decided at the time of the passing through the beam-splitter but at the time detecting.

This experiment shows us that we have to stop ask what is happening when we are not looking. What has happened is decided at the time of the measurement. All we can imagine is that the photon go through every possible history simultaneously and calculate the probabilities of possible outcomes. To highlight this fact, let me offer you the real (not only thought)

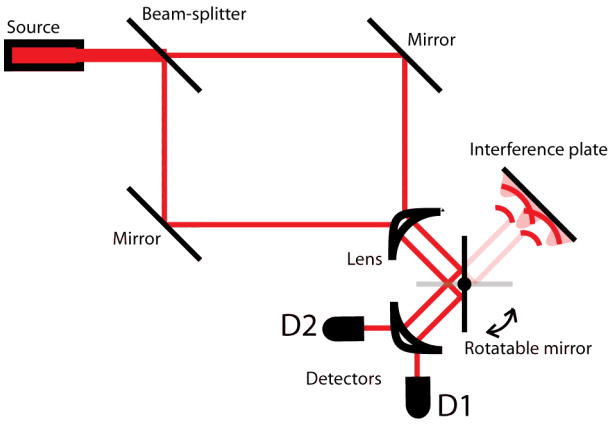


Figure 4.2: Wheeler's original delayed choice experiment

experiment which has been done [8]. In this experiment there a photon which decays with some probability at a given time (when passing the BBO crystal) into two other lower-energy photons. So we have one photon in the beginning and two at the end. The interesting thing is that there are two possible times when the decay can happen. And the two photons created at this first time interfere with those created at the second time. If you start to ask what is actually happening in the system, you will get to the conclusion that the energy conservation law has been violated! Because there must have been a time where the original photon and two created photons existed simultaneously! In fact, the conservation law has not been violated, because always when you look at the system, you will find either two less-energy photons or one high-energy. The conservation laws apply to measurements only, not on what is happening between them!

These experiments show us that we need to throw out the old comprehending of reality and thinking in a new way. All systems are evolving through every possible history and the only relevant questions are those referring to some measurement. A little insight about the measurement may offer you chapter **??**.

Chapter 5

Free Will Experiment

5.1 The idea

The Free Will Experiment is a Delayed Choice Quantum Eraser experiment, which combine both Delayed Choice and Quantum Eraser. We have it based on the work of [5] and using very similar experimental setup but for a different purpose. The main idea is the following. Having an entangled pair of particles, keeping one and letting the other for Alice, could we predict by measuring the first of the pair what Alice will do with the second? If this is possible then it would mean that the Alice has not the free will. She would be only the actress who plays her role, but everything she does would be predetermined.

The scheme is on figure 5.1.

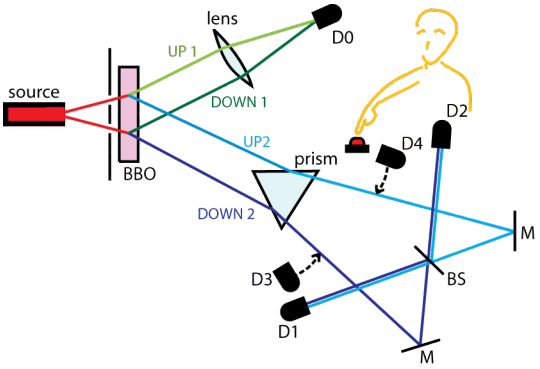


Figure 5.1: Free Will experiment. The tested person decides whether to push the button or not.

A high-energy photon goes through a double-slit and hits the BBO[1] crystal. Then it splits into an entangled (in polarizations) pair of photons of which one (1) goes up and one (2) goes down. But we will not use the entanglement in polarizations but rather in position. The state of the original photon just after it passed the double-slit and before it hits the crystal is superposition of states "Went through the upper slit"—$|Upper\rangle$ and "Went through the lower slit"—$|Lower\rangle$:

$$\frac{1}{\sqrt{2}}(|Upper\rangle + |Lower\rangle) \ . \tag{5.1}$$

Then it hits the crystal and since the superposition is hereditary, the position-entangled pair is produced:

$$\frac{1}{\sqrt{2}}(|UP1\rangle|UP2\rangle + |DOWN1\rangle|DOWN2\rangle) \ . \tag{5.2}$$

Where $|UP1\rangle$ means the state where the first photon went through the upper path (Light green on the picture), and $|DOWN1\rangle$ the state where the first photon went through the lower path (Dark green). The entanglement (5.2) just means that the two photons had to have the common place of creation. If the first was created in the place just after the upper slit, the second would have been created in the same place. We measure whether the first photon will interfere or not by detector $D0$.[2] From the (5.2) we know that if we try to measure which way the second photon went (person-Alice decides to push the red button which pushes the detectors $D3$ and $D4$ in both possible ways of the second photon) we will also find out the path of the first photon. If the second photon went through the lower path $|DOWN2\rangle$, the first would have to go through the lower path too $|DOWN1\rangle$. So the first photon cannot interfere, because the paths are in principle distinguishable. But if Alice decides not to push the red button, the information about the path would be erased by a beam-splitter. By hits in detectors $D1$ and $D2$ we cannot say, which way did the second and thus also the first photon went. So the first photon should to interfere again.

Notice one thing. Our reasoning did not depend on the time at all. It does not matter if you make the path of the first photon longer or shorter than the path of the second. Also the time evolution was only some irrelevant phase factor. So we can make the trajectory of the first photon much smaller than the trajectory of the second (We put the detector $D0$ closer than detectors

[1] Beta Barium Borate

[2] In fact, we measure position of hits of single photons, which may or may not form an interference pattern.

$D1, 2, 3, 4$). Now, we are able to predict what the Alice do! If we will see the interference pattern, we know that Alice will decide not to push the button and erase the information. If we will not see the interference pattern, Alice will decide to push the button and find out which way the second photon went and destroy the interference of the first. The problem is, it does not work. Why it does not work can be easily seen from the calculations.

Let $|D1\rangle$, $|D2\rangle$ be the states representing the hit on the detectors $D1$, $D2$ respectively. The final state after the second photon went through beam-splitter and the "which-path" information has been erased is

$$\frac{1}{2}(|UP1\rangle + |DOWN1\rangle)|D1\rangle + \frac{1}{2}(|UP1\rangle - |DOWN1\rangle)|D2\rangle \ . \tag{5.3}$$

So if the second photon hit the detector $D1$ the state of the first is

$$\frac{1}{\sqrt{2}}(|UP1\rangle + |DOWN1\rangle) \tag{5.4}$$

and it interferes. If the second photon hit the detector $D2$ the state of the first is

$$\frac{1}{\sqrt{2}}(|UP1\rangle - |DOWN1\rangle) \tag{5.5}$$

and it interferes too. The problem is the minus sign. Because the interference pattern cannot be seen as a whole with a single photon, we need to use many of them. But half of the second photons will be detected at the detector $D1$ and half of them will be detected at the detector $D2$, so half of the first photons will form the interference pattern in the shape of (5.4) and half of the first photons will form the interference pattern in the shape of (5.5) and we will not see the interference pattern at all. Since these interference patterns are exactly $\pi/2$ shifted, they add together and form the very same pattern as the non-interference one. The interference could be seen only when we filter out the first photons which refers to the hits on the detector $D1$ and then we will see the pattern in the shape of (5.5), but that cannot be done without some information from Alice! That ruins the Free Will test entirely. Can be this experiment corrected somehow, in order to work? No. As we will see in the next section, every Free Will test based on this entanglement idea is predetermined to fail.

But this experiment shows another nice thing. It shows that past is correlated with future and vice versa, future is correlated with past. It does not really matter if you measure the interference of the first photon first, or if Alice is deciding first. The correlations only mean that if you know some information from the measurement about the second, you are able to predict better the measurement outcomes of the first particle. The old

view of the collapse of the wave function as a time-bounded event totaly fail. Usually the collapse of the wave function is considered at the time of the measurement. This approach is illegitimate, since it really does not matter what measurement happened first and for some (for example moving) observers it may appear that the measurement number one happened first and for other (for example moving in the opposite way) observers it may appear that the measurement number two happened first. In fact, the only relevant time, when the state function collapses, is the time of hearing some new information about the system. Suppose that I hear from the Alice that she has detected the second particle in the detector $D1$. Then I know I should expect a hit somewhere where it is allowed to by (5.4). This is the time when the state function collapses. As we will see in the next section, the collapse of the wave function is generally the change from using one density operator describing our system to the conditioned one, which describes our system better. Or, in this special case of free will, from the density operator to a pure state.

5.2 Is there any general rule, why does it fail? Correlations are not causality, information cannot be send through entanglement

Suppose again that the Alice is the tested person and we are measuring the first photon in order to find out what she decides to do. The whole system before Alice decides what to do is described by a density operator

$$\rho = |\psi\rangle\langle\psi| \ , \tag{5.6}$$

where $|\psi\rangle$ is the state (5.2). Denote the system consisting of the first photon System A and the Alice's photon System B. We have access to the system A only, so our system is described by a density operator

$$\rho_A = \mathrm{Tr}_B(\rho) \ . \tag{5.7}$$

For now, suppose that Alice is deciding to apply some Unitary U to her photon=System B. This unitary will act on the whole system ρ as a tensor product, unitary

$$\tilde{U} = I_A \otimes U \ . \tag{5.8}$$

Does it affect our system somehow? Will we notice that Alice have just done something? No! Because our system is now described by a density operator

$$\rho_{A,\text{After Alice's decision}} = \mathrm{Tr}_B(\tilde{U}\rho\tilde{U}^\dagger) = \mathrm{Tr}_B(\tilde{U}\rho\tilde{U}^{-1}) = \mathrm{Tr}_B(\tilde{U}^{-1}\tilde{U}\rho) = \rho_A \ , \tag{5.9}$$

So we will not recognize a thing, because the density operator describing our system is the very same as it was before.

Now, suppose that Alice is deciding to apply some non-unitary K to her photon. To prove that we will get the very same result as before we need to do a little trick - which is in fact correct. Now, because she herself is the tested person, she must be considered as a part of the tested system. She decides if she apply to one measurement device or the other. And the measurement devices are also part of tested system. We need to expand our Hilbert space! We expand the system B to the new system B_{new}, which is now

$$B_{\text{new}} = B \otimes (\text{Measurement devices}) \otimes (\text{Alice}) \ , \tag{5.10}$$

And the whole system is now Hilbert space

$$\mathbf{H} = A \otimes B_{\text{new}} \tag{5.11}$$

Now, whatever Alice decides to do, it must be some unitary U on the composite system B_{new}, since B_{new} is a closed system. The unitary U is the expansion of her decision operator K

$$K = \text{Tr}_{\text{Measurement devices and Alice}}(U) \tag{5.12}$$

and the unitary U acts on the whole system $\mathbf{H}$ again as the unitary $\tilde{U} = I_A \otimes U$. Let $|\varphi\rangle$ be an initial state of the measuring devices and Alice. The initial density operator before the Alice's decision is now

$$\rho = |\psi\rangle|\varphi\rangle\langle\varphi|\langle\psi| \tag{5.13}$$

and density operator describing our system after the decision is now

$$\rho_{A,\text{After Alice's decision}} = \text{Tr}_{B,\text{Measurement devices and Alice}}(\tilde{U}\rho\tilde{U}^{\dagger}) = \cdots = \rho_A \ . \tag{5.14}$$

We have proved that no matter what Alice do we do not see the difference on our system. That is because the entanglement is yielding some information about what happened only, what made the particles entangled, not about what will happen. The entanglement yield the information about the past, not about the future. Still, it can be used for example for quantum teleportation [9], which actually happen later *after the particles are entangled*, but to complete the teleportation successfully sending classical information is required.

Since the above derivation was entirely general (the example with Alice was used only for a better understanding), **we have proved effectively that no information can be send through entanglement**. That is because correlations are not causality.

Moreover, if the that was not true, we could predict the future and communicate with past. On the other hand, the presumption that the evolution is unitary process was the crucial part of the proof. If the evolution of the closed system had not been unitary, such free will experiments would have been possible.

Chapter 6

Wigner's friend as one of the possible explanations what is happening, relative reality

Wigner's friend is an extension of the Schrödinger's cat and is one of possible looks on the measurement. The cat, the experimenter and his friend (the Wigner's friend). The idea is that when the experimenter open the box, he will see the dead or alive cat. But when he opened the box, he stopped being the outer observer. He just entangled himself with the system, so now he is the part of the (bigger) system

$$\text{cat} \otimes \text{experimenter}. \tag{6.1}$$

Suppose he likes cats and his happy when the cat is alive and sad when the cat is dead. So from the outside the state of the composite system cat $\otimes$ experimenter will look like

$$|\text{alive cat}\rangle|\text{☺}\rangle + |\text{dead cat}\rangle|\text{☹}\rangle \ . \tag{6.2}$$

So when the friend come he will measure not only the system of the cat in the box, he will measure the whole system cat $\otimes$ experimenter and see the alive cat and happy experimenter or dead cat and sad experimenter. By this second measurement he also entangled himself with the system.

Why the measurement is not the unitary evolution is because both experimenter and his friend cannot see themselves as part of the superposition. But that does not mean that from the outside they cannot be seen in superpositions and could not for example interfere.

While the experimenter is opening the box, it seems to the friend that system cat $\otimes$ experimenter is evolving through Unitary U,

$$(|\text{alive cat}\rangle + |\text{dead cat}\rangle)|:-|\rangle \xrightarrow{U} |\text{alive cat}\rangle|\text{☺}\rangle + |\text{dead cat}\rangle|\text{☹}\rangle \ . \tag{6.3}$$

This is what we have used in the previous section. We have considered Alice as part of the system and from our point of view she and system B evolve through unitary U.

The Wigner's friend approach also predict the existence of relative realities, still, no logical paradoxes are not coming out. Each observer has his own reality. To prove that, suppose we have a double slit and the experimenter will look which way the photon went. Then he find out that the photon went through upper or lower slit only and from his point of view the photon cannot interfere. But for the (Wigner's) friend he is now in entangled state

$$U|experimenter\rangle|photon\rangle \ . \tag{6.4}$$

If the friend decide to look at the (composite) system he finds out that the photon went really through one slit only. So there is not a logical paradox. But if he will not decide to look at the system and applies the U^{-1}[1] to the system instead. He will regain interference. Because the system he looks at then is again

$$U^{-1}U|experimenter\rangle|photon\rangle = |experimenter\rangle|photon\rangle \ . \tag{6.5}$$

So by applying this inverse unitary he changed the system back again before the experimenter looked at the slits. But that completely erased his memory! So there might have been a time when he knew that the photon went through one slit only, and for you it seems that the photon went through both slits simultaneously. But at a given time there is not logical paradox, because before you have seen the interference you had to erased memory of all people who knew that the photon cannot interfere. It seems that changing the reality is OK as long as nobody notices.

This is actually a real Quantum Eraser experiment and it works for tiny particles (as photons or electrons) already. The problem is with a bigger system it might be almost impossible to apply the inverse unitary. Just consider Alice looking at the slits. How could you make her unlook when she is consisted of so many particles?

[1]Which is in principle possible to apply. For example for the Beam-splitter U the inverse unitary U^{-1} is again a Beam-splitter.

Chapter 7

Conclusion

In this work we have shown that the idea of Hidden Quantum Markov models is consistent, that they generalize the classical case and may model some real physical systems. We have also shown the basic statistical quantities for this models and how to calculate them, for example the density operator describing each segment of a chain and the word probabilities. What we do not know is how much these models generalize the classical case if some applications may be found. These are the main aims for the further research.

In the second part of the work we have talked about the Free Will test based on the delayed choice quantum eraser experiments and explained why this kind of experiments could not work. We have also proved effectively that the correlations are not causality and that any form of information cannot be send through the entanglement alone. At the very end we have shown how it is related to the Wigner's friend approach and what else can be derived from this approach.

Appendix A

Markov Models

In this chapter we present simplified classical Markov models in a way which is necessary for the understanding of the quantum Markov models. For better understanding the classical ones see, for instance, [4], [6] or [10].

Definition A.0.1. *Let $S = \{s_i\}_{i\in I}$ be a finite or countable set of different states, $I = \{1, 2, 3, ...\}$ is an index set. Makrov chain is an growing chain of such states, where probability of the next being in the state s_i depends only the previous state s_j, this is called Markovian property. For explicit formulation, let the $X_1, X_2, ..., X_n$ be the random variables describing elements 1,2,...,n in a Markov chain. The Markovian property can be now written as*

$$P(X_n = s_i | X_{n-1} = s_j, X_{n-2} = s_{j_{n-2}}, ..., X_1 = s_{j_1}) = P(X_n = s_i | X_{n-1} = s_j)$$

for all $n \in \mathbb{N}$.

The Markov model is the set of states S together with all conditional probabilities $P(X_n = s_i | X_{n-1} = s_j)$, $n \in \mathbb{N}, i, j \in I$.

The states in Markov chain are directly observable. For the analogy to the quantum Markov model we can consider that in each step a value i from the index set I is produced with some probability and then the state passes onto state s_i. So the index set takes place of the set of outputted values.

Usually, only Markov models with non-evolving probabilities are considered. By non-evolving we mean

$$\forall n \in \mathbb{N}, \quad P(X_n = s_i | X_{n-1} = s_j) = P(X_{n-1} = s_i | X_{n-2} = s_j), \tag{A.1}$$

so the probability of obtaining value i does not depend on the actual position in the Markov chain, it does depend only on the previous state. In this paper we will consider only the non-evolving Markov models with finite set of states.

A.1 Probability distribution after n iterations

Suppose we want to predict in what state will be the n-th element of the chain or equivalently, which value will be produced, knowing the initial state s_{ini}. Denote

$$P(X_n = s_i | X_{n-1} = s_j) = \mathbf{T}_{ij}, \quad \mathbf{p_n} = (P(X_n = s_1), ..., P(X_n = s_m))^T, [1] \tag{A.2}$$

From the Bayes' theorem we know that

$$\begin{aligned} P(X_n = s_i) = P(X_n = s_i | X_{n-1} = s_1) P(X_{n-1} = s_1) + \cdots + \\ + P(X_n = s_i | X_{n-1} = s_m) P(X_{n-1} = s_m) = \sum_{j=1}^{m} \mathbf{T}_{ij} (\mathbf{p_{n-1}})_j \end{aligned} \tag{A.3}$$

or in the more compact form

$$\mathbf{p_n} = \mathbf{T} \mathbf{p_{n-1}}. \tag{A.4}$$

So if we know probability distribution on the set of states in the previous segment of the chain, we can easily calculate the probability distribution in the actual segment only by multiplication of the matrix $\mathbf{T}$, usually called transition matrix. For example, if the initial state is s_2, then the initial probability distribution is $\mathbf{p_0} = (0, 1, 0, ..., 0)^T$. Probability distribution on the n-th state will be then

$$\mathbf{p_n} = \mathbf{T}^n \mathbf{p_0}. \tag{A.5}$$

From the construction of the Markov model it is obvious that

$$P(X_n = s_1 | X_{n-1} = s_j) + \cdots + P(X_n = s_m | X_{n-1} = s_j) = 1 \tag{A.6}$$

(each segment of the chain must be represented by some state from S), what implies

Theorem A.1.1. *Transition matrix* $\mathbf{T}$ *satisfyies following relations:*

$$\forall i, j \in I, \quad 0 \leq \mathbf{T}_{ij} \leq 1 \ , \tag{A.7}$$

$$\forall j \in I, \quad \sum_{i \in I} \mathbf{T}_{ij} = 1 \ . \tag{A.8}$$

From the above theorem can be easily derived that for all n is $\mathbf{p_n}$ probability distribution. Furthermore, we call every matrix satisfying the above theorem transition matrix.

[1] T is a transpose. We consider vectors as columns.

A.2 Stationary distribution

Definition A.2.1. *We define stationary distribution of the Markov model as a probability distribution* $\mathbf{p}_s$ *satisfying*

$$\mathbf{p}_s = \mathbf{T}\mathbf{p}_s \ .$$

The stationary distribution is distribution on the set of states S which does not change with iterations. In other words, if one element of the Markov chain is characterized by a stationary distribution $\mathbf{p}_s$, then the next one will also be characterized by the same distribution $\mathbf{p}_s$.

We have already mentioned the distribution after n iterations $\mathbf{p_n}$. What if the limit of such distributions exist? The limit would well characterize the distribution on the states after the machine has been running for a long time. Moreover, such limit is a stationary distribution and thus does not change anymore.

Theorem A.2.1. *Suppose that the limit* $\lim_{n\to\infty} \mathbf{p_n} = \lim_{n\to\infty} \mathbf{T}^n\mathbf{p_0}$ *exist. Then the limit is a stationary distribution.*

A.3 Word probabilities

We have already mentioned word in the definition 1.1.3. The word probabilities of the classical Markov model are much simpler than in the quantum case. For example, probability of obtaining of the word 436 with the initial state s_2 is

$$P(436, s_2) = P(4, s_2)P(3, s_4)P(6, s_3) = \mathbf{T}_{42}\mathbf{T}_{34}\mathbf{T}_{63} \ . \tag{A.9}$$

It happens little complicated when some uncertainties comes on the stage. For example, if we don't know initial state exactly or some character are omitted. Denote vector $\mathbf{p_{s_i}} = (0, ...0, 1, 0, ..., 0)^T$, where only i-th element is non-zero and equal to 1. Let the $\mathbf{p_0}$ be a probability distribution describing the first segment of the chain. Then

$$P(XXX43XX2) = (\mathbf{T}^4\mathbf{p_0})_4\mathbf{T}_{34}(\mathbf{T}^3\mathbf{p_{s_3}})_2 \ . \tag{A.10}$$

Appendix B

Hidden Markov Models

The difference between Markov models and Hidden Markov models is that in Hidden Markov models we do have states and the outputs, which are not the same. In fact, we consider states always hidden (that is the word in the name) so we cannot see them directly, but we still can observe outputs, which gives us at least some information about hidden states. Predicting the probability distribution over the set of states given the outputs is the main task which the people usually do and for that there are many algorithms, but we will not describe this whole problem. We will rather explain the main overview in order to make you better understand the "quantum" version and why this quantum version generalize the classical. There are two main types of the Hidden Markov models. Older and more discussed one is the Moore Hidden Markov model, or occasionally called state-emitting. Usually, when people talk about this Hidden Markov models they consider this Moore model. But it is not the only one. The other is Mealy Hidden Markov model, or edge-emitting, which is in fact more general and include the Moore model. More information can be found in many publications, for example cite.

B.1 Moore and Mealy Hidden Markov models

The schematic example of the Moore and Mealy model is on pictures B.1 and B.2 respectively. Both types are fully characterized by set of conditional probabilities as in the non-hidden case and also possess the Markovian-forgetful property, probability of being in the next state s from the set of states S with output y from the set of outputs O depend only on the previous state.

In Moore model the output is produced with some probability dependent on the initial state and then the state passes onto next, also with the probability dependent on the previous state only. Thus the Moore model is

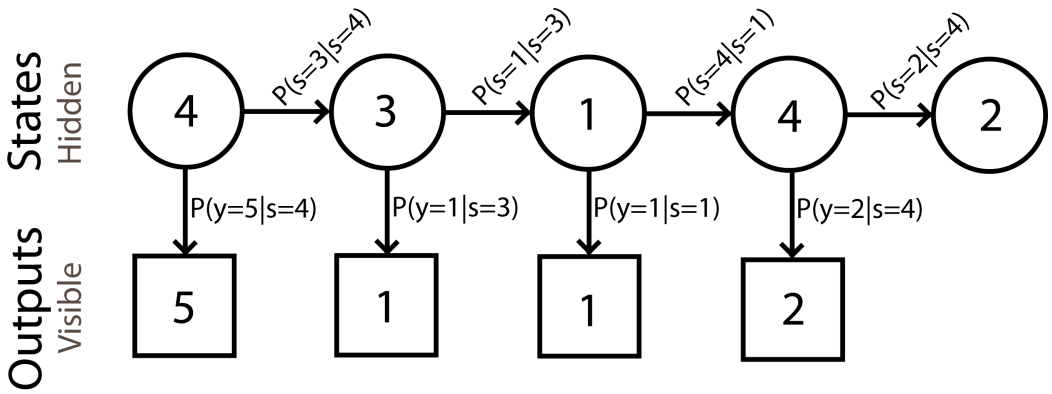


Figure B.1: Moore Hidden Markov model

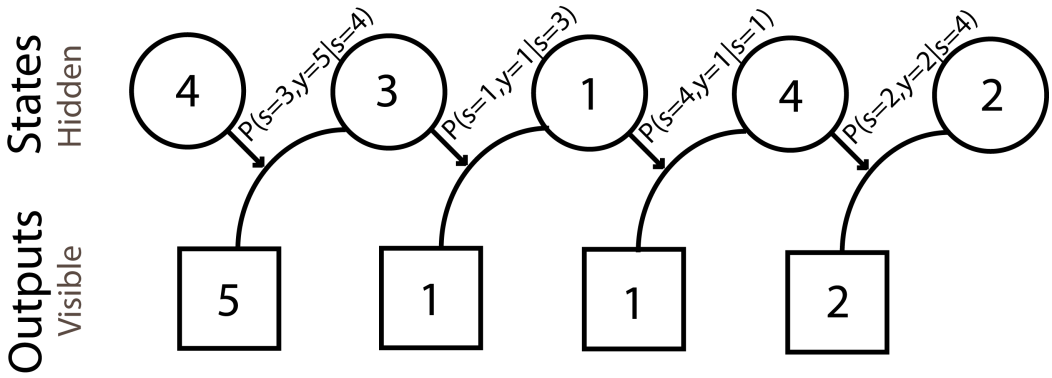


Figure B.2: Mealy Hidden Markov model

characterized by a set of conditional probabilities

$$P(Y = y|\tilde{X} = \tilde{s}) \equiv \mathbf{P}_{ys} \quad \text{and} \quad P(X = s|\tilde{X} = \tilde{s}) \equiv \mathbf{T}_{s\tilde{s}} \ . \tag{B.1}$$

In contrast, in Mealy model the output and the passing onto next state happens simultaneously with joint probability

$$P(Y = y, X = s|\tilde{X} = \tilde{s}) \equiv \mathbf{P}_{ys\tilde{s}} \tag{B.2}$$

From the pictures it is quite obvious that the Mealy model is just a generalization of the Moore. Given the Moore model we can define the Mealy model just by multiplying probabilites:

$$P(Y = y, X = s|\tilde{X} = \tilde{s}) = P(Y = y|\tilde{X} = \tilde{s}) \cdot P(X = s|\tilde{X} = \tilde{s}) \tag{B.3}$$

So the Moore model considers only cases where random values outputs and next states are independent.

Since the machine has to be in some state and some output has to be produced we gain simple conditions for the probabilities B.1 and B.2

$$\forall \tilde{s} \in S, \ \sum_{y \in O} \mathbf{P}_{y\tilde{s}} = 1, \ \sum_{s \in S} \mathbf{T}_{s\tilde{s}} = 1 \ , \tag{B.4}$$

$$\forall \tilde{s} \in S, \ \sum_{y \in O, s \in S} \mathbf{P}_{ys\tilde{s}} = 1 \ , \tag{B.5}$$

from which we can easily derive number of free real parameters which characterize the whole model. Suppose that we have Hidden Markov model with n possible states and m outputs. The Moore model is then fully characterized by $n \cdot (n + m - 2)$ and Mealy model by $n \cdot (m \cdot n - 1)$ free parameters.

One last notion for the Hidden Markov model in this section: non-hidden Markov models are subsets of the Hidden ones. That is because we can always identify each "hidden" state to one "non-hidden" outputted value and thus reveal the hidden states. Technically speaking, we simply use the Moore model and put $\mathbf{P}_{ys} = \delta_{ys}$.

B.2 Word probabilities

Since the Moore model is a trivial subset of the Mealy, we can simply derive the word probabilities for the Mealy model and Moore model will be included. Since in Hidden Markov model the states remain hidden, we do not know the initial state. We can only estimate probability with which the initial state

is the state s_0. This probability is characterized by a probability vector $\mathbf{p}$, $(\mathbf{p})_{s_0} = P(X_0 = s_0)$. We usually consider the uniform distribution

$$\mathbf{p}_0 = \left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right) , \tag{B.6}$$

but other distributions may be more suitable - for example one of the stationary probability distributions, which is more in accordance with our knowledge about the system[1] The probability of obtaining word $y_1 y_2 ... y_k$ is a simple multiple of the probabilities summed over the every possible way of obtaining this word.

$$P(y_1 y_2 ... y_k) = \sum_{s_0, s_1, ..., s_k \in S} \mathbf{P}_{y_k s_k s_{k-1}} \mathbf{P}_{y_{k-1} s_{k-1} s_{k-2}} \cdots \mathbf{P}_{y_1 s_1 s_0} \mathbf{p}_{s_0} \tag{B.7}$$

We may also ask what is the most probable distribution over the set of states given the observed outputs. In this derivation we use only the ordinary definitions of joint and conditional probabilities.

$$\begin{aligned} &P(X_0 = s_0, X_1 = s_1, ... X_k = s_k | Y_1 = y_1, Y_2 = y_2, ..., Y_k = y_k) = \\ &= \frac{P(X_0 = s_0, X_1 = s_1, ... X_k = s_k, Y_1 = y_1 | Y_2 = y_2, ..., Y_k = y_k)}{P(Y_1 = y_1 | Y_2 = y_2, ..., Y_k = y_k)} = \cdots = \\ &= \frac{P(X_0 = s_0, X_1 = s_1, ... X_k = s_k, Y_1 = y_1, Y_2 = y_2, ..., Y_k = y_k)}{P(Y_1 = y_1, Y_2 = y_2, ..., Y_k = y_k)} = \\ &= \frac{\mathbf{P}_{y_k s_k s_{k-1}} \mathbf{P}_{y_{k-1} s_{k-1} s_{k-2}} \cdots \mathbf{P}_{y_1 s_1 s_0} \mathbf{p}_{s_0}}{P(y_1 y_2 ... y_k)} , \end{aligned} \tag{B.8}$$

where $P(y_1 y_2 ... y_k)$ is the normalization constraint and thus has to be equal to sum of the numerators over s's. [2]

[1] We usually know the parameters of the model B.1 and B.2. In Mealy model we define transition matrix between states as $\mathbf{T}_{s\tilde{s}} = P(X = s | \tilde{X} = \tilde{s}) = \sum_{y \in O} P(Y = y, X = s | \tilde{X} = \tilde{s})$. The stationary states can be then calculated in the very same way as in the Markov model in section A.2.

[2] Which is an effective proof of the equation (B.7).

Appendix C

Examples of Quantum Markov models

Here we present some examples of Quantum Markov models together with the "forgetful" evolution of the density matrix $\rho_k = \mathsf{K}^k(\rho_0)$ and graphs showing the convergence of the density matrix to the stationary distribution. The initial matrix is denoted ρ_0.

1. The non-converging density matrix.

$$K_0 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$\rho_0 = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.3 \end{pmatrix}, \quad \rho_1 = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.7 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 0.7 & 0 \\ 0 & 0.3 \end{pmatrix}, \quad \rho_3 = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.7 \end{pmatrix}$$

2. The density matrix converging to a pure state

$$K_0 = \begin{pmatrix} 0.5 & 0 \\ -0.5 & -0.707 \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0.5 & 0 \\ -0.5 & 0.707 \end{pmatrix}$$

$$\rho_0 = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.3 \end{pmatrix}, \quad \rho_1 = \begin{pmatrix} 0.350 & -0.350 \\ -0.350 & 0.650 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 0.175 & -0.175 \\ -0.175 & 0.825 \end{pmatrix}$$

3. The density matrix converging to the ρ_s

$$K_0 = \begin{pmatrix} 0.966 & 0.105 \\ -0.237 & 0 \end{pmatrix}, \quad K_1 = \begin{pmatrix} -0.092 & 0.703 \\ -0.051 & 0.703 \end{pmatrix}, \quad \rho_s = \begin{pmatrix} 0.880 & -0.123 \\ -0.123 & 0.120 \end{pmatrix}$$

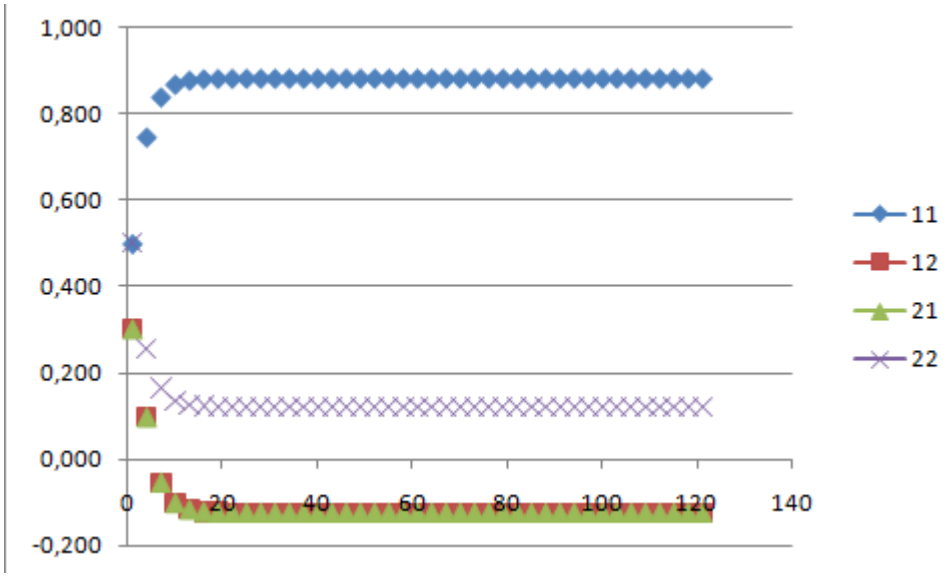


Figure C.1: Evolution of the density matrix 3.

$$\rho_0 = \begin{pmatrix} 0.500 & 0.300 \\ 0.300 & 0.500 \end{pmatrix}, \quad \rho_1 = \begin{pmatrix} 0.745 & 0.097 \\ 0.097 & 0.255 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 0.837 & -0.053 \\ -0.053 & 0.163 \end{pmatrix}$$

4. The density matrix converging to the totaly mixed state.

$$K_0 = \begin{pmatrix} 0.400 + 0.131\mathrm{i} & -0.569 + 0.699\mathrm{i} \\ 0.143 - 0.890\mathrm{i} & -0.364 - 0.211\mathrm{i} \end{pmatrix}$$

$$K_1 = \begin{pmatrix} 0.049 & 0.085 + 0.024\mathrm{i} \\ -0.085 - 0.021\mathrm{i} & 0.049 - 0.002\mathrm{i} \end{pmatrix}$$

5. The density matrix converging to the ρ_s

$$K_0 = \begin{pmatrix} 0.849 & 0.423 + 0.113\mathrm{i} \\ -0.364 - 0.364\mathrm{i} & 0.264 - 0.687\mathrm{i} \end{pmatrix}$$

$$K_1 = \begin{pmatrix} 0.031 & 0.024 + 0.225\mathrm{i} \\ 0.1030.048\mathrm{i} & -0.298 + 0.355\mathrm{i} \end{pmatrix}$$

$$\rho_s = \begin{pmatrix} 0.334 & -0.049 - 0.164\mathrm{i} \\ -0.049 + 0.164\mathrm{i} & 0.666 \end{pmatrix}$$

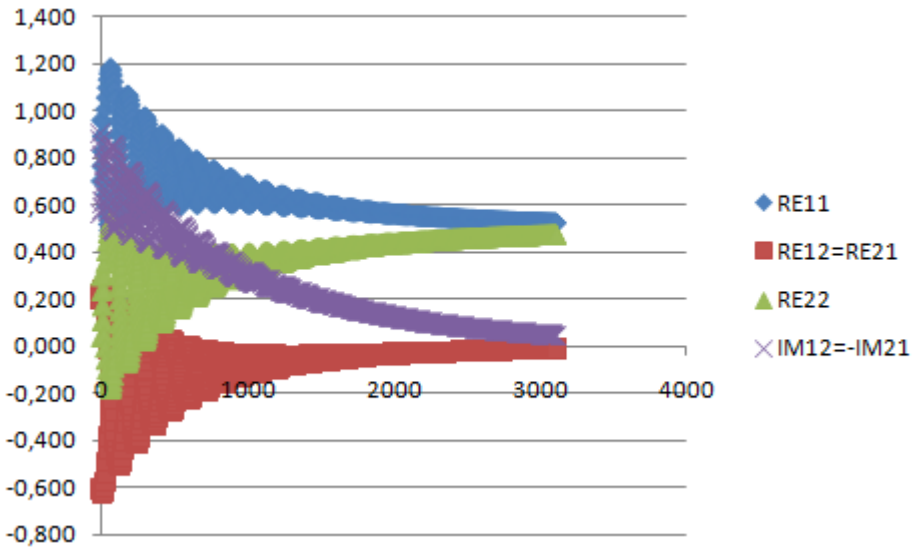


Figure C.2: Evolution of the density matrix 4.

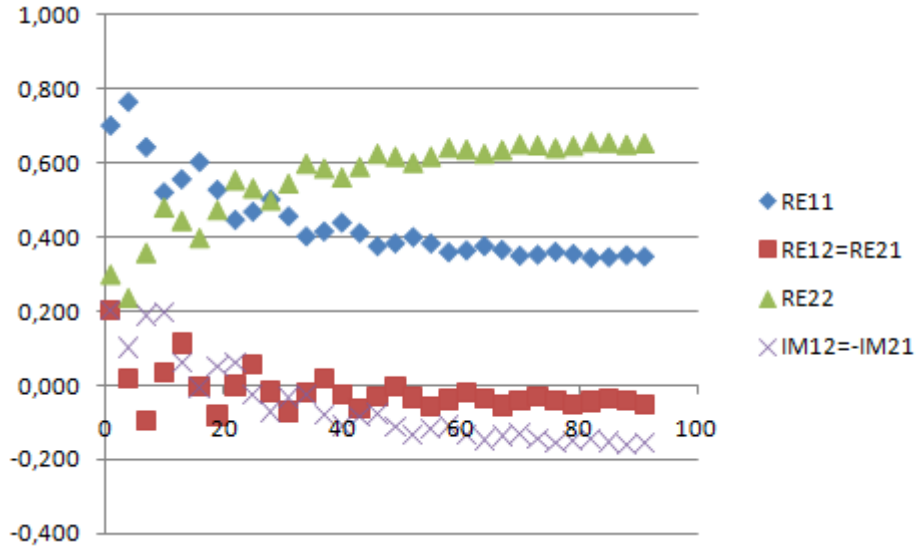


Figure C.3: Evolution of the density matrix 5.

6. Always converging density matrix, but the limit stationary density ma-

trix depends on the initial state.

$$K_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0.707 \\ 0 & 0.707 & 0 \end{pmatrix}, \quad K_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.707 & 0 \\ 0 & 0 & 0.707 \end{pmatrix}$$

Bibliography

[1] References. In K. Kraus, A. BĂśhm, J. Dollard, and W. Wootters, editors, *States, Effects, and Operations Fundamental Notions of Quantum Theory*, volume 190 of *Lecture Notes in Physics*, pages 150–152. Springer Berlin / Heidelberg, 1983.

[2] A. Beige A. Monras and K. Wiesner. Hidden quantum markov models and non-adaptive read-out of many-body states. 2010.

[3] B. V. Rajarama Bhat, G. Ramesh, and K. Sumesh. Stinespring's theorem for maps on Hilbert C*-modules. 2010.

[4] Wai-Ki Ching and Michael K. Ng. *Markov Chains: Models, Algorithms and Applications (International Series in Operations Research & Management Science)*. Springer, 1 edition, December 2005.

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

[6] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. Academic Press, 1992.

[7] P. Gmeiner. Equality conditions for internal entropies of certain classical and quantum models. Technical Report arXiv:1108.5303, Aug 2011. Comments: 19 pages, 4 figures.

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

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

[10] Sheldon M. Ross. *Introduction to Probability Models, Ninth Edition*. Academic Press, Inc., 2006.

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