acta physica slovaca vol. 62 No. 3, 275 – 346 **June 2012**

QUANTUM CHANNELS WITH MEMORY

Tomáš Rybár[1]

Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, SK-845 11 Bratislava, Slovak Republic

Received 16 November 2012, accepted 28 November 2012

Quantum memory channels represent a very general, yet simple and comprehensible model for causal processes. As such they have attracted considerable research interest, mostly aimed on their transfer capabilities and structure properties. Most notably it was shown that memory channels can be implemented via physically naturally motivated collision models. We also define the concept of repeatable channels and show that only unital channels can be implemented repeatably with pure memory channels. In the special case of qubit channels we also show that every unital qubit channel has a repeatable implementation. We also briefly explore the possibilities of stroboscopical simulation of channels and show that all random unitary channels can be stroboscopically simulated. Particularly in qubit case, all indivisible qubit channels are also random unitary, hence for qubit all indivisible channels can be stroboscopically simulated. Memory channels also naturally capture the framework of correlated experiments. We develop methods to gather and interpret data obtained in such setting and in detail examine the two qubit case. We also show that for control unitary interactions the measured data will never contradict a simple unitary evolution. Thus no memory effects can be spotted then.

DOI: 10.2478/v10155-012-0004-3

PACS: 03.65.Yz, 03.67.Lx, 03.67.Hk, 03.65.Ta

KEYWORDS: Memory Channels, Quantum Information, Estimation, Open Systems, Collision Model, Stroboscopic Simulation

Contents

[1]E-mail address: tomas.rybar@savba.sk

1 Introduction

Any physical process can be viewed as some sort of communication between sender and receiver. Such information processing protocol is always affected by noise whose source is the environment. The noise usually hinders us from acquiring certain information about the process or message. It is thus crucial to identify the noise, and its consequences. When repeating some experiment, the noise affecting each try is taken independently. This is usually justified by assuming that the environment is not significantly affected by the experiment. Such assumption allows us to approach the noise in a systematic way. However in many real world applications this assumption cannot be justified.

If the relaxation of environment is not perfect, the information from previous experiments may still roam around inside the environment. This greatly hardens the evaluation of the experiment, because repeated experiments cannot be treated independently anymore. The experiment explicitly depends on the past experiments but naturally is still independent of the upcoming experiments in future. Such scheme is called causal.

Quantum channels successfully describe independent experiments, whereas memory channels stand as model for the latter case when the experiments are affected by the history of the experiment. The important structure of causality underlies, and defines the concept of memory channels as stands proven in the paper of Werner and Kretschmann [38]. They show that every causal process is a collision model. That is a repetition of simpler processes, e.g. the single experiments, which share a common system, the environment. The environment serves then as the memory, it "remembers" the actions of previous processes. Memory channels are a more general concept for describing repeating processes and experiments as compared to quantum (memoryless) channels, and may better reflect the reality inside the laboratories. Most of the recent research was naturally aimed towards the information transfer capacity of certain classes of memory channels (see references in the introduction to Chapter 4). Our aim was to investigate the structural properties of quantum memory channels and analyze the possibilities of their estimation. In memoryless setting the information leaked into environment is irrevocably lost, whereas in memory setting, this leaked information can leak back in future experiments, thus is in principle available to the experimenter. On the other hand, the experimenter can obtain less information about the initial state of environment, since he has only one copy of it, as opposed to the memoryless case, where he has in principle infinitely many copies of the environment.

The work is organized as follows. Chapters 2 and 3 establish the language and basic notions of quantum information and estimation methods. Chapter 4 introduces the model of memory channels . In Chapter 5 structural properties of memory channels are proposed and analyzed. Chapter 6 develops estimating protocols for memory channels, and illustrates them on a simple two dimensional example.

2 Basics of quantum information

In this we will introduce the basic language of quantum information. All this can be found in standard textbooks on this subject [45, 50] and [29, 49, 6] for more in depth mathematical foundations. The notation and language will be mostly compatible with [29].

2.1 Notations and symbols

Definition of these symbols can be found in Appendix A.1.
$\mathcal{H}, \mathcal{H}_i, \mathcal{H}_{[a,b]}, \mathcal{M}$ - Hilbert space of some system,
$\mathcal{T}(\mathcal{H})$ - set of trace class operators on Hilbert space $\mathcal{H}$
$\mathcal{L}(\mathcal{H})$ - set of bounded operators on $\mathcal{H}$
T^*, T - some transformation in Schrödinger picture/ Heisenberg picture
$\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{M}$ - a C^*-algebra

2.2 States and effects

A quantum mechanical experiment (see Figure 2.1) can be divided into three parts. Preparation, evolution and measurement. In preparation stage we feed our experimental apparatus with some set of (classical) parameters $\vec{\lambda}$ in order to produce some quantum mechanical state ρ. The same set of these parameters will always produce the same quantum mechanical state of the system $\rho(\vec{\lambda})$. In the sense that whenever we push the preparation button, a quantum state will be produced, independent of previous preparations and the ensemble of all states prepared with this fixed preparation is described by $\rho(\lambda)$. The evolution stage is to large extent arbitrary and can be made a part of preparation or measurement, depending on our needs. For now we let the evolution to be trivial, until the next section. At the end we decide which measurement to use and read out the outcome $O \in \Omega$, where Ω is the set of all possible outcomes. Since QM is a statistical theory, it predicts only probabilities $p(O|\rho)$. We have to repeat such experiment many times to acquire relevant statistics.

A quantum state ρ is then associated with certain preparation procedure and an effect E_O will then attach to this state the probability of outcome O when we conduct the measurement

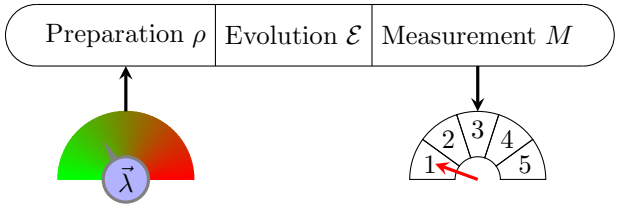

Fig. 2.1. General experimenting scheme. Experimenter sets with some knob parameters $\vec{\alpha}$ of preparation procedure which in turn specifies a quantum mechanical state ρ. The state is then optionally evolved with evolution $\mathcal{E}$ and finally measured with measurement M with specific outcome, in this case 1. The borders between preparation evolution and measurement can be laid almost arbitrarily, for the purpose of description.

$E_O(\rho) = p(O|\rho)$. The situation when a preparation procedure ρ was used and outcome O was obtained is called an *event*.

Example 2.1 (Identity and zero effect)**.** The identity effect I assigns probability 1 to every state ρ, $I(\rho) = 1$. This describes a measurement with single outcome that will be always registered. Similarly zero effect assigns zero probability to an event that never happens.

Suppose that we fix the measurement procedure but we alternate between two different preparation procedures randomly. This should also define a possibly different but valid preparation procedure. We choose the first preparation procedure associated to state ρ_1 with probability q and the second procedure associated to state ρ_2 with probability $1-q$. The resulting preparation procedure will be associated with a convex mixture of the original preparations, $\rho_3 = q\rho_1 + (1-q)\rho_2$. The probability assigned to effect E by the preparation ρ_3 has to be the convex mixture of probabilities assigned by the single preparations,

$$E(q\rho_1 + (1-q)\rho_2) = qE(\rho_1) + (1-q)E(\rho_2) \tag{2.1}$$

for any effect E and arbitrary $0 \leq q \leq 1$. To summarize:

- an effect is an affine mapping from the set of states to the interval $[0, 1]$
- the set of states is a convex set.

2.2.1 States

The set of states, in the Hilbert space formulation of QM, is described by positive trace class operators of unit trace, also called density matrices

$$\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{T}(\mathcal{H}) | \rho \geq O, \mathrm{Tr}(\rho) = 1\}, \tag{2.2}$$

we identify the set of states in quantum mechanics with $\mathcal{S}(\mathcal{H})$. This is a convex set. Every state from this set has a canonical decomposition

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \tag{2.3}$$

where $|\psi_i\rangle$ are eigenvectors and $\sum_i p_i = 1$.

A convex set has the property that whenever $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ then also their convex linear combination $\lambda\rho_1 + (1-\lambda)\rho_2$ for every $\lambda \in (0, 1)$ is from this set. Every convex set is then fully described by its extremal elements. Those are the elements for which no nontrivial decomposition into a convex combination of different elements is possible.

The extremal points of $\mathcal{S}(\mathcal{H})$ are the 1-dimensional projections $|\psi\rangle\langle\psi|$ also called the *pure states*. Each 1D projection can be identified with a normalized vector in $\mathcal{H}$, thus pure states can be also viewed as vectors on $\mathcal{H}$ with unit norm up to global phase. Every other state is then a *mixed state*. As can be seen from the canonical decomposition every state can be written as convex combination of 1-dimensional orthogonal projections. This is however not the only convex decomposition possible. In fact there are uncountably many other decompositions into convex combination of non-orthogonal projections or some non extremal elements of the state space which yield the same state.

Remark 2.2 (Classical state)**.** In classical statistical mechanics we describe the state space in a similar fashion but with an added requirement that all physically relevant states have to be diagonal.

In finite d-dimensional $\mathcal{H}$ the states of quantum system can be represented as positive, self-adjoint $d \times d$ square matrices with unit trace. Self-adjoint matrices form a vector space over real numbers. Given a basis in $\mathcal{H}$ as $\{|i\rangle\}$, there is a standard way how to construct basis in $\mathcal{L}_S(\mathcal{H}) \supset \mathcal{S}(\mathcal{H})$ called a *traceless operator basis*.

Example 2.3 (Self-adjoint operator basis)**.** Operators τ_{mn} form a basis in $\mathcal{L}_S(\mathcal{H})$, where

$$\tau_{mm} = \frac{1}{\sqrt{m(m+1)}}\Big(\sum_{k=0}^{m-1} |k\rangle\langle k| - m|m\rangle\langle m|\Big), \quad m \geq 1 \tag{2.4}$$

and for $m = 0$

$$\tau_{00} = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} |k\rangle\langle k|, \tag{2.5}$$

and for $m < n$

$$\tau_{mn} = \frac{1}{\sqrt{2}}(|m\rangle\langle n| + |n\rangle\langle m|) \tag{2.6}$$

$$\tau_{nm} = \frac{1}{\sqrt{2}} i(-|m\rangle\langle n| + |n\rangle\langle m|). \tag{2.7}$$

One can check that $\tau_{ab} = \tau_{ab}^\dagger$, $\mathrm{Tr}(\tau_{ab}) = 0\,\forall (a,b) \neq (0,0)$ and that $\langle \tau_{ab}|\tau_{cd}\rangle_{\mathrm{HS}} = \mathrm{Tr}(\tau_{ab}\tau_{cd}) = \delta_{ac}\delta_{bd}$. Every state ρ in $\mathcal{S}(\mathcal{H})$ can then be written as

$$\rho = \frac{1}{\sqrt{d}}\Big(\tau_0 + \sum_{a=1}^{d^2-1} r_a \tau_a\Big) \tag{2.8}$$

where $\tau_{00} =: \tau_0$ and τ_a are τ_{ij} in some fixed order.

Example 2.4 (Qubit)**.** The simplest nontrivial Hilbert space is of dimension 2. Let the basis of this space be denoted by normalized orthogonal vectors $|0\rangle, |1\rangle$. The traceless operator basis then is

$$\tau_{00} = \frac{1}{\sqrt{2}}\mathbb{I} \quad \tau_{01} = \frac{1}{\sqrt{2}}X \quad \tau_{10} = \frac{1}{\sqrt{2}}Y \quad \tau_{11} = \frac{1}{\sqrt{2}}Z \tag{2.9}$$

where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.10}$$

are the Pauli matrices. A state ρ is then expressed as $\frac{1}{\sqrt{2}}(\tau_{00} + r_{01}\tau_{01} + r_{10}\tau_{10} + r_{11}\tau_{11})$ or more common as $\rho = \frac{1}{2}(\mathbb{I} + \vec{r}\vec{\sigma})$, where $\vec{\sigma} = (X, Y, Z)$ is the vector of Pauli matrices. In order to $\rho \geq O$ we find out that $\| \vec{r} \| \leq 1$. The qubit state is defined by 3 parameters r_i such that

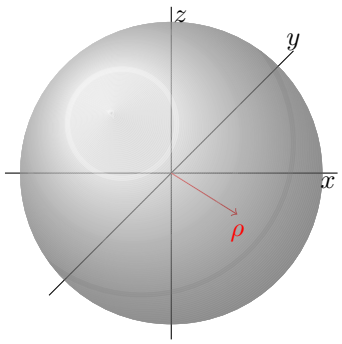

Fig. 2.2. States of qubit form a Bloch ball. Each point of this ball is a valid physical state from a 2-dimensional Hilbert space.

$\sum_i r_i r_i \leq 1$. The set of all qubit states thus forms a ball better known as Bloch ball where each state is specified by a vector lying within this unit ball called a Bloch vector, see Figure 2.2. Pure states are the states for which $\| \vec{r} \| = 1$ and lie on the surface of the ball. Two pure states are orthogonal if the Hilbert-Schmidt product of their density matrices is zero. Let $\rho_1 = \frac{1}{2}(\mathbb{I} + \vec{r}\vec{\sigma})$ and $\rho_2 = \frac{1}{2}(\mathbb{I} + \vec{s}\vec{\sigma})$.

$$\langle \rho_1 | \rho_2 \rangle_{\mathrm{HS}} = \mathrm{Tr}(\rho_1 \rho_2) = \frac{1}{2}(1 + \sum_{ij} r_i s_j \mathrm{Tr}(\sigma_i \sigma_j)) = \frac{1}{2}(1 + \vec{r}\vec{s}), \tag{2.11}$$

therefore two pure states are orthogonal if their Bloch vectors are antipodal $\vec{r} = -\vec{s}$.

Definition 2.5 (von Neumann entropy)**.** The *von Neumann entropy* $S(\rho)$ of a state $\rho \in \mathcal{S}(\mathcal{H})$ is given by following

$$S(\rho) := -\mathrm{Tr}(\rho \log \rho) = -\sum_k p_k \log p_k, \tag{2.12}$$

where p_k are the nonzero eigenvalues of ρ.

It is the quantum analog of the classical Shannon entropy. The von Neumann entropy has following important properties:

- S is a concave function on the set of states, i.e. $S(q\rho_1 + (1-q)\rho_2) \geq qS(\rho_1) + (1-q)S(\rho_2)$;
- S is invariant under unitary conjugation $S(U\rho U^\dagger) = S(\rho)$;
- $S(\rho) = 0$ iff ρ is a pure state.

A unique state ρ_M which maximizes the von Neumann entropy is called a *maximally mixed state*. Such state is proportional to identity $\rho_M = \frac{1}{d}\mathbb{I}$ and the von Neumann entropy of such state is

$$S(\rho_M) = -\sum_{k=0}^{d-1} \frac{1}{d} \log \frac{1}{d} = \log d, \tag{2.13}$$

where d is the dimension of underlying Hilbert space.

Pure states have an additional structure called the superposition. Having two pure states $|\psi\rangle$ and $|\phi\rangle$ we can form another pure state $|\theta\rangle = \alpha|\psi\rangle + \beta|\phi\rangle$ which is called a superposition of $|\psi\rangle$ and $|\phi\rangle$. Such combination of pure states is not allowed in classical mechanics and gives rise to many quantum phenomena.

2.3 Effects

When we defined the state space in previous subsection we also made a step toward defining the effects. From the beginning of this section we know that an effect is a linear mapping from $\mathcal{S}(\mathcal{H})$ to the interval $[0,1]$. Thus an effect is a linear functional on $\mathcal{S}(\mathcal{H})$ which by linearity can be spread over the whole $\mathcal{T}(\mathcal{H})$. The dual space of $\mathcal{T}(\mathcal{H})$ is $\mathcal{L}(\mathcal{H})$ and that is the native space of effects. So every effect can be associated with an operator $E \in \mathcal{L}(\mathcal{H})$ such that

$$E(\rho) = \mathrm{Tr}(E\rho), \tag{2.14}$$

where $\mathbb{I} \geq E \geq O$ in order to obtain a valid probability from interval $[0,1]$. Let denote the set of effects by $\mathcal{E}(\mathcal{H})$.

Example 2.6 (Qubit effects)**.** Since effects are positive operators we can expand them in the traceless operator basis, thus every qubit effect can be written as

$$E = (\mathbb{I}\alpha + \vec{a}\vec{\sigma}). \tag{2.15}$$

Eigenvalues of this operator are $\lambda_{\pm} = \alpha \pm \parallel a \parallel$ therefore E is an effect if and only if $\lambda_{-} \geq 0$ and $\lambda_{+} \leq 1$. Notice that this also implies $\parallel a \parallel \leq 1$.

2.3.1 Observables

Each measurement device can have n possible different outcomes. Each outcome is represented by an effect E_j such that the probability of outcome j when $\rho \in \mathcal{S}(\mathcal{H})$ was prepared is $\mathrm{Tr}(E_j\rho)$. Formally let Ω be a nonempty set of outcomes. A *σ-algebra* on Ω is a collection $\mathcal{F}$ of subsets of Ω such that

- $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$
- if $X \in \mathcal{F}$ then $\Omega \setminus X \in \mathcal{F}$
- if $X_1, X_2, \ldots \in \mathcal{F}$ then $\bigcup_i X_i \in \mathcal{F}$.

$\mathcal{F}$ represents all possible questions "Was the observed outcome in this subset of possible outcomes? ".

Definition 2.7 (POVM)**.** A *positive operator valued measure* is a mapping $A : \mathcal{F} \mapsto \mathcal{E}(\mathcal{H})$ such that

- $A(\emptyset) = O$
- $A(\Omega) = \mathbb{I}$

- $A(\bigcup_i X_i) = \sum_i A(X_i)$ for any set X_i of disjoint sets in $\mathcal{F}$.

A POVM is thus a prescription which assigns to every possible set of outcomes appropriate effect. However for all practical purposes in this work it is enough to think of POVM as a direct outcome - effect assignment. In this sense Ω is just the set of possible outcomes. Hence since $A(\Omega) = \mathbb{I}$ also $\sum_{i=1}^{n} E_i = \mathbb{I}$ for a measurement where every outcome is registered. An observable is then simply a POVM. In some courses of quantum mechanics observables are commonly defined as self-adjoint operators. Those are then the POVMs with all effects being orthogonal projections. They are sometimes referred to as projection valued measures, PVMs or sharp measurements. To every PVM A we can uniquely assign a self-adjoin operator. Let Ω be the set of all outcomes and let $X_i \in \mathcal{F}$ correspond to individual outcomes. Then

$$O_A := \sum_i a_i A(X_i), \tag{2.16}$$

where a_i are the values assigned to each outcome.

2.3.2 Bipartite systems

So far we have considered only one system in Hilbert space with arbitrary dimension. Let us have two independent quantum systems A and B in separate Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$. The Hilbert space of composite system A+B is then obtained by tensor product $\mathcal{A} \otimes \mathcal{B}$. If the state of system A is $\rho_{\mathcal{A}}$ and of the system B, $\rho_{\mathcal{B}}$, the joint system is then again obtained by tensor product $\omega_{\mathcal{A}B} = \rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}$. We call such states factorized. Similarly for effects. If we have an effect $E_{\mathcal{A}}$ and $E_{\mathcal{B}}$ assigned to outcomes O_A, O_B the composite effect on $\mathcal{A}\otimes\mathcal{B}$ is $E_{\mathcal{A}}\otimes E_{\mathcal{B}}$. The probability assigned to $E_{\mathcal{A}} \otimes E_{\mathcal{B}}$ should respect the independent nature of the events $p(O_A, O_B|\rho_{\mathcal{A}}, \rho_{\mathcal{B}}) = p(O_A|\rho_{\mathcal{A}})p(O_B|\rho_{\mathcal{B}})$ and this is true:

$$\begin{aligned} p(O_A, O_B|\rho_{\mathcal{A}}, \rho_{\mathcal{B}}) &= \mathrm{Tr}[(E_{\mathcal{A}} \otimes E_{\mathcal{B}})(\rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}})] = \mathrm{Tr}(E_{\mathcal{A}}\rho_{\mathcal{A}})\mathrm{Tr}(E_{\mathcal{B}}\rho_{\mathcal{B}}) \\ &= p(O_A|\rho_{\mathcal{A}})p(O_B|\rho_{\mathcal{B}}). \end{aligned} \tag{2.17}$$

However with the structure of the tensor product we have also introduced a state space which is by far larger than a state space of factorized density matrices. Not every density matrix in $\mathcal{A} \otimes \mathcal{B}$ is of the aforementioned form. States which cannot by written in this form describe correlated systems. If the state can be written as convex combination of factorized states, then it is called *separable*. Finally states which are not separable or factorized are called *entangled* and describe systems with quantum correlations Pure states can be only entangled or factorized.

We have thus a way how to describe joint systems, however we have to find a way how to describe single systems which are part of a larger system. If a state is not factorized the legitimate question is then what are the *local* states of respective subsystems?

Given that we have some $\omega_{\mathcal{A}B} \in \mathcal{S}(\mathcal{A} \otimes \mathcal{B})$ which is not factorized we denote $\omega_{\mathcal{A}}$ and $\omega_{\mathcal{B}}$ the local states of respective subsystems. We require that any measurement outcome probability obtained only on one of the subsystems has to be recovered by the density matrix of the subsystem:

$$\mathrm{Tr}[\omega_{\mathcal{A}B}(E_{\mathcal{A}} \otimes \mathbb{I})] = \mathrm{Tr}(\omega_{\mathcal{A}}E_{\mathcal{A}}) \quad \forall E_{\mathcal{A}}. \tag{2.18}$$

The state $\omega_{\mathcal{A}}$ is called a partial trace of $\omega_{\mathcal{A}B}$ over system B and is denoted by

$$\omega_A = \mathrm{Tr}_{\mathcal{B}}\omega_{AB} = \sum_i \langle i|_{\mathcal{B}}\omega_{\mathcal{A}B}|i\rangle_{\mathcal{B}} \tag{2.19}$$

with some basis in $\mathcal{B}$, $\{|i\rangle_{\mathcal{B}}\}$.

Let the basis of $\mathcal{A}\otimes\mathcal{B}$ be $\{|i\rangle_{\mathcal{A}}\otimes|j\rangle_{\mathcal{B}}\}$ where $\{|i\rangle_{\mathcal{A}/\mathcal{B}}\}$, are bases of respective subspaces and let the state $\omega_{\mathcal{A}B}$ have an expansion in this basis

$$\omega_{\mathcal{A}B} = \sum_{klmn} \omega_{kl,mn}|k\rangle_{\mathcal{A}}\langle l|\otimes|m\rangle_{\mathcal{B}}\langle n|, \tag{2.20}$$

then the density matrix $\omega_{\mathcal{A}}$ is

$$\begin{aligned}\sum_i \langle i|_{\mathcal{B}}\omega_{\mathcal{A}B}|i\rangle_{\mathcal{B}} &= \sum_{iklmn} \omega_{kl,mn}|k\rangle_{\mathcal{A}}\langle l|\otimes\langle i|m\rangle_{\mathcal{B}}\langle n|i\rangle \\ &= \sum_{ikl}\omega_{kl,ii}|k\rangle_{\mathcal{A}}\langle l|. \end{aligned}\tag{2.21}$$

Example 2.8 (Schmidt form)**.** Let us have a pure state $|\psi\rangle$ in $\mathcal{S}(\mathcal{H}\otimes\mathcal{H})$ with a basis in (both) $\mathcal{H}$: $\{|i\rangle\}$, such that $|\psi\rangle = \sum_j \sqrt{p}_j|jj\rangle$. Let us denote the partial trace over the first and second subsystem Tr_1 and Tr_2 respectively. The density matrix of $|\psi\rangle$ is

$$|\psi\rangle\langle\psi| = \sum_{ij}\sqrt{p_i p_j}|ii\rangle\langle jj|, \tag{2.22}$$

and the local density matrix of both subsystems is

$$\mathrm{Tr}_1(|\psi\rangle\langle\psi|) = \sum_i p_i|i\rangle\langle i| = \mathrm{Tr}_2(|\psi\rangle\langle\psi|). \tag{2.23}$$

Every pure state in $\mathcal{A}\otimes\mathcal{B}$ (where $\mathcal{A}$ doesn't have to be equal to $\mathcal{B}$) can be written for a suitably chosen basis in $\mathcal{A}$ and $\mathcal{B}$ $\{|i\rangle_{\mathcal{A}/\mathcal{B}}\}$ as (2.22) and the density matrices of subsystems are diagonal in these bases.

The entropy of a pure state is zero. However if this state is bipartite, the entropy of its separate parts needs not be zero if the composite state is not factorized. Thus the information inside the local states is not allways enough to describe the whole information of the bipartite state. Because in the description of individual local states we cannot include the bipartite correlations which might be present. Mathematically we say that the von Neumann entropy is *subadditive*:

$$S(\omega_{\mathcal{A}B}) \leq S(\rho_{\mathcal{A}}) + S(\rho_{\mathcal{B}}), \tag{2.24}$$

end the equality arises only when systems A and B are not correlated.

2.4 Evolution

In quantum mechanics we have learned that Schrödinger equation governs the time evolution of a quantum state

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle, \tag{2.25}$$

where H is a self-adjoint operator, *hamiltonian*. If this hamiltonian is time independent we can give a formal solution of (2.25)

$$|\psi(t)\rangle = e^{iHt/\hbar}|\psi\rangle =: U(t)|\psi\rangle, \tag{2.26}$$

where $U(t)$ is some unitary operator for any time t. For a density matrix $\rho \in \mathcal{S}(\mathcal{H})$ this yields unitary conjugation $\rho_t = U(t)\rho U^\dagger(t)$, where $U(0) = \mathbb{I}$ and $\rho_0 = \rho$. Evolution of states forward in time is called the Schrödinger picture of quantum mechanics. Alternatively we can look on the evolution of observables: $B_0 = U^\dagger(t)B_tU(t)$ what is called the Heisenberg picture. A map in Schrödinger picture T^* (with an asteriks) such that $T^*(\rho_0) = \rho_t$ will have a Heisenberg picture equivalent T (without asteriks) connected via identity

$$\mathrm{Tr}[\rho_t B_t] = \mathrm{Tr}[T^*(\rho_0)B_t] =: \mathrm{Tr}[\rho_0 T(B_t)] = \mathrm{Tr}[\rho_0 B_0], \tag{2.27}$$

for any bounded observable B_t and any state ρ_0. As can be seen, the evolution in Heisenberg picture goes against the flow of time in Schrödinger picture because the map T transforms observable B from the *output* space of T^* to its *input* space. Both pictures are equivalent, however some times it is more insightful to use Heisenberg picture, mostly when infinite dimensional systems are considered and sometimes the Schrödinger picture is more intuitive. Another asymmetry which favors the Heisenberg picture is that the states in big Hilbert spaces are hard to describe in terms of localization. This is in general also true for observables, however physically for us it usually makes sense to think about measurements and measurement apparatuses which are precisely localized, and physically feasible to construct. It seems that nonlocal states are far more easily produced than nonlocal measurements. In this work we will use both pictures at our advantage.

If we have our state space defined as $\mathcal{S}(\mathcal{H})$ then possibly every map $T^* : \mathcal{S}(\mathcal{H}) \mapsto \mathcal{S}(\mathcal{H})$ could be a valid evolution. Such evolution should have thus following properties

- linearity; $T^*(\rho_1 + \lambda\rho_2) = T^*(\rho_1) + \lambda T^*(\rho_2)$
- trace preserving; $\mathrm{Tr}[T^*(\rho)] = \mathrm{Tr}\rho$
- positive; $T^*(\rho) \geq O$.

This turns out to be not enough. Let's say we have a map T^* which is positive. This guarantees that $T^*(\rho) \geq O$ for every positive $\rho \in \mathcal{H}$. This does not guarantee us that $I_\mathcal{K} \otimes T^*(\omega) \geq O$ for every $\omega \in \mathcal{K} \otimes \mathcal{H}$ where $\mathcal{K}$ is a Hilbert space with arbitrary dimension and $I_\mathcal{K}$ is a trivial (identical) evolution in this space.

Maps that satisfy this condition $I_\mathcal{K} \otimes T^*(\omega)$ for arbitrary $\mathcal{K}$ are called *completely positive*. Due to linearity the action can be shifted from the state space $\mathcal{S}(\mathcal{H})$ to the trace class operators $\mathcal{T}(\mathcal{H})$.

Definition 2.9 (Channel)**.** A linear trace preserving mapping in Schrödinger picture $T^* : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ which is at the same time completely positive is called a channel.

In Heisenberg picture the trace preserving condition translates to unitality on observables. We call a mapping T *unital* if it preserves the identity operator: $T(\mathbb{I}) = \mathbb{I}$.

Definition 2.10 (Channel)**.** A linear unital mapping in Heisenberg picture $T : \mathcal{L}(\mathcal{B}) \mapsto \mathcal{L}(\mathcal{A})$ which is at the same time completely positive is called a channel.

In previous lines we have adopted convention that states entering some evolution (inputs) live in Hilbert space $\mathcal{A}$ and states which are produced after the evolution (outputs) live in Hilbert space $\mathcal{B}$. The Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$ can be different, but unless explicitly stated we will consider them as isomorphic. However we will still maintain the distinction in naming the input Hilbert space as $\mathcal{A}$ and the output space $\mathcal{B}$ for better readability. Note that Heisenberg picture describes the evolution *against* the flow of time from observables on outputs in $\mathcal{B}$ to observables on inputs in $\mathcal{A}$.

To show that not every positive map is also completely positive see the next example.

Example 2.11 (Positive but not completely positive)**.** Let us have a linear qubit map $E : \mathcal{S}(\mathcal{H}) \mapsto \mathcal{S}(\mathcal{H})$ defined on (not normed) traceless operator basis:

$$E(\mathbb{I}) = \mathbb{I} \quad E(X) = X \quad E(Y) = Y \quad E(Z) = O. \tag{2.28}$$

This map projects the states in Bloch ball onto its xy-plane and thus is positive. Lets take a bipartite pure state $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ whose density matrix is

$$|\psi\rangle\langle\psi| = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{1}{4}(\mathbb{I} + X \otimes X + Y \otimes Y - Z \otimes Z). \tag{2.29}$$

Then

$$I \otimes E(|\psi\rangle\langle\psi|) = \frac{1}{4}(\mathbb{I} + X \otimes X + Y \otimes Y) = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}. \tag{2.30}$$

However this is not positive because the operator on the right hand side of (2.30) has negative eigenvalues.

We do not know whether all channels are actually physical. Schrödinger equation gives us only unitary evolution on single system, however the structure of state space does not put such hard constraints on evolution. Stinespring's key result [64] tells us that every completely positive linear map T^* can be thought of as a unitary evolution on a larger Hilbert space. Hence channels will become the natural language for the evolution of open quantum systems.

Theorem 2.12 (Stinespring's dilation theorem)**.** *Let $\mathfrak{B}$ be a unital C^*-algebra and let $T : \mathfrak{B} \mapsto \mathcal{L}(\mathcal{A})$ be a completely positive unital map. Then there exists a Hilbert space $\mathcal{K}$, a unital $\dagger$-homomorphism[2] $\pi : \mathfrak{B} \mapsto \mathcal{L}(\mathcal{K})$ and an isometry $V : \mathcal{A} \mapsto \mathcal{K}$ such that*

$$T(b) = V^\dagger \pi(b) V. \tag{2.31}$$

for all $b \in \mathfrak{B}$. The triple $(\pi, V, \mathcal{K})$ is called Stinespring representation of channel T. If the closed linear span of $\pi(\mathfrak{B})V\mathcal{A}$ equals $\mathcal{K}$ then such representation is called minimal.

[2] $\pi(b^\dagger) = \pi(b)^\dagger$

The Stinespring's dilation theorem is very powerful and does not even require separability of Hilbert space.

For finite d-dimensional $\mathcal{B}$ where we identify $\mathcal{L}(\mathcal{B})$ with $\mathfrak{B}$ we get that $\mathcal{K} = \mathcal{M} \otimes \mathcal{B}$, for some finite dimensional Hilbert space $\mathcal{M}$, isometry V can be written as $V = U|0\rangle$ where U is unitary[3], $|0\rangle$ is some pure state in $\mathcal{M}$ and the mapping $\pi(b)$ will become $\pi(b) = \mathbb{I}_{\mathcal{M}} \otimes b$ so that

$$\begin{aligned} \mathrm{Tr}[\rho T(b)] &= \mathrm{Tr}[(|0\rangle\langle 0| \otimes \rho)U^\dagger(\mathbb{I}_{\mathcal{M}} \otimes b)U] \\ &= \mathrm{Tr}[U(|0\rangle\langle 0| \otimes \rho)U^\dagger(\mathbb{I}_{\mathcal{M}} \otimes b)] = \mathrm{Tr}[T^*(\rho)b]. \end{aligned} \tag{2.32}$$

In Schrödinger picture this has a nice interpretation. Any channel $T^* : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ can be realized as a composition of three maps:

1. take input state ρ and attach to it a pure state of some environment $\mathcal{M}$: $|0\rangle$, to get $|0\rangle\langle 0| \otimes \rho$;
2. let the composite system evolve with unitary interaction U, $U(|0\rangle\langle 0| \otimes \rho)U^\dagger$;
3. trace out the environmental degrees of freedom to obtain the output, $T^*(\rho) = \mathrm{Tr}_{\mathcal{M}}[U(|0\rangle\langle 0| \otimes \rho)U^\dagger]$.

The last step can can be seen from

$$\begin{aligned} &\mathrm{Tr}[T^*(\rho)b] = \mathrm{Tr}[U(|0\rangle\langle 0| \otimes \rho)U^\dagger(\mathbb{I}_{\mathcal{M}} \otimes b)] \\ &= \sum_{i\mu} \langle \mu i|U(|0\rangle\langle 0| \otimes \rho)U^\dagger(\mathbb{I}_{\mathcal{M}} \otimes b)|\mu i\rangle \\ &= \mathrm{Tr}[\mathrm{Tr}_{\mathcal{M}}\big(U(|0\rangle\langle 0| \otimes \rho)U^\dagger\big)b], \end{aligned} \tag{2.33}$$

for some basis $\{|\mu\rangle\}$ in $\mathcal{M}$ and $\{|i\rangle\}$ in $\mathcal{B}$, hence

$$T^*(\rho) = \mathrm{Tr}_{\mathcal{M}}[U(|0\rangle\langle 0| \otimes \rho)U^\dagger]. \tag{2.34}$$

Remark 2.13 (Kraus representation)**.** The unitary $U : \mathcal{M} \otimes \mathcal{A} \mapsto \mathcal{M} \otimes \mathcal{B}$ can be written as

$$U = |i\rangle\langle j| \otimes A_{ij}, \tag{2.35}$$

where A_{ij} are some linear operators $A_{ij} : \mathcal{A} \mapsto \mathcal{B}$. Then

$$\begin{aligned} &\mathrm{Tr}_{\mathcal{M}}[U(|0\rangle\langle 0| \otimes \rho)U^\dagger] = \sum_i \langle i|(U(|0\rangle\langle 0| \otimes \rho)U^\dagger)|i\rangle \\ &= \sum_{iklmn} \langle i|k\rangle\langle l|0\rangle\langle 0|m\rangle\langle n|i\rangle A_{kl}\rho A_{nm}^\dagger \\ &= \sum_i A_{i0}\rho A_{i0}^\dagger = T^*(\rho). \end{aligned} \tag{2.36}$$

This is the so called *Kraus representation* of a channel $T^* : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ with *Kraus operators* $A_k := A_{k0}$. This way we can completely remove the explicit environment from the description

[3] In case $\mathcal{A}$ and $\mathcal{B}$ were not isomorphic U would remain an isometry.

of channel and instead use only the set of Kraus operators $\{A_k\}$ which has to satisfy the trace preserving condition of channel

$$\mathrm{Tr}[\rho] = \mathrm{Tr}[T^*(\rho)] = \mathrm{Tr}[\sum_k A_k \rho A_k^\dagger] \Rightarrow \sum_k A_k^\dagger A_k = \mathbb{I}_{\mathcal{A}} \quad \forall \rho \in \mathcal{S}(\mathcal{A}). \tag{2.37}$$

In Heisenberg picture we get $T : \mathcal{L}(\mathcal{B}) \mapsto \mathcal{L}(\mathcal{A})$ through the duality relation $\sum_k \mathrm{Tr}[A_k \rho A_k^\dagger b] = \sum_k \mathrm{Tr}[\rho A_k^\dagger b A_k]$:

$$T(b) = \sum_k A_k^\dagger b A_k, \tag{2.38}$$

where we see that the trace preserving condition transforms to unitality

$$\sum_k A_k^\dagger \mathbb{I}_{\mathcal{B}} A_k = \mathbb{I}_{\mathcal{A}}. \tag{2.39}$$

The Kraus operators are not unique. As can be seen from the Stinespring's representation the unitary U can be replaced by $(w \otimes I)U$ with an arbitrary unitary $w : \mathcal{M} \mapsto \mathcal{M}$,

$$\begin{aligned} &\mathrm{Tr}[(w \otimes I)U(|0\rangle\langle 0| \otimes \rho)U^\dagger(w^\dagger \otimes I)(\mathbb{I}_{\mathcal{M}} \otimes b)] \\ &= \mathrm{Tr}[(w^\dagger \otimes I)(w \otimes I)U(|0\rangle\langle 0| \otimes \rho)U^\dagger(\mathbb{I}_{\mathcal{M}} \otimes b)] \\ &= \mathrm{Tr}[U(|0\rangle\langle 0| \otimes \rho)U^\dagger(\mathbb{I}_{\mathcal{M}} \otimes b)]. \end{aligned} \tag{2.40}$$

This freedom translates to freedom on Kraus operators:

$$\hat{A}_k = \sum_i w_{ki} A_i. \tag{2.41}$$

The new Kraus operators $\hat{A}_k$ form the same channel as operators A_k. It was shown by Kraus [37, 36] that any channel $T^* : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ can have a Kraus representation with number of Kraus operators n such that $n \leq d^2$ where $d = \max\big(\dim(\mathcal{A}), \dim(\mathcal{B})\big)$.

Remark 2.14 (Uniqueness of minimal Stinespring dilation)**.** Steinspring's representation is not unique. However it can be shown that minimal Stinespring's representation is unique up to isometry in the following sense. Let us have two different Stinespring's representations of the same channel T, $(\pi, V, \mathcal{K})$ and $(\pi_1, V_1, \mathcal{K}_1)$,

$$T(b) = V^\dagger \pi(b) V = V_1^\dagger \pi_1(b) V_1. \tag{2.42}$$

Assume that the first representation is minimal. Then we can conclude that $\dim \mathcal{K} \leq \dim \mathcal{K}_1$ and there exists a well defined isometry $W : \mathcal{K} \mapsto \mathcal{K}_1$ with prescription

$$W(\pi(b)V\psi) := \pi_1(b)V_1\psi, \tag{2.43}$$

for all $b \in \mathfrak{B}$ and $\psi \in \mathcal{A}$. By setting $b = \mathbb{I}$ we get

$$WV = V_1, \tag{2.44}$$

and the intertwining relation

$$W\pi = \pi_1 W \tag{2.45}$$

as a bonus.

Definition 2.15 (Choi-Jamiolkowski operator)**.** Let us have a channel T^* on d-dimensional Hilbert space $\mathcal{H}$ with some basis $\{|i\rangle\}$. We call the operator $\chi(T^*) = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes T^*(|i\rangle\langle j|)$ the *Choi-Jamiolkowski operator* of the channel T^*.

Note that

$$\mathrm{Tr}_2[\chi(T^*)] = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \mathrm{Tr}[T^*(|i\rangle\langle j|)], \tag{2.46}$$

due to the trace preserving quality of T^* we get

$$\begin{aligned} &\sum_{i,j=0}^{d-1} |i\rangle\langle j| \mathrm{Tr}[T^*(|i\rangle\langle j|)] = \sum_{i,j=0}^{d-1} |i\rangle\langle j|\delta_{ij} \\ &= \sum_{i=0}^{d-1} |i\rangle\langle i| = \mathbb{I}, \end{aligned} \tag{2.47}$$

where by Tr_2 we denote the partial trace over the second subsystem. Since T^* is completely positive, the Choi-Jamiolkowski operator is positive[4]. The virtue of Choi-Jamiolkowski operator is that it is unique, what makes it a good representation for optimization in the space of quantum channels. Moreover, the relation is isomorphic, meaning that every positive operator χ in $\mathcal{L}(\mathcal{H} \otimes \mathcal{H})$ whose partial trace over one subsystem is $\mathrm{Tr}_2\chi = \mathbb{I}$ is a Choi-Jamiolkowski operator of some channel.

A channel is a linear map acting on a real vector space of self-adjoint operators. Lets denote by $\{\tau_j\}$ the traceless operator basis. We can view a channel as an affine mapping on the generalized Bloch vector $\vec{r}$ from equation (2.8),

$$\vec{r} \mapsto \vec{a} + \hat{A}\vec{r}, \tag{2.48}$$

with some vector $\vec{a}$ and operator $\hat{A}$. It can be further encapsulated in one operator

$$\begin{pmatrix} 1 \\ \vec{r} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ \vec{a} & \hat{A} \end{pmatrix} \begin{pmatrix} 1 \\ \vec{r} \end{pmatrix}. \tag{2.49}$$

Let

$$A = \begin{pmatrix} 1 & 0 \\ \vec{a} & \hat{A} \end{pmatrix}, \tag{2.50}$$

then

$$A_{ij} = \mathrm{Tr}[\tau_i T^*(\tau_j)]. \tag{2.51}$$

For unitary channels A is from special orthonormal group, $A^T A = \mathbb{I}, \det(A) = 1$ and $\vec{a} = 0$ because of unitality.

[4]The operator $\sum_{ij} |i\rangle\langle j| \otimes |i\rangle\langle j|$ is positive, and remains positive whenever completely positive mapping acts on its subparts.

Example 2.16 (Contraction to a line)**.** Let us have a qubit channel $E^* : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ defined on the operator basis

$$E^*(\mathbb{I}) = \mathbb{I} \quad E^*(X) = X \quad E^*(Y) = O \quad E^*(Z) = O. \tag{2.52}$$

The vector representation is by far the easiest to acquire, let $\tau = 1/\sqrt{2}(\mathbb{I}, X, Y, Z)$, then

$$A_{ij} = \mathrm{Tr}[\tau_i E^*(\tau_j)] \Rightarrow$$
$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.53}$$

The Choi-Jamiolkowski operator is also quite straightforward. Knowing that

$$\sum_{i,j=0}^{1} |i\rangle\langle j| \otimes |i\rangle\langle j| = \frac{1}{2}(\mathbb{I} \otimes \mathbb{I} + X \otimes X - Y \otimes Y + Z \otimes Z), \tag{2.54}$$

we get

$$\chi(E^*) = \frac{1}{2}(\mathbb{I} \otimes \mathbb{I} + X \otimes X)$$
$$= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{2.55}$$

To get the Kraus representation is a bit more involving. Expanding the Kraus operators in operator basis

$$A_k = \sum_a a_{ka}\tau_a, \tag{2.56}$$

we get

$$E^*(\rho) = \sum_k \sum_{a,b=0}^{3} a_{ka}\overline{a_{kb}}\tau_a\rho\tau_b$$
$$= \sum_{a,b=0}^{3} x_{ab}\tau_a\rho\tau_b. \tag{2.57}$$

where $x_{ab} = \sum_k a_{ka}\overline{a_{kb}}$. Then observe

$$\chi(E^*) = \sum_k \sum_{a,b=0}^{3} a_{ka}\overline{a_{kb}} \sum_{i,j=0}^{1} |i\rangle\langle j| \otimes (\tau_a|i\rangle\langle j|\tau_b)$$
$$= \sum_{a,b=0}^{3} x_{ab} \sum_{i,j=0}^{1} |i\rangle\langle j| \otimes (\tau_a|i\rangle\langle j|\tau_b)$$
$$= \sum_{a,b=0}^{3} x_{ab}|\psi_a\rangle\langle\psi_b|, \tag{2.58}$$

where $|\psi_a\rangle = \sum_{i=0}^{1} |i\rangle \otimes \tau_a |i\rangle$ are orthonormal vectors. The matrix x is the Choi-Jamiolkowski operator of channel E^* in basis $\{|\psi_a\rangle\}$, therefore it is positive and diagonalizable, thus we can write

$$x_{ab} = \sum_{k=0}^{3} U_{ak} \lambda_k \overline{U_{bk}} \tag{2.59}$$

where λ_k are eigenvalues of $\chi(E^*)$, and

$$E^*(\rho) = \sum_{a,b=0}^{3} x_{ab} \tau_a \rho \tau_b = \sum_{k,a,b=0}^{3} \lambda_k U_{ak} \tau_a \rho \overline{U_{bk}} \tau_b. \tag{2.60}$$

We then get

$$A_k = \sum_{a=0}^{3} \sqrt{\lambda_k} U_{ak} \tau_a. \tag{2.61}$$

For channel E^* we obtain x_{ab} from the Choi-Jamiolkowski operator, $x_{ab} = \langle \psi_a | \chi(E^*) \psi_b \rangle$:

$$x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.62}$$

Since this is already diagonal with only one nonzero eigenvalue 1 with multiplicity 2, we know that Kraus operators will be first two members of operator basis:

$$E^*(\rho) = \tau_0 \rho \tau_0 + \tau_1 \rho \tau_1 = \frac{1}{2}(\rho + X \rho X). \tag{2.63}$$

Now we can write the Stinespring dilation of this channel. We have only two Kraus operators, thus a qubit Hilbert space $\mathcal{M}$ is sufficient. The isometry V will then be

$$V = \begin{pmatrix} \tau_0 \\ \tau_1 \end{pmatrix}$$
$$E^*(\rho) = \mathrm{Tr}_2[V \rho V^\dagger] \tag{2.64}$$

It can be checked that this dilation is minimal. This is not very surprising since the environmental overhead is the smallest nontrivial Hilbert space of dimension 2. To this V many unitaries can be formed with different $|O\rangle$ such that

$$U|O\rangle = V. \tag{2.65}$$

One such interesting unitary interaction is U_{cnot} together with $|O\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$ where

$$\begin{aligned} U_{\mathrm{cnot}} &= |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes X \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \end{aligned} \tag{2.66}$$

Remark 2.17. If the state of environment in some (not only minimal) unitary Stinespring dilation of channel $T^* : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ is complete mixture, then the channel T^* is unital. Let $T^*(\rho) := \mathrm{Tr}_{\mathcal{M}}\big(U(1/d_{\mathcal{M}}\mathbb{I}_{\mathcal{M}} \otimes \rho)U^\dagger\big)$ for some unitary $U : \mathcal{M} \otimes \mathcal{A} \mapsto \mathcal{M} \otimes \mathcal{B}$, where $d_{\mathcal{M}} = \dim \mathcal{M}$. Then $T^*(\mathbb{I}_{\mathcal{A}}) = 1/d_{\mathcal{M}}\mathrm{Tr}_{\mathcal{M}}\big(U(\mathbb{I}_{\mathcal{M}} \otimes \mathbb{I}_{\mathcal{A}})U^\dagger\big) = 1/d_{\mathcal{M}}\mathrm{Tr}_{\mathcal{M}}(\mathbb{I}_{\mathcal{M}} \otimes \mathbb{I}_{\mathcal{A}}) = \mathbb{I}_{\mathcal{A}}$.

Definition 2.18 (Contractive map)**.** A map T^* is called *contractive* if

$$\| T^*(\rho_1 - \rho_2) \|_{\mathrm{tr}} \leq k \| \rho_1 - \rho_2 \|_{\mathrm{tr}}, \tag{2.67}$$

for some $k < 1$. If this inequality holds only for $k \leq 1$, then the map is *non-expansive*.

Every channel is non-expansive, but not all channels are contractive. Due to Banach fixed point theorem, contractive channels have exactly one unique fixed point.

An important norm can be defined on the set of all linear maps.

Definition 2.19 (Norm of complete boundedness)**.** Let us have a linear mapping $T : \mathcal{L}(\mathcal{B}) \mapsto \mathcal{L}(\mathcal{A})$. The norm of complete boundedness, or the *cb*-norm is then defined as

$$\| T \|_{\mathrm{cb}} = \max_G \frac{\| (I_{\mathcal{H}} \otimes T)(G) \|_\infty}{\| G \|_\infty}, \tag{2.68}$$

where $G \in \mathcal{L}(\mathcal{H} \otimes \mathcal{A})$ for arbitrary $\mathcal{H}$ in Heisenberg picture and similar *diamond norm*

$$\| T^* \|_\diamond = \max_\rho \frac{\| (I_{\mathcal{H}} \otimes T^*)(\rho) \|_{\mathrm{tr}}}{\| \rho \|_{\mathrm{tr}}}, \tag{2.69}$$

where $\rho \in \mathcal{T}(\mathcal{H} \otimes \mathcal{A})$ for arbitrary $\mathcal{H}$ in Schrödinger picture.

The *cb*-norm can be also defined as the largest difference between the overall probabilities in two statistical quantum experiments differing only by one use of T. These experiments may involve entangling the systems on which the channels act with arbitrary further systems.

2.5 Measurement

Untill now we have considered only measurement as an observable which maps states to probability distributions on measurement outcomes. We might be interested in the state after the measurement has occurred. For this we will need the notions of a *measurement* model and *instruments*. These ideas were first formalized by Ozawa [47, 46]. A comprehensive reference on quantum measurements is the monograph by Busch et al [10]. A measurement of some quantum system can be realized by coupling the system to some other quantum system, called a probe, interaction and then measurement of the probe.

Definition 2.20 (Measurement model)**.** Let $A : \mathcal{F} \mapsto \mathcal{E}(\mathcal{H})$ be an observable (POVM) on our system of concern. A *measurement model* M is a quadruple $M = (\mathcal{K}, \rho_1, V^*, F)$, where

- $\mathcal{K}$ is a Hilbert space of the probe.
- ρ_1 is the initial state of the probe.
- V^* is a channel, $V^* : \mathcal{T}(\mathcal{H} \otimes \mathcal{K}) \mapsto \mathcal{T}(\mathcal{H} \otimes \mathcal{K})$, which describes the measurement interaction between the probe and the system.

- F is an observable of the probe with the outcome space $(\Omega, \mathcal{F})$ taking values in $\mathcal{E}(\mathcal{K})$. We call it the *pointer observable* describing the measurement of probe.

If the following *probability reducibility* condition holds, then M is a measurement model (memo) for the observable A:

$$\mathrm{Tr}[\rho A(X)] = \mathrm{Tr}[V^*\big(\rho \otimes \rho_1\big)\big(\mathbb{I} \otimes F(X)\big)] \quad \forall X \in \mathcal{F}, \rho \in \mathcal{S}(\mathcal{H}). \tag{2.70}$$

We could have just measured the system itself, without attaching the probe, however this is usually how the measurements work in reality. A direct measurement of a particle often completely destroys the particle, e.g. photon absorption.

A measurement model also defines the state of system after measurement. Let us have a measurement model $M = (\mathcal{K}, \rho_1, V^*, F)$ which defines an observable A with outcome space $(\Omega, \mathcal{F})$. The state of system after measuring event $X \in \mathcal{F}$ is

$$\widetilde{\rho}_X = \frac{1}{\mathrm{Tr}(V^*(\rho \otimes \rho_1)\mathbb{I} \otimes F(X))} \mathrm{Tr}_{\mathcal{K}}(V^*(\rho \otimes \rho_1)\mathbb{I} \otimes F(X)). \tag{2.71}$$

We call the $\widetilde{\rho}_X$ a *conditional output state*. We can introduce an operation

$$\mathcal{I}_X^M(\rho) = \mathrm{Tr}_{\mathcal{K}}(V^*(\rho \otimes \rho_1)\mathbb{I} \otimes F(X)), \tag{2.72}$$

and the conditional output state is

$$\widetilde{\rho}_X = \frac{1}{\mathrm{Tr}(\mathcal{I}_X^M(\rho))} \mathcal{I}_X^M(\rho). \tag{2.73}$$

The mapping $\mathcal{I}_X^M$ has following properties

(C1) for each $X \in \mathcal{F}$, $\mathcal{I}_X^M$ is linear, completely positive and trace non-increasing

(C2) $\mathrm{Tr}(\mathcal{I}_\Omega^M(\rho)) = 1$ and $\mathcal{I}_\emptyset^M(\rho) = O$ for all $\rho \in \mathcal{S}(\mathcal{H})$

(C3) If $\{X_j\}$ is a sequence of mutually disjoint sets, then

$$\mathrm{Tr}(\mathcal{I}^M_{\cup_j X_j}) = \sum_j \mathrm{Tr}(\mathcal{I}^M_{X_j}).$$

These properties can be abstracted and lead to following definition.

Definition 2.21 (Instrument)**.** A mapping $\mathcal{I}$ from outcome space $(\Omega, \mathcal{F})$ to the set of trace non-increasing CP-maps on $\mathcal{H}$ is called an *instrument* if it satisfies the properties (C1)-(C3).

We see that every measurement model M defines an instrument $\mathcal{I}^M$ and we say that $\mathcal{I}^M$ is instrument induced by M. Furthermore, due to Ozawa's theorem [47], for every instrument there exists a corresponding measurement model.

3 Process estimation

In this short chapter we will introduce some tools for interpreting the data obtained from quantum mechanical experiments. We will start by inverting the statistics to assess the preparation procedure, i.e. the state estimation. Then we will extend this notion to process estimation. The estimation of processes is essentially the same as estimation of states since every completely positive map can be via Choi-Jamiolkowski isomorphism connected with appropriate (unnormalized) state on a larger Hilbert space. Thus the estimation of quantum process can be viewed as estimation of its Choi-Jamiolkowski state. Since the statistics obtained in experiments are finite, we don't have exact probabilities, the inverted map is not necessarily completely positive. If we consider this just to be the effect of finite statistics we can use Bayesian approach and search for a channel with highest likelihood in the set of all channels. This will naturally yield a valid physical map. In reality this is a hard problem of finding global maximum/minimum and has to be solved numerically, with all disadvantages of this approach. It is important that all these approaches assume to have in principle infinitely many independent copies of the state or process. For more information and references about estimations see Ref. [48].

3.1 State estimation

Assume that we have an experimental setup capable of preparing some unknown state $\rho \in \mathcal{S}(\mathcal{H})$. To reconstruct the density matrix of this state we need to measure some observable A to get probability distribution over all effects of the observable. Let E_k correspond to primitive effects of the observable A. Then we try to measure the probabilities $p(\rho)_k = \mathrm{Tr}(\rho E_k)$ by counting the events when particular outcome has occurred. If we want to completely determine the state ρ we need to have a one to one correspondence between the state and the measured probability distribution. This leads to a notion of an informationally complete observable introduced by Prugovečki [55]

Definition 3.1 (Informationally complete observable)**.** Observable A is informationally complete if

$$\Phi_A(\rho_1) = \Phi_A(\rho_2) \Rightarrow \rho_1 = \rho_2, \tag{3.1}$$

where $\Phi_A(\rho)$ is probability distribution over outcomes of observable A for state ρ.

Informational completeness assures us that whenever we measure some particular probability distribution, then this probability distribution corresponds to a unique state. In such case we are able to invert the probabilities to obtain ρ. Suppose A is informationally complete. Every effect can be expanded in the traceless operator basis

$$E_k = \sum_j Q_{kj}\tau_j \tag{3.2}$$

and so can be the state ρ

$$\rho = \sum_j r_j\tau_j. \tag{3.3}$$

The probability p_k is then

$$p(\rho)_k = \mathrm{Tr}(\rho E_k) = \sum_{jl} Q_{kj} r_l \mathrm{Tr}(\tau_l \tau_j) = \sum_l Q_{kl} r_l, \tag{3.4}$$

where we used the orthonormality of the traceless operator basis. This can be written in matrix form as $\vec{p}(\rho) = Q.\vec{r}$. To reconstruct the state we make a left inversion

$$Q^{-1}\vec{p}(\rho) = \vec{r}. \tag{3.5}$$

The condition of informational completeness ensures that Q is left invertible.

Example 3.2 (S-G along three axes)**.** A common measurement is a Stern Gerlach measurement along some axis, lets say z-axis. In ideal case this is a PVM with two outcomes with effects being the density matrices of eigenstates of Pauli matrix Z:

$$E_{z+} = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_{z-} = |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.6}$$

This is not an informationally complete observable. For example the eigenstates of X, $|x+\rangle = 1/\sqrt{2}(|0\rangle+|1\rangle)$ and $|x-\rangle = 1/\sqrt{2}(|0\rangle-|1\rangle)$ yield equal probability distributions over E_{z+}, E_{z-}, $p_{z+}(x+) = p_{z+}(x-) = p_{z-}(x+) = p_{z-}(x-) = 1/2$. We can add two S-G measurements along remaining orthogonal axes to get an informationally complete observable. The added effects are

$$E_{x+} = |x+\rangle\langle x+| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad E_{x-} = |x-\rangle\langle x-| = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \tag{3.7}$$

along the x-axis and

$$E_{y+} = |y+\rangle\langle y+| = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}, \quad E_{y-} = |y-\rangle\langle y-| = \frac{1}{2}\begin{pmatrix} 0 & i \\ -i & 1 \end{pmatrix}, \tag{3.8}$$

along the y axis. We have to normalize the effects when combining more measurements into one observable. Let us use the qubit traceless operator basis from (2.9) $\{\tau_0, \tau_1, \tau_2, \tau_3\}$. The matrix Q is then

$$Q = \frac{1}{3\sqrt{2}}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix}, \tag{3.9}$$

where we assume the order

$$\vec{p} = (p_{x+}, p_{x-}, p_{y+}, p_{y-}, p_{z+}, p_{z-})^T. \tag{3.10}$$

This observable is in fact informationally overcomplete and we need any four linearly independent rows from the matrix. This is a bit inconvenient since in real application various rows will give slightly different results, depending on the size of your statistics.

The class of informationally complete observables is broad. Some observables might be better then others simply because they use less outcomes. Minimal informationally complete observables are those which use the smallest outcome space. In d-dimensional Hilbert space, the smallest informationally complete observable has d^2 outcomes - the dimension of the operator space. From the set of all minimal informationally complete observables we may select a special class of symmetric informationally complete observables.

Definition 3.3 (SIC observable)**.** An observable A on Hilbert space $\mathcal{H}$ is *symmetric informationally complete* (SIC) if:

1. A is minimal, ie. has d^2 outcomes $k \in (1, 2, \dots d^2)$;
2. each effect E_k is a rank-1 operator;
3. $\mathrm{Tr}(E_k) = \alpha$, where α is constant for all k;
4. $\mathrm{Tr}(E_k E_l) = \beta$, where β is constant for all $k \neq l$.

SIC observables are interesting for the high symmetry and efficiency in estimation protocols. It is not known whether there exists a SIC observable in every finite dimensional Hilbert space. Numerical studies have shown that there are SIC observables in all dimensions up to 67 [63]. The constants α and β are fixed by the dimension of Hilbert space $\mathcal{H}$: $\alpha = 1/d$ and $\beta = 1/(d^2(d+1))$.

Example 3.4 (SIC qubit observable)**.** The most efficient qubit observable needs at least four effects, three of them are independent, for three independent parameters characterizing the qubit state. The effects have trace $\mathrm{Tr}E_k = 1/2$ and have rank 1, thus they can be parametrized as $E_k = \frac{1}{4}(\mathbb{I} + \vec{r}_k\vec{\sigma})$. The second condition $\mathrm{Tr}(E_k E_l) = 1/12$ gives

$$\frac{1}{12} = \mathrm{Tr}(E_k E_l) = \frac{1}{8}(1 + \vec{r}_k\vec{r}_l) \Rightarrow \vec{r}_k\vec{r}_l = -\frac{1}{3}. \tag{3.11}$$

As a result vectors $\vec{r}_k$ point into corners of a tetrahedron. One possible choice can be

$$\begin{array}{ll} \vec{r}_1 = \frac{1}{\sqrt{3}}(1,1,1) & \vec{r}_2 = \frac{1}{\sqrt{3}}(1,-1,-1) \\ \vec{r}_3 = \frac{1}{\sqrt{3}}(-1,1,-1) & \vec{r}_4 = \frac{1}{\sqrt{3}}(-1,-1,1), \end{array} \tag{3.12}$$

and the matrix Q from (3.2) is $Q_{kj} = \vec{r}_{kj}$.

Sometimes we don't need to fully estimate a state, but instead just test a discrete set of alternatives. For example we have a state ρ which is known to be either ρ_1 or ρ_2. The problem is to answer which of these two it is. If we have only single copy of this state, we can either decide for a strategy where the probability of a wrong identification will be minimal or the probability of wrong identification will be zero, but the probability of inconclusive outcome will be greater than zero. The former strategy is called *minimal error discrimination* and the latter *unambiguous discrimination*.

In minimal error discrimination strategy the probability of error is proportional to the trace norm of operator $\rho_1 - \rho_2$ (see cf. chapter 3 in [29])

$$p_{\mathrm{error}} = \frac{1}{2}(1 - \frac{1}{2}\| \rho_1 - \rho_2 \|_{\mathrm{tr}}), \tag{3.13}$$

where

$$\frac{1}{2}\| \rho_1 - \rho_2 \|_{\mathrm{tr}} =: D(\rho_1, \rho_2) \tag{3.14}$$

is the trace distance between two states. Trace distance has a clear operational meaning. The larger the trace distance is, the smaller the probability of error becomes and hence the easier it is to discriminate between the two states. Naturally, if we are given more copies of the state ρ, it becomes easier to discriminate between the two possibilities, hence

$$D(\rho_1^{\otimes n}, \rho_2^{\otimes n}) \geq D(\rho_1, \rho_2), \tag{3.15}$$

where n is the number of copies of ρ we are given.

Another common way of measuring the difference between quantum states is the *fidelity* of two states ρ_1 and ρ_2

$$F(\rho_1, \rho_2) = \mathrm{Tr}(|\sqrt{\rho_1}\sqrt{\rho_2}|) = \mathrm{Tr}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}. \tag{3.16}$$

This quantity provides an upper bound on the optimal success probability for the unambiguous discrimination of pair of states.

3.2 Process estimation

Let us have an unknown process $T^* : \mathcal{T}(A) \mapsto \mathcal{T}(B)$. If we can prepare state $|I\rangle \in \mathcal{S}(\mathcal{A} \otimes \mathcal{A})$ such that

$$|I\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |j\rangle|j\rangle, \tag{3.17}$$

then

$$(I \otimes T^*)(|I\rangle\langle I|) = \frac{1}{d}\chi(T^*) \tag{3.18}$$

is the scaled Choi-Jamiolkowski operator of channel T^*. This can be measured using state estimation on the bipartite system. More generally, any estimation procedure of $T^* : \mathcal{T}(A) \mapsto \mathcal{T}(B)$ consists of preparation of a state $\rho \in \mathcal{S}(\mathcal{K} \otimes \mathcal{A})$, application of the map $\mathbb{I}_{\mathcal{K}} \otimes T^*(\rho)$ and some measurement of observable A on the Hilbert space $\mathcal{K} \otimes \mathcal{B}$ with effects E_n. Repeating this procedure many times for various input states ρ_k and measurement observables A_l with effects E_{kn}, one then measures conditional probabilities of events $p(E_{ln}|\rho_k, A_l)$

$$p(E_{ln}|\rho_k, A_l) = \mathrm{Tr}[\tag{3.19}$$

$$big(idc_{\mathcal{K}} \otimes T^*(\rho_k)\big)E_{kl}] = \frac{p(E_{ln,}\rho_k, A_l)}{p(\rho_k, A_l)}, \tag{3.20}$$

where $p(\rho_k, A_l)$ is the probability that we have chosen preparation k and measurement l and $p(E_{ln,}\rho_k, A_l)$ is the overall probability of observing event composed of preparation k and effect E_{ln}. The overall probabilities still contain information about details of the estimation scheme. You may choose to repeat the procedure for certain preparations more often then for the other

as well as you can use for each preparation a different measurement. The conditional probabilities $p(E_{ln}|\rho_k, A_l)$ are free of this information, they depend only on the process. From these probabilities we can then reconstruct desired parameters.

It is possible to do the complete process estimation using only local states and one local, informationally complete measurement. Let us have some fixed set of states $\{\rho_k\} \in \mathcal{S}(\mathcal{A})$ and an informationally complete observable A in $\mathcal{B}$ with effects E_n. Using the channel many times one measures the probabilities

$$p_{n|k} := p(E_n|\rho_k) = \mathrm{Tr}\big(T^*(\rho_k)E_n\big). \tag{3.21}$$

Again we can define the matrix Q as in (3.2) and a matrix S such that

$$\begin{aligned} E_n &= \sum_j Q_{nj}\tau_j \\ \rho_k &= \sum_i S_{ki}\tau_i \end{aligned} \tag{3.22}$$

where $\{\tau_i\}$ is the traceless operator basis in $\mathcal{A}$ and in $\mathcal{B}$ if they have same dimensions. One can then write

$$\begin{aligned} p_{n|k} &= \sum_{ij} Q_{nj}S_{ki}\mathrm{Tr}\big(T^*(\tau_i)\tau_j\big) = \sum_{ij} Q_{nj}S_{ki}A_{ij} \\ &\Rightarrow \quad p = SAQ^T \end{aligned} \tag{3.23}$$

where A_{ij} is the vector representation of channel T^*. The matrix A can be reconstructed by inverting the matrices Q and S as in previous section. For this we need that the set of states $\{\rho_k\}$ spans the whole operator space, then the matrix S is invertible.

As with states, we can also do process discrimination. Minimum error discrimination of two processes, T_1^*, T_2^* then falls down to discrimination of the output states $T_1^*(\rho)$ and $T_2^*(\rho)$. We can use a strategy when we prepare an input state $\rho \in \mathcal{S}(\mathcal{H} \otimes \mathcal{A})$ and discriminate states $\mathbb{I}_\mathcal{H} \otimes T_1^*(\rho)$ and $\mathbb{I}_\mathcal{H} \otimes T_2^*(\rho)$. Thus the distinguishability of two processes will be proportional to the largest trace distance of states $\mathbb{I}_\mathcal{H} \otimes T_1^*(\rho)$ and $\mathbb{I}_\mathcal{H} \otimes T_2^*(\rho)$, that is the cb-norm (2.68).

3.3 Maximum likelihood

Both methods from previous two sections require measuring probabilities. This can never be done perfectly with finite statistics. In real experiment we measure the number of occurrences n_k of some event k. For large enough statistics the probability is then assumed to be n_k/N where N is the number of all events. Due to the small errors in probabilities the resulting inversion might not be a physical state or process. We can then ask what process or state describes the measured data best. In such case we can use the maximum likelihood method to estimate the parameters of a process or state. We construct the likelihood function as

$$L(R, D) = p(D \mid R) \tag{3.24}$$

where R is the set of parameters of the statistical model we try to estimate and D represents the measured data. Likelihood function is then the probability of measuring data D given that

the statistical model has parameters R. Maximizing this function over the allowed set of model parameters will give us the parameters R such that this probability is maximal.

Main problem with this method is that the likelihood function tends to be quite complex, and finding the global maximum is a very hard task, especially when the maximum lies on the border of the parameter space. On the other side it has lot of nice properties. It uses all measured data, no post selection is needed. It behaves well asymptotically, for large statistics it converges to the real parameters of the model. By construction, the result is always a physical map or state.

3.4 Teleportation experiment

In an actual experiment of group in Heidelberg under supervision of Prof. Jian-Wei Pan, a scheme for teleportation was tested. Teleportation is a scheme for communicating one quantum state from site A to site B. Without going into details of the protocol, the resulting transformation is an ideal channel from sender to receiver. At the end one has to assess how much the practical realization did deviate from an ideal channel, and thus how successful the teleportation was. In this experiment, six preparations were used: $\{H, V, P, N, R, L\}$, corresponding to various light polarizations. Ideally ρ_H, ρ_V should be the horizontal and vertical polarizations, $|0\rangle$ and $|1\rangle$, ρ_P, ρ_N should be the $|+\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle), |+\rangle = 1/\sqrt{2}(|0\rangle - |1\rangle)$ and finally ρ_R, ρ_L the $|R\rangle = 1/\sqrt{2}(|0\rangle + i|1\rangle), |L\rangle = 1/\sqrt{2}(|0\rangle - i|1\rangle)$. However the preparations are not ideal and we have to estimate them. Three projective measurements with outcomes H, V, P, N and R, L polarizations were used where $E_H = 1/2(\mathbb{I} + Z)$, $E_V = 1/2(\mathbb{I} - Z)$, $E_P = 1/2(\mathbb{I} + X)$,$E_N = 1/2(\mathbb{I} - X)$, $E_R = 1/2(\mathbb{I} + Y)$ and $E_L = 1/2(\mathbb{I} - Y)$, this is equivalent to the ones in example 3.2 . The measured data is summarized in table 1.

Two datasets are shown. One for estimating the preparations, where no teleportation of the inputs has occurred and one for teleportation of the preparations. The counts for the teleportation are much lower due to probabilistic nature of the protocol.

Since the data for preparation estimation is overcomplete for each preparation, it is best to choose the maximal likelihood method to find the prepared states. Each preparation k is a qubit state parametrized by a three dimensional vector $\vec{r_k}$ such that

$$\rho_k = \frac{1}{2}(\mathbb{I} + r_{k,x}X + r_{k,y}Y + r_{k,z}Z). \tag{3.25}$$

The likelihood function is then the probability that preparation k with parameters $\vec{r_k}$ will yield measured set of counts $\{n_{k,l}\}$, $l \in \{V, H, N, P, R, L\}$:

$$L(\vec{r_k}, \{n_{k,l}\}) = \prod_l p_{l|k}^{n_{k,l}}, \tag{3.26}$$

where

$$p_{l|k} = \mathrm{Tr}(\rho_k E_l) \tag{3.27}$$

is the probability of measuring effect E_l when preparation k with parameters $\vec{r_k}$ was input. Maximizing over the parameters $\vec{r_k}$ we obtain the result[5]. Alternatively it is numerically easier to,

[5] Actually one can use directly the overall probabilities of events, since this just changes the multiplicative constant in front of the likelihood function, but doesn't change the position of maxima in the parameter space.

Prep.	Out.	Count	Tel. count	Prep.	Out.	Count	Tel. count
H	H	1305	53	V	H	33	7
	V	29	0		V	1067	56
	P	613	30		P	709	27
	N	670	20		N	516	42
	R	718	30		R	497	30
	L	592	28		L	652	38
P	H	759	25	N	H	548	20
	V	596	23		V	557	32
	P	1426	55		P	26	7
	N	25	6		N	1052	47
	R	674	36		R	586	45
	L	686	51		L	542	20
R	H	635	21	L	H	657	25
	V	672	29		V	506	41
	P	777	45		P	588	33
	N	576	26		N	591	40
	R	1369	49		R	39	4
	L	49	7		L	1084	49

Tab. 1. Experimental data for teleportation protocol. (H,V), (P,N) and (R,L) correspond to qubit preparations and measurements along three perpendicular axes. In column "Count" data for measurement on preparations is placed, $n_{k,l}$, and in column "Tel. count" measured data for teleported preparations, $n_{(k,l)}$.The teleported data have much lower count due to probabilistic nature of the protocol used.

instead of maximizing the function L, which contains very small numbers for large counts to minimize $-\log L = \sum_l n_{k,l} \log 1/p_{l|k}$. Either way, using this method we obtain following results

$$\rho_H = \begin{pmatrix} 0.98 & -0.02 - i0.05 \\ -0.02 + i0.05 & 0.02 \end{pmatrix} \quad \rho_V = \begin{pmatrix} 0.03 & 0.08 + i0.07 \\ 0.08 - i0.07 & 0.97 \end{pmatrix}$$
$$\rho_P = \begin{pmatrix} 0.56 & 0.48 \\ 0.48 & 0.44 \end{pmatrix} \quad \rho_N = \begin{pmatrix} 0.5 & -0.48 - i0.02 \\ -0.48 + i0.02 & 0.5 \end{pmatrix}$$
$$\rho_R = \begin{pmatrix} 0.49 & 0.07 - i0.47 \\ 0.07 + i0.47 & 0.51 \end{pmatrix} \quad \rho_L = \begin{pmatrix} 0.56 & +i0.47 \\ -i0.47 & 0.44 \end{pmatrix}. \tag{3.28}$$

Then we can continue in same manner with process estimation. We parametrize the process T^* by its Choi-Jamiolkowski state

$$\chi(T^*) = \sum_{a,b=0}^{1} 1/2|a\rangle\langle b| \otimes T^*(|a\rangle\langle b|) \tag{3.29}$$

and the probability of event (k, l) is then

$$p(k, l) = \mathrm{Tr}(T^*(\rho_k)E_l) = \mathrm{Tr}(\chi(T^*)\rho_k^T \otimes E_l), \tag{3.30}$$

with the likelihood function

$$L(\chi, n_{(k,l)}) = \prod_{(k,l)} p(k,l)^{n_{(k,l)}}. \tag{3.31}$$

This yields

$$\chi(T^*) = \begin{pmatrix} 0.46 & 0.04 - i0.05 & -0.02 - i0.04 & 0.42 - i0.04 \\ 0.04 + i0.05 & 0.04 & 0.03 + i0.01 & 0.02 + i0.04 \\ -0.02 + i0.04 & 0.03 - i0.01 & 0.06 & -0.04 + i0.05 \\ 0.42 + i0.04 & 0.02 - i0.04 & -0.04 - i0.05 & 0.44 \end{pmatrix}. \tag{3.32}$$

For comparison the Choi-Jamiolkowski state of ideal transfer in this representation is

$$\chi(I) = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{3.33}$$

For numerical minimization of the function $-\log L(\chi, n_{(k,l)})$ we used downhill simplex method, also known as Nelder-Maeds method.

A separate problem is to introduce "error bars" on the result. A commonly used procedure is to perform Monte-Carlo simulation of the experiment, with events distributed according to Poisson distribution with mean values of the actual measured data. Then from these datasets estimate the state or process, and the error in terms of a distance measure, for example fidelity of the Choi-Jamiolkowski state.

To obtain the error bars of the preparation procedures we made 10000 simulations. Results are summarized in table 2.

Preparation	Mean fidelity [%]	5σ interval [%]
H	98	± 2
V	97	± 3
P	98	± 2
N	98	± 2
R	97	± 2
L	97	± 3

Tab. 2. Fidelity of preparation procedures.

The fidelity of the teleportation was $85 \pm 12\%$(5 sigma) with respect to ideal transfer, where we sampled 1044 simulations to obtain the distribution, see figure 3.1. This is well above the classical threshold of $2/3$, which can be achieved without entanglement.

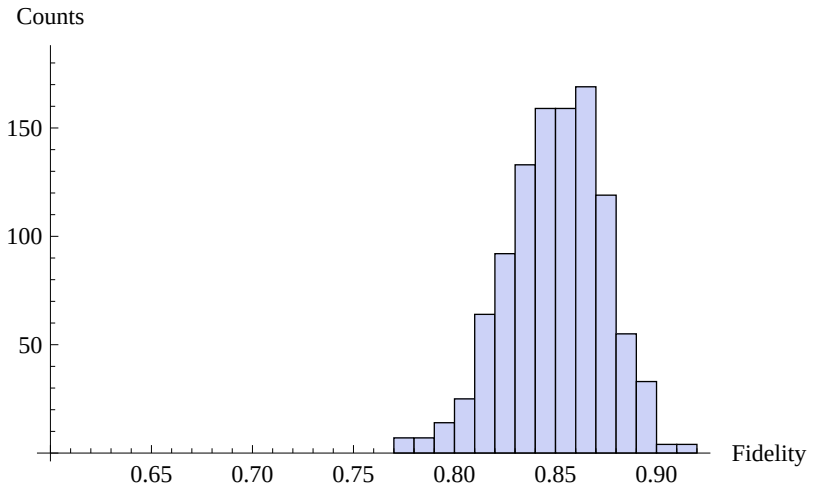

Fig. 3.1. Distribution of fidelities of 1000 simulations of the experiment. The 5 sigma interval covers the whole data set.

4 Quantum memory channels

Schrödinger equation governs the evolution of a closed quantum system. Such evolution is unitary. It is difficult if not impossible to experimentally achieve such ideal situation. Unavoidable interaction between the system under consideration and environment occurs leading to a non-unitary evolution of the system. Such non-unitary evolution is described by a quantum channel - a linear, completely positive and trace preserving map in Schrödinger picture on quantum states or equivalently by linear, completely positive and unital map in Heisenberg picture on observables. Stinespring theorem gives us full justification for this step, because it tells us that every quantum channel can be extended to a unitary evolution on a larger Hilbert space.

In this chapter we will introduce the model of quantum channels with memory, memory channels. Quantum memory channel is a simple quantum channel of some system, where the system has an inner structure and additional physical requirement of causality is placed on the channel. The system is an ordered, infinite sequence of subsystems and the causality condition requires that information can move only in one way. Information stored in a particular subsystem can then affect evolution only of subsystems which follow in sequence after. This can be interpreted as a simple collision model where some quantum device sequentially processes a sequence of quantum signals. The research of memory channels is mostly focused on investigation of transmission rates for particular classes of memory channels [38, 43, 44, 3, 5, 24, 33, 32, 14, 17, 16, 11, 69, 20]. Recently, attention has been paid to an interesting class of the so-called bosonic memory channels [25, 12, 59, 58, 52, 42, 40, 61, 41] and also to memory effects in the transmission of quantum states over spin chains [53, 4, 54, 57].

4.1 Memory channel as general causal process

As we have already mentioned a memory channel is just a special case of a simple quantum channel on a specially structured input. The aforementioned structure of input is as follows. The input is a chain of ordered quantum systems each with a finite dimensional Hilbert space $\mathcal{A}_k$ where $k \in \mathbb{Z}$ denotes the ordering. This can be understood as infinite sequence of quantum particles. Since this sequence is potentially infinite, we will start by describing the channel in Heisenberg picture.

Every subsystem comes with an algebra of observables $\mathfrak{A}_k = \mathcal{L}(\mathcal{A}_k)$. If we would like to denote algebras associated to sets of subsystems $\Lambda \subset \mathbb{Z}$ we follow this notation: if Λ is a finite subset of $\mathbb{Z}$ then $\mathfrak{A}_\Lambda = \bigotimes_{k\in\Lambda} \mathfrak{A}_k$. $\mathfrak{A}_\Lambda$ is then just a simple tensor product of selected subsystems. In case of infinite $\Lambda \subset \mathbb{Z}$ we associate with $\mathfrak{A}_\Lambda$ the C^*-closure of increasing family of finite dimensional algebras $\mathfrak{A}_{\Lambda_f}$ for finite $\Lambda_f \subset \Lambda$. This is also called a *quasi-local algebra* [6]. From now on $\mathfrak{A}_\mathbb{Z}$ will be abbreviated as $\mathfrak{Z}$ and we will also use the shorthand $\mathfrak{A}_-$ and $\mathfrak{A}_+$ for the left and right hand chain halves $\mathfrak{A}_{(-\infty,0]}$ and $\mathfrak{A}_{[1,\infty)}$ when it comes in handy. For clarity we will also make difference between input algebras $\mathfrak{A}$ and output algebras $\mathfrak{B}$ even if we consider them isomorphic. The terms $\mathfrak{B}_{[a,b]}$,$\mathfrak{B}_\pm$ and similar are then defined analogously. Also we will denote the Hilbert space of outputs with $\mathcal{B}_k$ even though it is isomorphic to $\mathcal{A}_k$

Definition 4.1 (Causal process)**.** Causal process is a channel (in Heisenberg picture) $T : \mathfrak{Z} \mapsto \mathfrak{Z}$ such that for every $z \in \mathbb{Z}$

$$T(b_{(-\infty,z]} \otimes \mathbb{I}^{\mathcal{B}}_{[z+1,\infty)}) = T(b_{(-\infty,z]}) \otimes \mathbb{I}^{\mathcal{A}}_{[z+1,\infty)}, \tag{4.1}$$

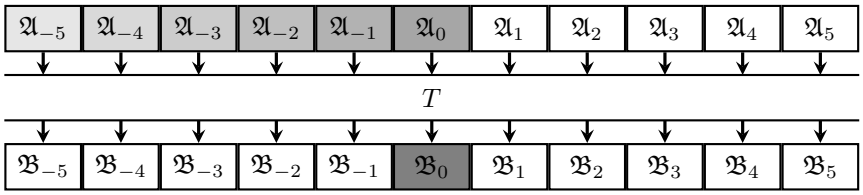

Fig. 4.1. A causal process has the property that any information that is stored at the output $\mathfrak{B}_0$(bottom gray) could only originate from inputs on the left from this site (upper gray region).

where $B_{(-\infty,z]} \in \mathfrak{B}_{(-\infty,z]}$ and $\mathbb{I}^{\mathcal{B}}_{[z+1,\infty)}$ is the identity operator in $\mathfrak{B}_{[z+1,\infty)}$ and $\mathbb{I}^{\mathcal{A}}_{[z+1,\infty)}$ is the identity operator in $\mathfrak{A}_{[z+1,\infty)}$.

Remark 4.2 (Notation)**.** To avoid too many identical subscripts we adopted following convention. For any causal channel $T(\mathfrak{B})$, if we would like to address its transformation on a subpart localized in region $[a,b]$ we could write $T_{[a,b]}(\mathfrak{B}_{[a,b]})$, however we omit the first subscript and write only $T(\mathfrak{B}_{[a,b]})$. When we see $T(\cdot)$ we need to know where the $\cdot$ is localized to correctly understand the transformation. This goes also for Schrödinger picture where the transformation will be specified by the localization of inputs.

The equation (4.1) just reflects the causality condition placed on the channel T. It means that whatever you can measure on the output with $b_{(-\infty,z]}$ can be measured on inputs up to the z-th particle with $T(b_{(-\infty,z]}) \in \mathfrak{A}_{(-\infty,z]}$. This is illustrated on Figure 4.1 where it is shown that a single localized output cell is smeared over infinitely many input cells localized left from the original cell.

We can also write this definition in Schrödinger picture. A causal process is a sequence of channels (in Schrödinger picture) $T^* : \mathcal{T}(\mathcal{A}_{(-\infty,z]}) \mapsto \mathcal{T}(\mathcal{B}_{[-\infty,z]})$ for every $z \in \mathbb{N}$ such that

$$T^*(\omega_{(-\infty,z]}) = \mathrm{Tr}_{[z+1,\infty)}[T^*(\omega_{(-\infty,\infty)})], \tag{4.2}$$

for every $z \in \mathbb{Z}$, and every $\omega_{(-\infty,z]}$, $\omega_{(-\infty,\infty)}$ such that

$$\omega_{(-\infty,z]} = \mathrm{Tr}_{[z+1,\infty)}(\omega_{(-\infty,\infty)}), \tag{4.3}$$

where $\mathrm{Tr}_{[a,b]}$ denotes partial trace over the interval.

Example 4.3 (Shift of a chain)**.** The channel σ_x (in Heisenberg picture) is defined on a translationally invariant chain of quantum systems with Hilbert spaces $\mathcal{A}_j \equiv \mathcal{A}$ and appropriate observable algebras $\mathfrak{A}_j \equiv \mathfrak{A}$ and $\mathfrak{B}_j \equiv \mathfrak{B}$. The σ_x is defined by

$$\sigma_x(\lambda^a_j) = \lambda^a_{j-x} \tag{4.4}$$

for all $\lambda^a_y \in \mathfrak{A}_y$ where $\{\lambda^a_y\}_a$ form the operator basis in $\mathfrak{A}_y$. This shift channel just shifts the chain by x sites to the left. In Schrödinger picture it can be defined as

$$\sigma^*_x(\omega_{[1,n]}) = \omega_{[1-x,n-x]}. \tag{4.5}$$

A memory channel is a general causal process. A physically more compelling insight might be conveyed through the view of collision models in next section.

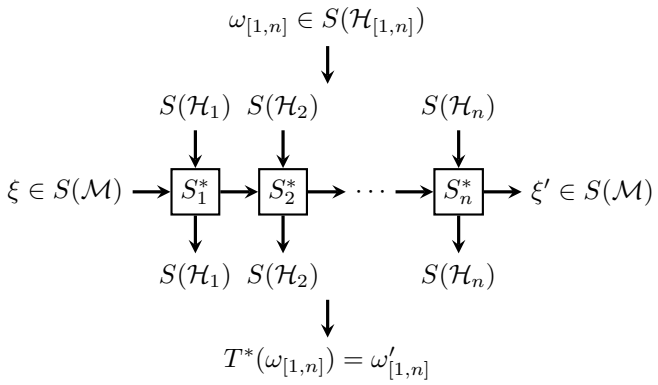

Fig. 4.2. Environment particle ξ is colliding causally with n particles of input sequence, where each collision is described by a channel S_i

4.2 Collision model

Let us look on a situation where an infinite sequence of quantum systems is colliding one at a time with some environment. This situation is illustrated on figure 4.2.

The environment is described by a single quantum system in Hilbert space $\mathcal{M}$ in state ξ. This environment sequentially collides with subparts of the input signal of length n. The i-th collision is described by a channel $S_i^* : \mathcal{T}(\mathcal{M}) \otimes \mathcal{T}(\mathcal{A}_i) \mapsto \mathcal{T}(\mathcal{B}_i) \otimes \mathcal{T}(\mathcal{M})$. The input signal is then processed with a channel T^*:

$$T^*(\omega_{[1,n]}) = \mathrm{Tr}_{\mathcal{M}}[S^*_{[1,n]}(\xi \otimes \omega_{[1,n]})], \tag{4.6}$$

where $S^*_{[1,n]}$ is an n-fold concatenation of subsequent collisions,

$$\begin{aligned} S^*_{[1,n]} &= (\mathbb{I}_{[1,n-1]} \otimes S_n^*) \cdots (\mathbb{I}_1 \otimes S_2^* \otimes \mathbb{I}_{[3,n]}) \cdot (S_1^* \otimes \mathbb{I}_{[2,n]}) \\ &= S_n^* \cdots S_2^* \cdot S_1^*. \end{aligned} \tag{4.7}$$

Note that in the definition of collision a swap of environmental system and input is included by default. Though this makes things a tiny bit less readable for the first time, it simplifies significantly the notation for concatenation of collisions as can already be appreciated in equation (4.7).

If $S_i^* = S^*$ for all i the we call such collision model translationally invariant. If all S_i^* are unitary we call such collision model pure. In previous lines we have assumed implicitly that the environment and the input signals are initially uncorrelated. We would like to stress that this does not need to be the case, however it won't change the picture and the message of this section. One thing needs to be said though, that the description by channel T^* is possible only when the memory is initially uncorrelated with the input signal, otherwise T^* wont be a channel, though the way how to acquire the output of the memory channel will remain the same.

It is easy to see that this channel is causal from the construction of the model. Each subsequent collision can change only the environment or the colliding signal, all signals which collided

before cannot be affected by subsequent inputs. However the inputs which follow after can be affected, where the effect is mediated through the environmental particle ξ. This is also the motivation to call this environment memory, because it "remembers" the inputs from "past" and mediates their effect in "present". From now on *memory* will refer to the environment system in the appropriate collision model.

Let us continue with a very basic example of a simple swap interaction where the memory and colliding input are simply swapped.

Example 4.4 (Swap collision)**.** Let $\sigma^* : \mathcal{T}(\mathcal{M}) \otimes \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B}) \otimes \mathcal{T}(\mathcal{M})$ be a translationally invariant pure collision model with a qubit memory and a chain of qubits where the unitary interaction is defined as

$$U_\sigma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4.8}$$

with $U_\sigma : \mathcal{A} \otimes \mathcal{M} \mapsto \mathcal{B} \otimes \mathcal{M}$. Let the initial state of memory be $\xi \in \mathcal{T}(\mathcal{M}) \equiv \mathcal{T}(\mathcal{A})$, then the transformation on first qubit in state $\omega_1 \in \mathcal{T}(\mathcal{A}_1)$ reads

$$\begin{aligned} T^*(\omega_1) &= \mathrm{Tr}_{\mathcal{M}}[{\sigma^*}_1(\xi \otimes \omega_1)] = \mathrm{Tr}_{\mathcal{M}}[U_\sigma(\xi \otimes \omega_1)U_\sigma^\dagger] \\ &= \mathrm{Tr}_{\mathcal{M}}[\omega_1 \otimes \xi] = \xi. \end{aligned} \tag{4.9}$$

The state of memory after first collision is

$$\xi' = \mathrm{Tr}_1[U_\sigma(\xi \otimes \omega_1)U_\sigma^\dagger] = \omega_1. \tag{4.10}$$

The transformation on first n qubits is then straightforward:

$$\begin{aligned} T^*(\omega_{[1,n]}) &= \mathrm{Tr}_{\mathcal{M}}[{\sigma^*}_{[1,n]}(\xi \otimes \omega_{[1,n]})] \\ &= \mathrm{Tr}_{\mathcal{M}}[(\mathbb{I}_{[1,n-1]} \otimes FU_\sigma)\dots(FU_\sigma \otimes \mathbb{I}_{[2,n]}) \\ &\quad (\xi \otimes \omega_{[1,n]})(U_\sigma^\dagger F^\dagger \otimes \mathbb{I}_{[2,n]})\dots(\mathbb{I}_{[1,n-1]} \otimes U_\sigma^\dagger F^\dagger)] \\ &= \mathrm{Tr}_{\mathcal{M}}[\xi \otimes \omega_{[1,n]}] = \xi \otimes \omega_{[1,n-1]}, \end{aligned} \tag{4.11}$$

where F swaps the memory and the system and is included for the nice concatenation properties. This way the memory system is the last subsystem after transformation and first at the begining. Note that in this case $F = U = U^\dagger = F^\dagger$ and the interaction seems trivial, however the position of memory system has changed. Thus the state of memory system after n collisions will be $\xi' = \omega_n$. This is in fact the shift channel σ_1^* from Example 4.3.

Another interesting example consists of control-U interaction.

Example 4.5. [Controlled U collision] Let us have again a translationally invariant pure collision model $\gamma^* : \mathcal{T}(\mathcal{M}) \otimes \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B}) \otimes \mathcal{T}(\mathcal{M})$ with memory living in Hilbert space $\mathcal{M}$, $\dim \mathcal{M} = d_{\mathcal{M}}$ and chain of qudits in Hilbert space $\mathcal{A}$, $\dim \mathcal{A} = d$. Where the collision is described by unitary

$$U_\gamma = \sum_{i=0}^{d_{\mathcal{M}}-1} |i\rangle\langle i| \otimes U_i, \tag{4.12}$$

with $U_\gamma : \mathcal{A} \otimes \mathcal{M} \mapsto \mathcal{B} \otimes \mathcal{M}$ and where $\{|i\rangle\}$ is the basis of $\mathcal{M}$ and U_i are some unitaries on $\mathcal{A}$. The transformation on inputs reads

$$\begin{aligned} T^*(\omega_{[1,n]}) &= \mathrm{Tr}_{\mathcal{M}}[\gamma^*{}_{[1,n]}(\xi \otimes \omega_{[1,n]})] \\ &= \mathrm{Tr}_{\mathcal{M}}[(U_i{}^{\otimes n}\omega_{[1,n]}U_j^{\dagger\otimes n}) \otimes (|i\rangle\langle i|\xi|j\rangle\langle j|)] \\ &= \sum_{i=0}^{d_{\mathcal{M}}-1} \langle i|\xi|i\rangle U_i^{\otimes n}\omega_{[1,n]}U_i^{\dagger\otimes n}, \end{aligned} \tag{4.13}$$

and the state of ξ after nth collision will be

$$\langle i|\xi'|j\rangle = \langle i|\xi|j\rangle \mathrm{Tr}[U_i^{\otimes n}\omega_{[1,n]}U_j^{\dagger\otimes n}]. \tag{4.14}$$

Note that the diagonal elements of memory do not change because

$$\mathrm{Tr}[U_i^{\otimes n}\omega_{[1,n]}U_i^{\dagger\otimes n}] = \mathrm{Tr}\omega_{[1,n]} = 1. \tag{4.15}$$

Example 4.6 (Memoryless channel)**.** Let us have a collision model $\mathcal{E}_k^* : \mathcal{T}(\mathcal{M}) \otimes \mathcal{T}(\mathcal{A}_k) \mapsto \mathcal{T}(\mathcal{B}_k) \otimes \mathcal{T}(\mathcal{M})$ with memory living in Hilbert space $\mathcal{M}$, $\dim \mathcal{M} = d_{\mathcal{M}}$ where

$$\mathcal{E}_k^*(\xi \otimes \omega_k) = (\mathbb{I}_k \otimes P_k^*)\big(U_k(\xi \otimes \omega_k)U_k^\dagger\big), \tag{4.16}$$

for all $\xi \in \mathcal{T}(\mathcal{M})$, $\omega_k \in \mathcal{T}(\mathcal{A}_k)$ and some $U_k : \mathcal{M} \otimes \mathcal{A}_k \mapsto \mathcal{B}_k \otimes \mathcal{M}$ and P^* a completely depolarizing channel on $\mathcal{M}$ such that

$$P_k^*(m) = \xi_{k+1}, \tag{4.17}$$

for all $m \in \mathcal{T}(\mathcal{M})$ where ξ_k is a predefined sequence of states in $\mathcal{T}(\mathcal{M})$.

This interaction will produce a transformation on the input sequence

$$T^*(\omega_{[1,n]}) = (T_1^* \otimes \ldots \otimes T_n^*)(\omega_{[1,n]}), \tag{4.18}$$

where

$$T_k^*(\omega_k) = \mathrm{Tr}_{\mathcal{M}}[U_k(\xi_k \otimes \omega_k)U_k^\dagger] \tag{4.19}$$

are channels which act on every subsystem of input signal independently. We call such action memoryless since the behavior of channel is not affected by any of the inputs and the action is independent on every subsystem. If we would like to make such model pure, we would need to provide for each P_k^* a separate dilation space and that would result into infinite resources needed for memory system.

We see that every collision model then defines a causal process on input signals therefore a memory channel. It would be interesting to know whether all causal processes can be connected to an appropriate collision model. In [38] structure theorem is proved which tells us that such collision model exists for every translationally invariant causal process. A weaker structure theorem for finite sequences was also proved in [22] and this proof does not require translational invariance. In the next section we will present the structure theorem.

4.3 Structure theorem

We will present the structure theorem and its proof as it is stated in [38] with a minor change, we drop the assumption of translational invariance.

Theorem 4.7 (Structure theorem)**.** *Let $T : \mathfrak{Z} \mapsto \mathfrak{Z}$ be a causal channel. Ignore the outputs on its left half chain $\mathfrak{B}_-$. Then there exists a memory observable algebra $\mathfrak{M}$ and an initializing channel $R : \mathfrak{M} \mapsto \mathfrak{A}_-$ such that for all $n \in \mathbb{N}$*

$$T(\mathbb{I}_- \otimes b_{[1,n]}) = (R \otimes \mathbb{I}_{[1,n]})S_{[1,n]}(b_{[1,n]} \otimes \mathbb{I}^{\mathcal{M}}) \tag{4.20}$$

for all $b_{[1,n]} \in \mathfrak{B}_{[1,n]}$, where $S_{[1,n]}$ is the n-fold concatenation of a collision model (in Heisenberg picture) $S_i : \mathfrak{B}_i \otimes \mathfrak{M} \mapsto \mathfrak{M} \otimes \mathfrak{A}_i$

Proof. Let $\mathcal{H}_-$ be the Hilbert space associated with universal representation of the left half chain $\mathfrak{B}_-$. Further let $(\mathcal{K}, \pi, V)$ be the minimal Stinespring dilation for $T\mid_{\mathfrak{B}_-}$:

$$T(b_-) = V^\dagger \pi(b_-)V \quad \forall b_- \in \mathfrak{B}_- \tag{4.21}$$

with Stinespring isometry $V : \mathcal{H}_- \mapsto \mathcal{K}$. From Stinespring's representation (4.21) and the causality property (4.1) we can conclude that

$$\begin{aligned} V^\dagger \pi(b_- \otimes \mathbb{I}^{\mathcal{B}}_{[1,n]})V &= T(b_- \otimes \mathbb{I}^{\mathcal{B}}_{[1,n]}) := T(b_-) \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]} \\ &= (V^\dagger \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]})(\pi(b_-) \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]})(V \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]}) \end{aligned} \tag{4.22}$$

for all $b_- \in \mathfrak{B}_-$. Since V is minimal dilation for T so is $V \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]}$ for $T \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]}$. We know then, due to uniquiness of the minimal dilation, that there exists an isometry $W_{[1,n]} : \mathcal{K} \otimes \mathcal{H}_{[1,n]} \mapsto \mathcal{K}$ defined by

$$\begin{aligned} &W_{[1,n]}(\pi(b_-) \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]})(V \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]})\psi_- \otimes \psi_{[1,n]} \\ &\qquad := \pi(b_- \otimes \mathbb{I}^{\mathcal{B}}_{[1,n]})V\psi_- \otimes \psi_{[1,n]} \end{aligned} \tag{4.23}$$

for all $b_- \in \mathfrak{B}_-$, $\psi_- \in \mathcal{H}_-$ and $\psi_{[1,n]} \in \mathcal{H}_{[1,n]}$ such that

$$\pi(b_- \otimes \mathbb{I}^{\mathcal{B}}_{[1,n]})W_{[1,n]} = W_{[1,n]}(\pi(b_-) \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]}) \tag{4.24}$$

for all $b_- \in \mathfrak{B}_-$, and

$$W_{[1,n]}(V \otimes \mathbb{I}^{\mathcal{A}}_{[1,n]}) = V. \tag{4.25}$$

Now let $\mathfrak{M} := \pi'(\mathfrak{B}_-)$ be the commutant of the observable algebra $\mathfrak{B}_-$ and let $S_{[1,n]} : \mathfrak{B}_{[1,n]} \otimes \mathfrak{M} \mapsto \mathcal{L}(\mathcal{K}) \otimes \mathfrak{A}_{[1,n]}$ be defined by

$$S_{[1,n]}(b \otimes m) := W^\dagger_{[1,n]}\pi(b_-)mW_{[1,n]} \tag{4.26}$$

for all $b \in \mathfrak{B}_{[1,n]}$ and $m \in \mathfrak{M}$. The memory initializing channel $R : \mathfrak{M} \mapsto \mathfrak{A}_-$ is then given by

$$R(m) := V^\dagger mV \quad \forall m \in \mathfrak{M}. \tag{4.27}$$

In order to justify this choice we need to show that

$$S_{[1,n]}(\mathfrak{B}_{[1,n]} \otimes \mathfrak{M}) \subset \mathfrak{M} \otimes \mathfrak{A}_{[1,n]}. \tag{4.28}$$

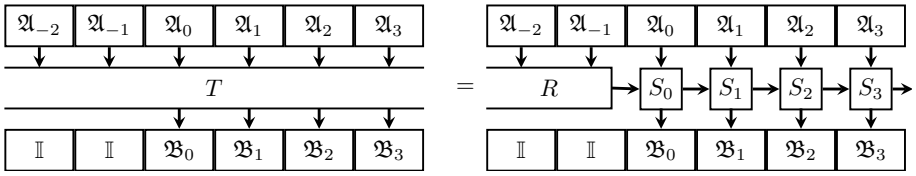

Fig. 4.3. Representation of structure theorem where causal process T is constructed as a collision model with memory initializer R. Note that the figure is in Heisenberg picture and therefore has to be read *against* the arrows which represent time.

Since $\mathfrak{B}_{[1,n]}$ and $\mathfrak{M}$ commute with $\mathfrak{B}_-$ we see from (4.24) that

$$\begin{aligned}
& S_{[1,n]}(b_{[1,n]}\otimes m)\cdot(\pi(b_-)\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]}) \\
= \;& W^{\dagger}_{[1,n]}\pi(\mathbb{I}^{\mathcal{B}}_{-}\otimes b_{[1,n]})mW_{[1,n]}(\pi(b_-)\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]}) \\
= \;& W^{\dagger}_{[1,n]}\pi(\mathbb{I}^{\mathcal{B}}_{-}\otimes b_{[1,n]})m\pi(b_-\otimes\mathbb{I}^{\mathcal{B}}_{[1,n]})W_{[1,n]} \\
= \;& W^{\dagger}_{[1,n]}\pi(\mathbb{I}^{\mathcal{B}}_{-}\otimes b_{[1,n]})\pi(b_-\otimes\mathbb{I}^{\mathcal{B}}_{[1,n]})mW_{[1,n]} \\
= \;& W^{\dagger}_{[1,n]}\pi(b_-\otimes\mathbb{I}^{\mathcal{B}}_{[1,n]})\pi(\mathbb{I}^{\mathcal{B}}_{-}\otimes b_{[1,n]})mW_{[1,n]} \\
= \;& (\pi(b_-)\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]})W^{\dagger}_{[1,n]}\pi(\mathbb{I}^{\mathcal{B}}_{-}\otimes b_{[1,n]})mW_{[1,n]} \\
= \;& (\pi(b_-)\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]})\cdot S_{[1,n]}(b_{[1,n]}\otimes m) \qquad (4.29)
\end{aligned}$$

for all $b_{[1,n]\in\mathfrak{B}_{[1,n]}}$, $b_-\in\mathfrak{B}_-$ and $m\in\mathfrak{M}$, implying that $S_{[1,n]}(b_{[1,n]}\otimes m)$ commutes with $\pi(b_-)\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]}$ from which (4.28) follows. The next thing we need to show is that $S_{[1,n]}$ has the right concatenation properties:

$$\begin{aligned}
R(m) &= (R\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]})S_{[1,n]}(\mathbb{I}^{\mathcal{B}}_{[1,n]}\otimes m) \quad \text{and} \\
T(b_{[1,n]}) &= (R\otimes\mathbb{I}^{\mathcal{A}}_{[1,n]})S_{[1,n]}(b_{[1,n]}\otimes\mathbb{I}^{\mathcal{M}}), \qquad (4.30)
\end{aligned}$$

however this directly follows from definitions of $S_{[1,n]}$, R and (4.25). For the translationally invariant case to obtain the result we set $S_i = S := S_{[1,1]}$. In the translationally non-invariant case we need to parcel the $S_{[1,n]}$ into separate S_i, what can be done via the weaker structure theorem presented in [22]. The proof again uses the uniqueness of minimal Stinespring's dilation and the same tools used in this proof. The structure theorem is visualized on figure 4.3. □

For every causal channel there exists a collision model $S_{[1,n]}$ and appropriate channel initializer R which has an equivalent input output behavior. This is an important result which proves that collision models are a highly general concept.

5 Memory

In this chapter we will stress out the nature and strength of the memory effects in memory channels introduced in previous chapter. We will define forgetful memory channels and strictly forgetful memory channels [70]. Finally we will address repeatability of channels in the quantum memory channel setting and stroboscopic simulation of evolution of open systems.

5.1 Forgetfulness

In the proof of structure theorem in Section 4.3 we needed to establish a channel initializer R which described the influence of the inputs from the remote past on the memory system. This influence is in generally nontrivial however for certain class of memory channels it becomes irrelevant in the long run. These memory channels should have the property that information that is localized far in the past will not affect the behavior of channel in present. Channels where the initializer R will become irrelevant are forgetful memory channels. In this chapter we will assume translational invariance, however the generalization to non invariant case is straightforward.

Definition 5.1 (Forgetful memory channel, Schrödinger picture)**.** Let $S^* : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}) \mapsto \mathcal{T}(\mathcal{B} \otimes \mathcal{M})$ be a translationally invariant collision model of a quantum memory channel. Suppose $\rho_1, \rho_2 \in \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_{[1,n]})$ such that $\mathrm{Tr}_{\mathcal{M}}(\rho_1) = \mathrm{Tr}_{\mathcal{M}}(\rho_2)$. Then $S^*_{[1,n]}$ is forgetful iff

$$\lim_{n\to\infty} \| \mathrm{Tr}_{\mathcal{B}_{[1,n]}}[S^*_{[1,n]}(\rho_1 - \rho_2)] \|_{\mathrm{tr}} = 0, \tag{5.1}$$

for all such ρ_1, ρ_2 and where $\| A \|_{\mathrm{tr}} = \mathrm{Tr}\sqrt{A^\dagger A}$ is the trace-norm.

This only means that the state of memory after many collisions depends only on the state of inputs $\mathrm{Tr}_{\mathcal{M}}(\rho_i)$ and not on the initial state of the memory $\mathrm{Tr}_{\mathcal{B}_{[1,n]}}(\rho_i)$. The definition in Heisenberg picture makes this only more evident:

Definition 5.2 (Forgetful memory channel, Heisenberg picture)**.** Let $S : \mathfrak{A} \otimes \mathfrak{M} \mapsto \mathfrak{M} \otimes \mathfrak{A}$ be a translationally invariant collision model of a quantum memory channel. Then let $\hat{S}_{[1,n]} : \mathfrak{M} \mapsto \mathfrak{M} \otimes \mathfrak{A}_{[1,n]}$ be its concatenation where the outputs are ignored $\hat{S}_{[1,n]}(m) := S_{[1,n]}(\mathbb{I}_{[1,n]} \otimes m)$ for all $m \in \mathfrak{M}$. Then S is forgetful iff there exists a sequence of quantum channels $\tilde{S}_{[1,n]} : \mathfrak{M} \mapsto \mathfrak{A}_{[1,n]}$ such that

$$\lim_{n\to\infty} \| \hat{S}_{[1,n]} - \mathbb{I}_{\mathcal{M}} \otimes \tilde{S}_{[1,n]} \|_{\mathrm{cb}} = 0. \tag{5.2}$$

It can be proven that forgetful channels are dense in the set of all memory channels, see [38]. The main idea is to add a infinitesimally small amount of white noise on the memory system and the channel will become forgetful. Such noise is present usually in real life applications hence generally the channels will be forgetful. For such channels coding theorems can be proven and channel capacities calculated. However it is rather hard to prove that some memory channel is forgetful.

Channels for which the limit in (5.1) and (5.2) is attained for finite $n < \infty$ are called *strictly forgetful* memory channels or channels with finite depth of memory, where the depth of memory is the number of times you have to use it to completely forget the state of memory.

Remark 5.3. In the case of pure memory channels which are strictly forgetful with depth of memory δ, we can observe that the dimension of memory system and the dimension of the system of δ consequent input cells have to be co-divisible in order to be strictly forgetful. Let $m \in \mathcal{L}(\mathcal{M})$ have $\dim \mathcal{M} = d_{\mathcal{M}}$ distinct eigenvalues. Then the operator $\hat{S}_{[1,\delta]}(m) = S_{[1,\delta]}(m \otimes \mathbb{I}_{[1,\delta]})$ has the same eigenvalues with multiplicity $\dim \mathcal{A}_{[1,\delta]} = d_\delta$ because $S_{[1,\delta]}$ is unitary. Since we know that the channel has finite depth, we can write $S_{[1,\delta]}(m \otimes \mathbb{I}_{[1,\delta]}) = \mathbb{I}_{\mathcal{M}} \otimes \tilde{S}_{[1,n]}(m)$. The operator $\mathbb{I}_{\mathcal{M}} \otimes \tilde{S}_{[1,n]}(m)$ has d_δ eigenvalues with multiplicity $d_{\mathcal{M}}$ but also it has $d_{\mathcal{M}}$ eigenvalues with multiplicity d_δ. Hence $d_{\mathcal{M}}$ and d_δ have to be co-divisible.

Forgetfulness describes the ability of the memory channel to completely reset the state of memory. If we are interested only in the input-output relations, this might be a too strict or counterintuitive approach. Memory channels where there is no interaction with memory, hence no memory effects are present, but the memory system is still included in the description of model, will be non-forgetful. Eventually any forgetful channel can become non-forgetful if we are "creative" enough to include some noninteracting memory into the model. We might relax the definition of forgetfulness in following way.

Definition 5.4 (I-O forgetfulness)**.** Let $S^* : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}) \mapsto \mathcal{T}(\mathcal{B} \otimes \mathcal{M})$ be a translationally invariant collision model of a quantum memory channel. Suppose $\rho_1, \rho_2 \in \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_{[1,n+m]})$ such that $\mathrm{Tr}_{\mathcal{M}}(\rho_1) = \mathrm{Tr}_{\mathcal{M}}(\rho_2) = \omega_{[1,n+m]}$ for some finite $m \in \mathbb{N}$. Then $S^*_{[1,n+m]}$ is I-O forgetful iff

$$\lim_{n \mapsto \infty} \| \, T_1^*(\omega_{[n,n+m]}) - T_2^*(\omega_{[n,n+m]}) \, \|_{\mathrm{tr}} = 0 \tag{5.3}$$

for every finite $m \in \mathbb{N}$ where $T_j^*(\omega_{[n,n+m]}) = \mathrm{Tr}_{\mathcal{M} \otimes \mathcal{B}_{[1,n]}}[S^*_{[1,n+m]}(\rho_j)]$. We can speak about *strict* I-O forgetfulness, if this limit is attained for every $n \geq \delta$, where $0 \leq \delta < \infty$.

I-O forgetfulness is weaker than forgetfulness. The set of all forgetful channels S_{ff} is a subset of I-O forgetful channels $S_{\mathrm{I-O}}$. Following example is here to show that the difference between I-O forgetfulness and normal forgetfulness is not only caused by the "noninteracting ancilla".

The memory channel from Example 4.5 is not I-O forgetful, since the transformation explicitly depends on diagonal elements of memory state. However we can add some noise to the memory system to make it I-O forgetful but not forgetful.

Example 5.5 (I-O Forgetful but not forgetful)**.** Let us have a translationally invariant qubit-qubit collision model with collision $S^* : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}) \mapsto \mathcal{T}(\mathcal{B} \otimes \mathcal{M})$

$$S^*(m \otimes \omega) = E^* \otimes \mathbb{I}\big(U_{\mathrm{cnot}}(m \otimes \omega){U_{\mathrm{cnot}}}^\dagger\big) \tag{5.4}$$

where E^* is the contraction to the x-axis from Example 2.16 and U_{cnot} is also defined there. It is easy to see that this channel is not forgetful. The states $|+\rangle_{\mathcal{M}} \otimes |+\rangle$ and $|-\rangle_{\mathcal{M}} \otimes |+\rangle$ are fixed points of channel S^*. Thus the x component of state in $\mathcal{M}$ will stay the same if the input sequence is $\omega_{[1,n]} = |+\ldots+\rangle$.

It is a bit tedious to prove that it is I-O forgetful, the idea is that U_{cnot} takes into account only the diagonal elements of memory and that E^* resets them to $\frac{1}{2}$ irrespective of the original state, hence the effect of initial memory state will be zero after first use and all subsequent uses will be independent of the initial state.

Here is the detailed proof. Let's write the joint state of memory and inputs $\rho \in \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_{1,n})$ as

$$\rho = \sum_{ij} |i\rangle_{\mathcal{M}}\langle j| \otimes \Omega_{ij}. \tag{5.5}$$

This way the state of input sequence is $\mathrm{Tr}_{\mathcal{M}}\rho = \Omega_{00} + \Omega_{11} = \omega_{[1,n]}$. The first collision will transform ρ into

$$\begin{aligned} S_1^*(\rho) &= E^* \otimes \mathbb{I}_{[1,n]}[(U_{\mathrm{cnot}} \otimes \mathbb{I}_{[2,n]})\rho({U_{\mathrm{cnot}}}^\dagger \otimes \mathbb{I}_{[2,n]})] \\ &= E^*(|0\rangle_{\mathcal{M}}\langle 0|) \otimes \Omega_{00} + E^*(|0\rangle_{\mathcal{M}}\langle 1|) \otimes (\Omega_{01}X_1) \\ &\quad + E^*(|1\rangle_{\mathcal{M}}\langle 0|) \otimes (X_1\Omega_{10}) + E^*(|1\rangle_{\mathcal{M}}\langle 1|) \otimes (X_1\Omega_{11}X_1) \\ &= \frac{1}{2}[\mathbb{I}_{\mathcal{M}} \otimes (\Omega_{00} + X_1\Omega_{11}X_1) + X_{\mathcal{M}} \otimes (\Omega_{01}X_1 + X_1\Omega_{10}), \end{aligned} \tag{5.6}$$

because $E^*(|i\rangle\langle i|) = \frac{1}{2}\mathbb{I}_{\mathcal{M}}$ and $E^*(|0\rangle\langle 1|) = E^*(|1\rangle\langle 0|) = \frac{1}{2}X_{\mathcal{M}}$. The output state clearly depends on the initial memory state. If the initial memory state was $|0\rangle_{\mathcal{M}}\langle 0|$, which would mean that $\Omega_{11} = 0$, the transformation would read

$$T^*(\omega_{[1,n]}) = \omega_{[1,n]}, \tag{5.7}$$

however if the initial state was $|1\rangle_{\mathcal{M}}\langle 1|$, then $\Omega_{00} = 0$, the transformation would be

$$T^*(\omega_{[1,n]}) = X_1\omega_{[1,n]}X_1. \tag{5.8}$$

As we can already see the Ω_{01} and Ω_{10} will never enter the channel T^* because they will be always attached to a traceless operator on memory system, we can ignore them ($\Omega_{01} = \Omega_{10} = 0$). Now if we look on the composite state of memory and inputs sequence except first input we find that

$$\begin{aligned} \mathrm{Tr}_{\mathcal{B}_1}[S_1^*(\rho)] &= \frac{1}{2}\mathrm{Tr}_{\mathcal{B}_1}[\mathbb{I}_{\mathcal{M}} \otimes (\Omega_{00} + X_1\Omega_{11}X_1)] \\ &= \frac{1}{2}\mathbb{I}_{\mathcal{M}} \otimes (\mathrm{Tr}_{\mathcal{B}_1}[\Omega_{00}] + \mathrm{Tr}_{\mathcal{B}_1}[X_1\Omega_{11}X_1]) \\ &= \frac{1}{2}\mathbb{I}_{\mathcal{M}} \otimes (\mathrm{Tr}_{\mathcal{B}_1}[\Omega_{00} + \Omega_{11}]), \end{aligned} \tag{5.9}$$

where the last equality is due to invariance of partial trace under unitary conjugation on subsystem which is being traced over, a consequence of invariance of trace under cyclic permutations. This state is independent of the initial memory and hence the initial memory state will have no effects after the first collision. Thus the whole memory channel is I-O forgetful, even strictly I-O forgetful with $\delta = 1$.

Lemma 5.6 (Equivalence class of collision models)**.** *Let $S_k^* : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_k) \mapsto \mathcal{T}(\mathcal{B}_k \otimes \mathcal{M})$ be a collision model of a quantum memory channel. Then memory channel with collisions*

$$\hat{S}_k^* = (\mathbb{I}_k \otimes W^\dagger)S_k^*(W \otimes \mathbb{I}_k), \tag{5.10}$$

where $W : \mathcal{M} \mapsto \mathcal{M}$ is an arbitrary unitary and $\mathbb{I}_k$ is the identity operation on k-th subsystem, has the same I-O relation as the former memory channel.

Proof. This directly follows from the concatenation properties of the collision model. The concatenation of $\hat{S}_k^*$ will be

$$\begin{aligned}
&\hat{S}^*_{[1,n]} = (\mathbb{I}_{[1,n]} \otimes W^\dagger)(\mathbb{I}_{[1,n-1]} \otimes S_n^*)(\mathbb{I}_{[1,n-1]} \otimes W \otimes \mathbb{I}_n) \\
&\cdots (\mathbb{I}_{[1,2]} \otimes W^\dagger \otimes \mathbb{I}_{[3,n]})(\mathbb{I}_1 \otimes S_2^* \otimes \mathbb{I}_{[3,n]})(\mathbb{I}_1 \otimes W \otimes \mathbb{I}_{[2,n]}) \\
&(\mathbb{I}_1 \otimes W^\dagger \otimes \mathbb{I}_{[2,n]})(S_1^* \otimes \mathbb{I}_{[2,n]})(W \otimes \mathbb{I}_{[1,n]}) = \\
&(\mathbb{I}_{[1,n]} \otimes W^\dagger)(\mathbb{I}_{[1,n-1]} \otimes S_n^*) \cdots (\mathbb{I}_1 \otimes S_2^* \otimes \mathbb{I}_{[3,n]})(S_1^* \otimes \mathbb{I}_{[2,n]})(W \otimes \mathbb{I}_{[1,n]}).
\end{aligned} \tag{5.11}$$

Since the partial trace is invariant under unitary conjugations of the system being traced over, the trace over memory system will be the same as for plain S_k^* without the unitary W. □

5.2 Finite depth memory channels

We will add another slightly different notion of forgetfulness. The notion of memory depth was already used for strictly forgetful channels, where it marks the maximal number of uses of the memory channel needed to forget. Now we will formalize memory channels with finite depth as those possessing this property, but we will require only I-O strict forgetfulness and require it only for factorized input sequences. This will also implicitly bind us to the Schrödinger picture, where the factorization of input sequence is more natural.

Definition 5.7 (Finite depth)**.** Let $S_k^* : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_k) \mapsto \mathcal{T}(\mathcal{B}_k \otimes \mathcal{M})$ be a collision model of a quantum memory channel. Let the input sequence restrict to $\omega_{[1,n]} = \bigotimes_{k=1}^n \omega_k$. Assume two different initial states of memory $\xi_1, \xi_2 \in \mathcal{T}(\mathcal{M})$ Then $S^*_{[1,n]}$ has finite depth δ if

$$\| \, T_1^*(\omega_{[\delta+1,n]}) - T_2^*(\omega_{[\delta+1,n]}) \, \|_{\mathrm{tr}} = 0 \tag{5.12}$$

for every finite $n > \delta$ where $T_j^*(\omega_{[\delta+1,n]}) = \mathrm{Tr}_{\mathcal{M} \otimes \mathcal{B}_{[1,\delta]}}[S^*_{[1,n]}(\xi_j \otimes \omega_{[1,n]})]$.

Strictly I-O forgetful, hence also strictly forgetful, memory channels have automatically finite depth. Note that finite depth is defined only on factorized inputs. However, next lemma shows that this can be trivialy shifted to any state of inputs and memory, hence finite depth also implies strict I-O forgetfulness.

Lemma 5.8. *If a memory channel has finite depth, then it is strictly I-O forgetful.*

Proof. Let us have a collision model ${S^*}_{[1,n]} : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_{[1,n]}) \mapsto \mathcal{T}(\mathcal{B}_{[1,n]} \otimes \mathcal{M})$ such that it has finite depth δ. Then

$$\mathrm{Tr}[S^*_{[1,\delta+1+m]}\big((\xi_1 - \xi_2) \otimes \omega_1 \otimes \cdots \otimes \omega_{\delta+1+m}\big)(\mathbb{I}_{[1,\delta]} \otimes b_{[\delta+1,\delta+1+m]} \otimes \mathbb{I}_{\mathcal{M}})] = 0 \tag{5.13}$$

holds for every $\xi_1, \xi_2 \in \mathcal{S}(\mathcal{M})$, $\omega_k \in \mathcal{S}(\mathcal{A}_k)$, $b_{[\delta+1,\delta+1+m]} \in \mathfrak{B}_{[\delta+1,\delta+1+m]}$ and $m \in \mathbb{N}$. Any $\rho_1 \in \mathcal{S}(\mathcal{M} \otimes \mathcal{A}_{[1,\delta+1+m]})$ can be written as a (not necessary convex or positive) sum of factorized sequences $\xi_1 \otimes \omega_1 \otimes \cdots \otimes \omega_{\delta+1+m}$ since they provide an over-complete operator basis. Due to linearity it has to be true that

$$\mathrm{Tr}[S^*_{[1,\delta+1+m]}\big(\rho_1 - \rho_2\big)(\mathbb{I}_{[1,n]} \otimes b_{[\delta+1,\delta+1+m]} \otimes \mathbb{I}_{\mathcal{M}})] = 0 \tag{5.14}$$

for any $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{M} \otimes \mathcal{A}_{[1,\delta+1+m]})$ such that $\mathrm{Tr}_{\mathcal{M}}\rho_1 = \mathrm{Tr}_{\mathcal{M}}\rho_2$. □

This is a very interesting result since it greatly simplifies the conditions for for strict I-O forgetfulness. It is enough to check only all factorized sequences.

As we have seen in previous section, the memory might be composed of some part which does not affect the transformation on inputs, we will call this the irrelevant part of memory. Let us fix a collision model $S_k^* : \mathcal{T}(\mathcal{M} \otimes \mathcal{A}_k) \mapsto \mathcal{T}(\mathcal{B}_k \otimes \mathcal{M})$ with the initial state of memory $\xi = 1/\sqrt{d_{\mathcal{M}}}(\tau_0 + \vec{m}\vec{\tau})$ and constrain the input sequence to factorized states, $\omega_{[1,n]} = \bigotimes_{k=1}^{n} \omega_k$. Then we can define the irrelevant subspace.

Definition 5.9 (Irrelevant degrees of freedom)**.** A traceless operator $\varsigma = \vec{s}\vec{\tau}$ is called *irrelevant* if for every $n \in \mathbb{N}$

$$\mathrm{Tr}_{\mathcal{M}}[S^*{}_{[1,n]}(\varsigma \otimes \omega_{[1,n]})] = O \tag{5.15}$$

holds. The space spanned by all all such ς is called the *irrelevant subspace* of memory and we will denote it by I_{S^*}.

One can see that any two initial memory states ξ_1, ξ_2 such that $\xi_1 - \xi_2 \in I_{S^*}$ have the same I-O relation, ie. produce the same channel $T^*(\omega_{[1,n]})$. Note that the space I_{S^*} can be only spanned by traceless operators because for any operator with nonzero trace the equation (5.15) does not hold because of trace preserving property of S^*.

Since the input is factorized, the memory evolves through composition of series of concurrent channels $C_k^* : \mathcal{T}(\mathcal{M}) \mapsto \mathcal{T}(\mathcal{M})$

$$\xi \mapsto C_n^* \cdots C_1^*(\xi). \tag{5.16}$$

The finite depth property can be then expressed as

$$C_\delta^* \cdots C_1^*(\xi_1 - \xi_2) \in I_{S^*} \tag{5.17}$$

for all $\xi_1, \xi_2 \in \mathcal{S}(\mathcal{M})$.

For every memory channel with finite depth δ, inputs separated by δ uses evolve under factorized transformation

$$T^*(\omega_k \otimes \omega_{k+\delta+1}) = T^*(\omega_k) \otimes T^*(\omega_{k+\delta+1}). \tag{5.18}$$

This has to be so because neither initial state of memory nor ω_k can have an effect on the transformation.

In non-translational invariant case the depth of the memory channel is independent of the size of the memory. One can simply think of a memory channel with all collisions trivial except every nth which would be a swap. Such channel has depth n and this can be an arbitrarily large number.

In the translational invariant case the maximum number of relevant degrees of freedom is $d_{\mathcal{M}}^2 - 1$, where $d_{\mathcal{M}}$ is the dimension of memory Hilbert space. Thus this should be also the bound on the memory depth, since at every collision one should erase at least one relevant degree of freedom. Indeed this is the case and the depth can be as high as $d_{\mathcal{M}^2-1}$ in the translational invariant case. This was proven in [1].

Theorem 5.10 (Bound on the depth)**.** *Let us have a strictly forgetful and translational invariant memory channel* $S : \mathfrak{M}\otimes\mathfrak{A} \mapsto \mathfrak{A}\otimes\mathfrak{M}$ *with depth* δ *and* $\dim(\mathcal{M}) = d_{\mathcal{M}} < \infty$*. Then* $\delta \leq d^2_{\mathcal{M}} - 1$.

Proof. We work now in Heisenberg picture. The condition of strict forgetfulness reads

$$\| \hat{S}_{[1,n]} - \mathbb{I}_{\mathcal{M}} \otimes \tilde{S}_{[1,n]} \|_{\mathrm{cb}} = 0, \quad \forall n \geq \delta \tag{5.19}$$

where $\hat{S}_{[1,n]}$ and $\tilde{S}_{[1,n]}$ are as in (5.2). Now define

$$\mathcal{V}_k = \{m \in \mathfrak{M} : \tilde{S}_{[\delta-k,\delta]}(m) \in \mathbb{I}_{\mathcal{M}}\mathfrak{A}_{[\delta-k,\delta]}\} \tag{5.20}$$

which is the subset of all $m \in \mathfrak{M}$ such that after k collisions m is fully localized on inputs. Since the map is linear these sets are subspaces of $\mathfrak{M}$. We define the

$$\mathcal{V}_0 = \mathbb{I}_{\mathcal{M}}, \tag{5.21}$$

and by definition we have

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \ldots \subset \mathcal{V}_\delta = \mathfrak{M}. \tag{5.22}$$

However these inclusions must be strict. Suppose that $\mathcal{V}_k = \mathcal{V}_{k-1}$. We can decompose $\tilde{S}_{[1,\delta]}(m)$ as

$$(\tilde{S}_{[1,\delta-k]} \otimes \mathbb{I}_{\mathcal{A}_{[\delta-k,\delta]}})\tilde{S}_{[\delta-k,\delta]}(m) \subset \mathcal{V}_0 \otimes \mathfrak{A}_{[1,\delta]}. \tag{5.23}$$

From this we see that

$$\tilde{S}_{[\delta-k,\delta]}(m) \subset \mathcal{V}_k \otimes \mathfrak{A}_{[\delta-k,\delta]}. \tag{5.24}$$

By assumption this is equal to $\mathcal{V}_{k-1} \otimes \mathfrak{A}_{[\delta-k,\delta]}$. Hence the depth of the channel can be reduced to $\delta - 1$ because

$$\big((\tilde{S}_{[2,\delta-k]} \otimes \mathbb{I}_{\mathcal{A}_{[\delta-k,\delta]}})(\mathcal{V}_{k-1})\big) \otimes \mathfrak{A}_{[\delta-k,\delta]} \subset \mathcal{V}_0 \otimes \mathfrak{A}_{[1,\delta]}. \tag{5.25}$$

The longest chain of strict inclusions of subspaces of a d^2 dimensional space is $d^2 - 1$ from which the claim follows. □

In the same paper they also prove that this bound is optimal, thus there exist such strictly forgetful memory channels with depth $\delta = d^2_{\mathcal{M}} - 1$.

5.2.1 Qubit-qubit case study

In this case study, we will consider only pure, translationally invariant memory channels with qubit memory interacting with qubit subsystem, $\dim \mathcal{M} = \dim \mathcal{A} = \dim \mathcal{B} = 2$. The collision will be described by an unitary interaction U. The U can be parametrized in following way [35]:

$$U = (W_2 \otimes V_2)D(W_1 \otimes V_1), \tag{5.26}$$

where $W_i : \mathcal{M} \mapsto \mathcal{M}$ and $V_i : \mathcal{A} \mapsto \mathcal{A}$ are unitaries on appropriate Hilbert spaces and D is 2-qubit unitary of a special form

$$D = e^{i\sum_{k=1}^{3} \frac{1}{2}\alpha_k \sigma_k \otimes \sigma_k}, \tag{5.27}$$

where $\sigma_1 = X$, $\sigma_2 = Y$ and $\sigma_3 = Z$ are the Pauli matrices and $\alpha_k \in \mathbb{R}$. Since $\sigma_k \otimes \sigma_k$ commute the D can be written as

$$D = \prod_{k=1}^{3} (\cos(\alpha_k/2)\mathbb{I} + i \sin(\alpha_k/2)\sigma_k \otimes \sigma_k). \tag{5.28}$$

This is a very nice parametrization for our task since it separates the local transformations from the interacting part. We can substitute W_1 for identity for our purposes, because of the I-O invariance of collision models from Lemma 5.6 by suitably choosing $W = W_1^\dagger$. Since W_2 can be any unitary this will not change the description.

Lets fix initial memory in state $\xi \in \mathcal{S}(\mathcal{M}) =: \xi_1$. We restrict our inputs to be of factorized form, $\omega_{[1,n]} = \bigotimes_{k=1}^{n} \omega_k$ uncorrelated to ξ. Firstly we concentrate our attention to identifying the relevant and irrelevant subspace. The transformation on first input will be

$$\begin{aligned} T^*(\omega_1) &= \mathrm{Tr}_{\mathcal{M}}[U(\xi_1 \otimes \omega_1)U^\dagger] = V_2 \mathrm{Tr}_{\mathcal{M}}[D\big(\xi_1 \otimes (V_1\omega_1 V_1^\dagger)\big)D^\dagger]V_2^\dagger \\ &=: V_2 \hat{T}^*(V_1\omega_1 V_1^\dagger)]V_2^\dagger. \end{aligned} \tag{5.29}$$

The irrelevant subspace is fully determined by $\hat{T}^*$ since only there interaction with memory system occurs. Lets fix the operator basis for qubit as usual $\tau = 1/\sqrt{2}(\mathbb{I}, X, Y, Z)$. Assume that $\xi_1 = 1/\sqrt{2}(\tau_0 + \vec{m}\vec{\tau})$. Then channel $\hat{T}^*$ is

$$\hat{T}^*(\omega_1) = \mathrm{Tr}_{\mathcal{M}}[D(\xi_1 \otimes \omega_1)D^\dagger], \tag{5.30}$$

what in vector representation looks $A(\hat{T}^*)_{ij} = \mathrm{Tr}[D(\xi_1 \otimes \tau_j)D^\dagger(\mathbb{I} \otimes \tau_i)]$:

$$A(\hat{T}^*) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ m_1 s_2 s_3 & c_2 c_3 & m_3 c_2 s_3 & -m_2 s_2 c_3 \\ m_2 s_1 s_3 & -m_3 c_1 s_3 & c_1 c_3 & m_1 s_1 c_3 \\ m_3 s_1 s_2 & m_2 c_1 s_2 & -m_1 s_1 c_2 & c_1 c_2 \end{pmatrix}, \tag{5.31}$$

where $c_i = \cos\alpha_i$ and $s_i = \sin\alpha_i$. We see that for any nonzero operator $\varsigma = \vec{m}\vec{\tau}$ to be irrelevant we need $s_i = 0$ for all $i = 1, 2, 3$. This implies that $\alpha_i = k_i\pi$ where $k_i \in \mathbb{N}$. However if this holds then D is a product of two equal Pauli matrices, i.e. no interaction at all. We conclude that the irrelevant subspace is trivial unless the whole interaction is of factorized form. Then the irrelevant subspace is maximal, meaning that the relevant subspace is spanned only by identity.

Now we will move on to the concurrent channel. Lets denote

$$\xi_{k+1} := C_k^* \cdots C_1^*(\xi_1) = C_k^*(\xi_k) \tag{5.32}$$

the state of memory after k-th collision, where C_k^* are the concurrent channels

$$C_k^*(\xi_k) = \mathrm{Tr}_k[U(\xi_k \otimes \omega_k)U^\dagger]. \tag{5.33}$$

After applying the parametrization (5.26) we get that

$$C_k^*(\xi_k) = W_2 \mathrm{Tr}_k[D\big(\xi_k \otimes (V_1\omega_k V_1^\dagger)\big)D^\dagger]W_2^\dagger, \tag{5.34}$$

where we see that V_2 doesn't affect the transformation on memory. Since the irrelevant subspace is trivial we require

$$C_\delta^* \cdots C_1^*(\varsigma) = O \tag{5.35}$$

for any $\varsigma = \vec{m}\vec{\tau}$ a memory channel of depth δ. In vector representation this means that if we define

$$A(C_k^*) = \begin{pmatrix} 1 & 0 \\ \vec{c}_k & \widetilde{C}_k \end{pmatrix} \tag{5.36}$$

then

$$\vec{m} \mapsto \widetilde{C}_\delta \cdots \widetilde{C}_1 \vec{m} + \sum_{k=1}^{\delta-1} \widetilde{C}_\delta \cdots \widetilde{C}_{k+1} \vec{c}_k + \vec{c}_\delta. \tag{5.37}$$

If we require the equation (5.35) to hold we need that

$$\widetilde{C}_\delta \cdots \widetilde{C}_1 = O \tag{5.38}$$

because only then the right hand side of (5.37) will be independent of $\vec{m}$. A necessary condition for that is that any $\widetilde{C}_k$ has to be singular, since we can have input sequences $\omega_{[1,n]} = \omega^{\otimes n}$ with arbitrary ω. Any $\widetilde{C}_k$ looks in vector representation as

$$\widetilde{C}_k = R_2 \begin{pmatrix} c_2c_3 & w_{k,3}c_2s_3 & -w_{k,2}s_2c_3 \\ -w_{k,3}c_1s_3 & c_1c_3 & w_{k,1}s_1c_3 \\ w_{k,2}c_1s_2 & -w_{k,1}s_1c_2 & c_1c_2 \end{pmatrix}, \tag{5.39}$$

where we have used that $V_1\omega_k V_1^\dagger = 1/\sqrt{2}(\mathbb{I} + \vec{w}_k\vec{\tau})$ and $R_2 = A(W_2)$ is the rotation due to unitary W_2. Since $\det(\widetilde{C}_k) = (c_1c_2c_3)^2 + w_{k,1}(s_1c_2c_3)^2 + w_{k,2}(c_1s_2c_3)^2 + w_{k,3}(c_1c_2s_3)^2$, this matrix is singular for arbitrary $\vec{w}_k$ only when $c_j = 0$ for two different j. Lets say that $c_1 = c_2 = 0$ and $c_3 \neq 0$, then

$$\begin{aligned} \widetilde{C}_k &= R_2 \begin{pmatrix} 0 & 0 & \pm w_{k,2}c_3 \\ 0 & 0 & \pm w_{k,1}c_3 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \pm \begin{pmatrix} 0 & 0 & w_{k,1}c_3R_{2,12} \pm w_{k,2}c_3R_{2,11} \\ 0 & 0 & w_{k,1}c_3R_{2,22} \pm w_{k,2}c_3R_{2,21} \\ 0 & 0 & w_{k,1}c_3R_{2,32} \pm w_{k,2}c_3R_{2,31} \end{pmatrix}. \end{aligned} \tag{5.40}$$

Under further investigation we find that condition (5.38) can hold for arbitrary factorized input sequence only if $R_{2,32} = R_{2,31} = 0$. Since R_2 is rotation this also implies that $R_{2,33} = \pm 1$ and

$$R_2 = \begin{pmatrix} S & 0 \\ 0 & \pm 1 \end{pmatrix}, \tag{5.41}$$

where S is rotation in the xy-plane. Such unitaries commute with Z thus

$$W_2 = e^{\beta Z}. \tag{5.42}$$

Furthermore the depth of such memory channel is necessary $\delta = 2$ since we can check that

$$\widetilde{C}_{k+1}\widetilde{C}_k = O, \tag{5.43}$$

for any factorized input sequence. So if $c_1 = c_2 = 0$ and $c_3 \neq 0$ and $W_2^z = e^{\beta Z}$ then any $U_{\delta=2}^z$ of form

$$U_{\delta=2}^z = (W_2^z \otimes V_2)D(\mathbb{I}_{\mathcal{M}} \otimes V_1), \tag{5.44}$$

with arbitrary V_k defines a collision model with memory depth $\delta = 2$: $S_k^*(\xi \otimes \omega_k) = FU_{\delta=2}(\xi \otimes \omega_k)U_{\delta=2}^\dagger F$. Similarly you will get the same thing for $U_{\delta=2}^y$ with $c_1 = c_3 = 0$ and $c_2 \neq 0$ and $W_2^y = e^{\beta Y}$ and $U_{\delta=2}^x$ with $c_2 = c_3 = 0$ and $c_1 \neq 0$ and $W_2^x = e^{\beta X}$. Furthermore it is easy to check that if an two qubit unitary U corresponds to a memory channel with depth of memory 2 then also $U^\dagger$ has depth of memory 2. Lets say that U is of the $U_{\delta=2}^z$ type, then

$$U^\dagger = (\mathbb{I}_{\mathcal{M}} \otimes V_1^\dagger)D^\dagger(W_2^{z\dagger} \otimes V_2^\dagger). \tag{5.45}$$

Since the I-O relation is invariant under conjugation of memory, we are able to move the inverse of W_2 back to the left side with suitable conjugation. Also the inverses of V_i can be arbitrary and hence are of no interest now, so we can omit the daggers on them. Thus

$$U^\dagger = (W_2^{z\dagger} \otimes V_1)D^\dagger(\mathbb{I}_{\mathcal{M}} \otimes V_2). \tag{5.46}$$

We can also ignore the dagger on W_2^z since it does not change the type of the unitary, it only changes the sign in front of β. If D is such that $c_1 = c_2 = 0$ then also $D^\dagger$ will fulfill this condition, because the dagger operation only introduces sign changes in front of α_i and this won't affect the cosines in condition, thus

$$U^\dagger = (W_2^z \otimes V_1)D(\mathbb{I}_{\mathcal{M}} \otimes V_2), \tag{5.47}$$

which is of depth 2 again.

We will also get finite depth if we set $c_1 = c_2 = c_3 = 0$, then $\widetilde{C}_k = O$ automatically and $\delta = 1$. It follows then that

$$U_{\delta=1} = (W_2 \otimes V_2)F(W_1 \otimes V_1), \tag{5.48}$$

with arbitrary W_k and V_k where F is the two qubit swap unitary. Trivially since the local unitaries are arbitrary and F self-adjoint, $U_{\delta=1}^\dagger$ has also memory depth 1.

If we look onto classical case we have discrete set of possible interactions - permutations of a set of four elements, which can be represented by 24, 4×4 permutation matrices, and states are diagonal, hence described by only one traceless parameter $\xi = 1/2(\mathbb{I} + pZ)$. It can be easily checked that these interactions can be grouped to three groups. First group would consist of factorized interactions - independent evolution of memory and input. This group has $\delta = 0$. Second group would be the swap like interactions, with $\delta = 1$ and third group are control unitary interactions (see Example 4.5), where the memory is control and the input target or the other way round. Such interactions have not finite depth. Thus the quantum case is slightly richer.

But the bound given by theorem 5.10 suggests that there could also exist a pure memory channel with depth 3. But such memory channel does not exist. For the case of pure memory channels the inclusion of subspaces in the proof of the aforementioned theorem will turn into inclusion of algebras, because if $X, Y \in \mathcal{V}_k$ then by unitarity of the evolution also $XY \in \mathcal{V}_k$. Thus $\mathcal{V}_k$ has to be closed under multiplication of its elements. In case of depth 3 there would need to be a chain of nesting subalgebras of qubit algebra such that

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \mathcal{V}_3 \tag{5.49}$$

where $\mathcal{V}_0$ is the trivial subalgebra including only the identity and $\mathcal{V}_3$ is the full qubit operator algebra. Since the inclusions are strict, the dimension of $\mathcal{V}_k$ has to be $k+1$. However there is no 3-dimensional subalgebra of a qubit operator algebra. Thus the longest possible chain of inclusions is 2 and not 3.

5.2.2 Memory channels and quantum cellular automata

In this place we will outline the close connection and interplay between one dimensional quantum cellular automata and memory channels. The results obtained here with proper proofs are still not published, but are the merit of [28].

The idea of generalizing the notion of classical cellular automata to quantum settings, can be traced back to Feynman where he in his paper [23] argued that quantum computation might out power the classical one. Various approaches have been considered for generalizing cellular automata to quantum regime [66, 15, 65, 21, 27]. In this section we will follow the approach of Werner et al. [56, 62, 2, 26], since this approach has a clear connection to memory channels.

Cellular automata are transformations of certain cell structure by a set of local rules. In one dimension the cell structure is an array of cells, where each cell has a finite number of possible states, for example 0 or 1. The automaton then transforms this state according to rules which depend on the states of cells in defined neighbourhood. Reversible automaton is when the transformation described by the rules is reversible, and the original state of the array cells can be reconstructed from the transformed cells.

In quantum setting a single cell will be a single quantum system in finite dimensional Hilbert space $\mathcal{A}_k$ with $d(k) = \dim \mathcal{A}_k$ and the cell state will be an operator from the algebra $\mathcal{L}(\mathcal{A}_k) = \mathfrak{A}_k$. The state of the infinite array is then from the quasi-local algebra as in Section 4.1. Thus quantum cellular automata (QCA) are transformations of the same structure as memory channels are. Whereas for memory channels the defining concept was causality, for reversible QCA's it will be locality, for existence of local rules and automorphism property for reversibility.

Definition 5.11 (Reversible 1D quantum cellular automaton)**.** A reversible one dimensional quantum cellular automaton C is a local automorphisms of the quasi-local algebra $C : \mathfrak{Z} \mapsto \mathfrak{Z}$. By automorphism we mean that it preserves the algebraic structure of $\mathfrak{Z}$:

$$C(x \cdot y) = C(x) \cdot C(y), \tag{5.50}$$

for all $x, y \in \mathfrak{Z}$ and by locality[6] we understand that

$$C(\mathfrak{B}_{\Lambda_1}) \subset \mathfrak{A}_{\Lambda_2}, \tag{5.51}$$

where Λ_2 is finite for every finite Λ_1.

Causality does not contradict to locality nor to the automorphic property. Thus there can exist reversible automata such that

$$C(b_{(-\infty,z]}) \in \mathfrak{A}_{(-\infty,z]}, \tag{5.52}$$

[6]The notion of locality is in some papers referred to as "causality" since it creates a cone where the information can propagate after several time steps. However since the term causality is already used for causal processes and moreover we would like to speak also about causal QCA we stick to the term local. This problem arises only because the natural time in QCA is usually the time step used to transform the whole cell structure. But in here we are interested only in one such time step, and introduce the time parameter *within* the cell structure, in the ordering of cells. This is completely natural from the point of memory channels and causal processes.

for every $z \in \mathbb{Z}$ which we call causal, motivated by (4.1). Moreover, since QCAs are local there exists for any automaton C a finite number τ such that

$$C(b_{(-\infty,z]}) \in \mathfrak{A}_{(-\infty,z+\tau]}, \tag{5.53}$$

for every $z \in \mathbb{Z}$. This means that every QCA can be made causal by a suitable shift:

$$S_\tau \circ C(b_{(-\infty,z]}) \in \mathfrak{A}_{(-\infty,z]}, \tag{5.54}$$

where S_τ is such that

$$S_\tau(b_z) = b_{z-\tau} \in \mathfrak{A}_{z-\tau}, \tag{5.55}$$

and is called a causal shift. Note that the shift transformation cannot be defined for arrays with cells of unequal size, thus we are bound to a translationally invariant cell structure. A shift transformation is also a causal QCA.

Causal QCA are memory channels as any other causal process. Due to locality, they also have finite depth. It turns out that every causal, 1D reversible QCA corresponds to a pure memory channel with finite depth and conversely each pure finite depth memory channel on a translationally invariant chain, defines a 1D reversible QCA.

For any QCA we can calculate a local invariant, so called index of a QCA, $\mathrm{ind}(C)$. The index theory of reversible 1D QCA was introduced in [26]. The index can be calculated from any sufficiently large, but finite portion of the automaton and represents the information flow inside the transformation. This index is always a ratio of two positive integers. For causal automata it is just an integer. Surprisingly this integer is equal to the smallest memory requirement of the memory channel which corresponds to such causal QCA. This is particularly nice, because it gives us a way to calculate the memory overhead needed to perform the transformation only from a finite part of the transformation in a well defined way. There have been some attempts to address this topic for the memory overhead in quantum convolutional codes [31] where the result is quite complicated and requires to find a longest path in a non-comutativity graph. Non-catastrophic convolutional codes, ie codes where errors spread to finite range, are equivalent to memory channels with finite depth.

5.3 Repeatable channels

In this section we will turn our attention in a slightly different direction. Until now we have discussed the effects of memory and how to counter them. Now we will try to find such memory channels which from some perspective do not exhibit any memory. Generally speaking a causal transformation T^* exhibits memory if

$$T^*(\omega_{[1,n]}) \neq T^* \otimes \cdots \otimes T^*(\omega_{[1,n]}). \tag{5.56}$$

Now let us have a pure memory channel with finite dimensional memory $\xi \in \mathcal{S}(\mathcal{M})$, with factorized input sequences $\omega_{[1,n]} = \omega_1 \otimes \cdots \omega_n \in \mathcal{S}(\mathcal{A}_1) \otimes \cdots \otimes \mathcal{S}(\mathcal{A}_n)$ and collision $U_k : \mathcal{M} \otimes \mathcal{A}_k \mapsto \mathcal{M} \otimes \mathcal{B}_k$. Since the input sequence is factorized, the transformation on any $\omega_{[a,b]}$, $T^*(\omega_{[a,b]})$ is a valid channel, ie. completely positive mapping. Let denote $T^*_{[a,b]}$ the channel T^* acting on $\omega_{[a,b]}$. Then we restate the relation 5.56 in a more readable way:

$$T^*_{[1,n]} \neq T^*_1 \otimes \cdots \otimes T^*_n \tag{5.57}$$

for a general memory channel. One can now be interested in the possibility of repeating some transformation G^* many times. The transformation may be a part of an experiment for example and experiments by their nature should be repeated many times. So there is a valid question which transformations G^* can be repeated in principle. From the example 4.6 of memoryless channels we have learned that if we need to repeat a channel G^* infinitely many times we need a memory system containing infinitely many particles at least of dimension of the minimal Stinespring's dilation for the channel G^*. This means that unless the dilation environment is trivial we need infinitely many particles and we consider this unphysical. However the dilation space is trivial only for unitary transformations. From this perspective the perfectly repeatable transformations G^* which require finite dimensional memory are only unitary.

There is not much we can do unless we lower the constraints on the repeatability of transformation. This is expressed in following definition:

Definition 5.12 (Repeatable transformation)**.** A channel G^* is *repeatable* if there exists a pure memory channel with finite dimensional memory, such that when constrained to factorized inputs $\omega_{[1,n]} = \omega_1 \otimes \cdots \omega_n$ we have

$$T^*_k = G^* \tag{5.58}$$

for all $k \in \mathbb{Z}$ where T^*_k is local transformation on k-th site.

Note that it stil holds that $T^*_{[a,b]} \neq G^* \otimes \cdots \otimes G^*$. We only require that *locally* the transformation on each site is equal G^*. As an example we can show that all random unitary channels are repeatable by explicitly constructing a corresponding memory channel. A random unitary channel $G^*_{\mathrm{ru}} : \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B})$ is of form:

$$G^*_{\mathrm{ru}}(\omega) = \sum_i p_i U_i \omega U_i^\dagger, \tag{5.59}$$

for any $p_i > 0$ such that $\sum_i p_i = 1$ and arbitrary set of unitaries $U_i : \mathcal{A} \mapsto \mathcal{B}$. The memory channel which employs these channels in repeatable fashion is the γ^* already used in Example 4.5. The memory channel exploited the unitary interaction $U_\gamma = F(\sum_{i=0}^{d_\mathcal{M}-1} |i\rangle\langle i| \otimes U_i)$ where we got

$$T^*{}_{\omega_{[1,n]}} = \sum_{i=0}^{d_\mathcal{M}-1} \langle i|\xi|i\rangle U_i^{\otimes n} \omega_{[1,n]} U_i^{\dagger \otimes n}, \tag{5.60}$$

where $\xi \in \mathcal{S}(\mathcal{M})$ was the initial state of memory. We compute $T^*_k = \mathrm{Tr}_{\forall l \neq k} T^*$ and get that

$$T^*_k(\omega_k) = \sum_i p_i U_i \omega_k U_i^\dagger, \tag{5.61}$$

where we assigned $p_i = \langle i|\xi|i\rangle$. Note that we do not even need the input sequence and the memory system to be factorized and this result will still hold true. Thus all random unitary channels are repeatable. Note that if a channel is repeatable it does not mean that it is always employed in repeatable fashion. The repeatability merely admits such option.

We can show that a necessary condition for a channel to be repeatable is that it has to be unital, ie. preserves the identity operator.

Theorem 5.13. *If a channel G^* is repeatable, then it is unital.*

Proof. Let us assume that we have a pure memory channel with finite dimensional memory system in initial state $\xi \in \mathcal{S}(\mathcal{M})$ with dimension $d_{\mathcal{M}} = \dim \mathcal{M}$ which employs a channel G^* in repeatable fashion. Assume factorized input sequence. Then

$$S(\xi) + \sum_{i=1}^{n} S(\omega_i) = S\big(U_{[1,n]}(\xi \otimes \omega_1 \otimes \cdots \omega_n)U_{[1,n]}^{\dagger}\big). \tag{5.62}$$

From equation 2.24 we know that the entropy of the whole is always less or equal to the entropy of the parts. Thus

$$S\big(U_{[1,n]}(\xi \otimes \omega_1 \otimes \cdots \omega_n)U_{[1,n]}^{\dagger}\big) \leq S\big(C_{[1,n]}^{*}(\xi)\big) + \sum_{i=1}^{n} S\big(G^*(\omega_i)\big), \tag{5.63}$$

where $C_{[1,n]}^{*}(\xi)$ is the final local state of memory after concurrent evolution and $G^*(\omega_i)$ is the local state of i-th particle after collision. Joining these two relations we get

$$\sum_{i=1}^{n} S(\omega_i) - S\big(G^*(\omega_i)\big) \leq S\big(C_{[1,n]}^{*}(\xi)\big) - S(\xi). \tag{5.64}$$

This means that the entropy *loss* on inputs cannot be greater than the entropy *gain* on memory system. The right hand side of (5.64) is bounded from above, because the dimension of memory is finite. The lowest entropy one can get on memory system is 0 and the highest entropy is $\log d_{\mathcal{M}}$ of a maximal mixture. Thus

$$\sum_{i=1}^{n} S(\omega_i) - S\big(G^*(\omega_i)\big) \leq \log d_{\mathcal{M}}. \tag{5.65}$$

Assume that all inputs are equal, $\omega_i = \omega$. Then

$$n\Big(S(\omega) - S\big(G^*(\omega)\big)\Big) \leq \log d_{\mathcal{M}}, \tag{5.66}$$

for every $n \in \mathbb{N}$. For $n \to \infty$ we have to get that $S(\omega) - S\big(G^*(\omega)\big) \leq 0$, what means that the transformation G^* cannot be entropy decreasing. This also means that such transformation has to preserve the complete mixture because complete mixture is the unique state which has the highest entropy. Since complete mixture is just scaled identity the map G^* has to preserve identity and therefore it has to be unital. It can be also shown that unital channels are entropy non-decreasing, see Appendix B.1. □

In qubit case all unital channels are also random unitary hence in qubit case unitality *implies* repeatability. Let us note that the concept of repeatability is similar to the concept of quantum cloning [60] in a sense that the channels (just like copies in quantum cloning) are not completely independent if measurements are taken into account. The impact of measurements on repeatability of quantum memory channels deserves further investigation, and will be partially addressed in the next chapter.

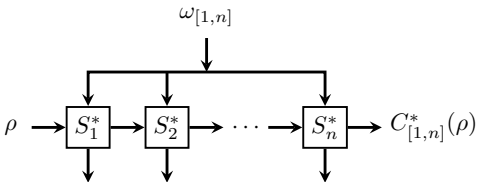

Fig. 5.1. The system ρ evolves in discrete time steps in a collision model effectively simulating some evolution of an open system interacting with an environment.

5.4 Simulation of indivisible qubit channels in collision models

Dynamics of open quantum systems is often modeled by the so-called master equations [39, 34]. For a comprehensive reference on evolution of open quantum systems see [18]. The idea is to get a time dependent channel describing the evolution, such that $\rho(t) = E_t^*(\rho)$, where ρ is the state of system at time $t = 0$ (thus $E^*_{t=0} = \mathbb{I}$). Often a Markovian approximation is made, where we assume that the environment is large and effectively doesn't change during the interaction with the system. In this approximation the one parametric class of channels E_t^* has a semigroup property: $E_s^* \circ E_t^* = E^*_{s+t}$. Such evolution can be stroboscopically simulated using a simple collision model.

Definition 5.14 (Stroboscopic simulation)**.** We say that a collision model $S^* : \mathcal{T}(\mathcal{M}) \otimes \mathcal{T}(\mathcal{A}) \mapsto \mathcal{T}(\mathcal{B}) \otimes \mathcal{T}(\mathcal{M})$ stroboscopically simulates time evolution E_t^* if

$$C^*_{[1,n]} = E^*_{n\Delta}, \tag{5.67}$$

for all $n \in \mathbb{N}$ and some $\Delta > 0$ and where $C^*_{[1,n]}$ is the concurrent channel on memory

$$C^*_{[1,n]}(\rho) = \mathrm{Tr}_{[1,n]}\big(S^*_{[1,n]}(\rho \otimes \omega[1,n])\big), \tag{5.68}$$

where $\mathrm{Tr}_{[1,n]}$ denotes partial trace over outputs $1, \ldots, n$. See for reference figure 5.1.

It is not important that the collision model is a memory channel. The inputs act only as an environment to the open system, which happens to be the memory system and we are not interested in the input-output relation of the memory channel. We focus only on the concurrent part. Stroboscopic simulation then simulates an evolution of open quantum system by discrete collisions with some structured environment, and approximates the continuous evolution with discrete time steps.

All Markovian evolutions are stroboscopically simulable. Let $E_t^* : \mathcal{T}(\mathcal{M}) \mapsto \mathcal{T}(\mathcal{M})$ be a Markovian evolution of some system with Hilbert space $\mathcal{M}$. Then let $U : \mathcal{M} \otimes \mathcal{A} \mapsto \mathcal{M} \otimes \mathcal{A}$ be a Stinespring's dilation of channel E^*_Δ:

$$E^*_\Delta(\rho) = \mathrm{Tr}_{\mathcal{A}}\big(U(\rho \otimes |0\rangle\langle 0|)U^\dagger\big), \tag{5.69}$$

for some $|0\rangle \in \mathcal{A}$. Then if we engineer environment in state $\omega_{[1,n]} = |0\rangle^{\otimes n}$ we get that

$$C^*_{[1,n]} = {E^*_\Delta}^{\otimes n} = E^*_{n\Delta}, \tag{5.70}$$

because of Markovianity of E_t^*. For any Markovian evolution we constructed a stroboscopic simulation with arbitrary small time steps. A natural property of Markovian evolutions which is also implicitly used in their stroboscopic simulation is their divisibility of any channel E_t^* (if it is not unitary evolution).

Definition 5.15 (Channel divisibility)**.** A channel E^* is called divisible if it can be written as a composition of two non-unitary channels:

$$E^* = E_1^* \circ E_2^*. \tag{5.71}$$

It is important that the channels E_1^* and E_2^* are not unitary. Otherwise we would get that all channels are trivially divisible.

We say that a channel E^* is stroboscopically simulable if there exists a stroboscopic simulation of a continuous time evolution E_t^* such that $E_{t=1}^* = E^*$. It is an interesting question whether also indivisible channels are stroboscopically simulable.

As a partial result we are able to say that every random unitary channel is stroboscopically simulable. For this we have to fix a right interaction and an appropriate state of environment. Let us have a random unitary channel

$$E^*(\rho) = \sum_{k=0}^{d-1} p_k V_k \rho V_k^\dagger, \tag{5.72}$$

where V_k is unitary and $\sum_k p_k = 1$. Let H_k be a Hamiltonian of V_k such that $e^{iH_k} = V_k$. Fix then $\dim \mathcal{A} = d$ and interaction $U : \mathcal{M} \otimes \mathcal{A} \mapsto \mathcal{M} \otimes \mathcal{A}$

$$U = \sum_{k=0}^{d-1} e^{\frac{i}{n} H_k} \otimes |k\rangle\langle k|. \tag{5.73}$$

Finally set the state of n environmental particles to be $\omega_{[1,n]} = \sum_{k=0}^{d-1} \sqrt{p_k} |k\rangle^{\otimes n}$. This will yield a stroboscopic evolution after m collisions

$$C_{[1,m]}^*(\rho) = \sum_k p_k e^{\frac{im}{n} H_k} \rho e^{-\frac{im}{n} H_k}, \tag{5.74}$$

giving us thus that

$$C_{[1,n]}^*(\rho) = \sum_k p_k e^{iH_k} \rho e^{-iH_k} = E^*. \tag{5.75}$$

One can replace this stroboscopic simulation by a continuous time evolution where

$$E_t^*(\rho) = \sum_k p_k e^{itH_k} \rho e^{-itH_k}, \tag{5.76}$$

thus the steps can be arbitrarily small.

As was first reported in [68], all indivisible qubit channels are of form:

$$E^*(\rho) = p_x X \rho X + p_y Y \rho Y + p_z Z \rho Z, \tag{5.77}$$

$p_x p_y p_z > 0$, hence all indivisible qubit channels are stroboscopically simulable. On figure 5.2 stroboscopic evolution of universal NOT evolution $E_{\mathrm{NOT}}^*(\rho) = 1/3(X\rho X + Y\rho Y + Z\rho Z)$ is shown.

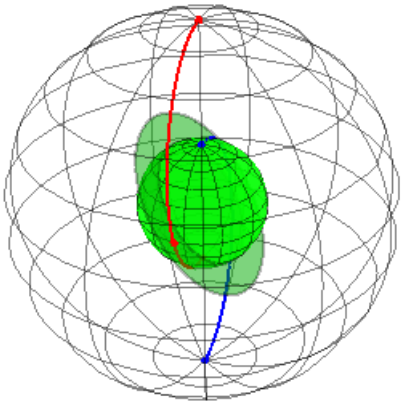

Fig. 5.2. The collision model simulating the continuous time evolution towards the universal NOT gate (shrunk Bloch sphere inversion). In particular, the transformation of the Bloch sphere (lines capture the time evolution of eigenstates of Z operator) is depicted for the time interval $t \in [0, 1]$. For $t = 2/3$ the channel $E^*_{t=2/3}$ is not invertible ($\det E^*_{t=2/3} = 0$) and at this time the Bloch sphere is mapped onto a two-dimensional disk. Let us note that images of eigenstates of Z operator are internal points of the disk. In fact, the whole disk is the image of pure states only. An animation of this evolution can be found at [71].

6 Estimation of memory channels

Memory channels differ from ordinary memoryless channels in the way of processing large messages sequentially. This has implications for coding and decoding algorithm used in communication when such memory channel is applied to process the data.

In a similar situation when an experimenter is faced with an unknown channel, which he would like to determine, he cannot assume that this channel is memoryless unless he has some a priori knowledge. Naturally the estimating technique has to differ from the memoryless one described in Section 3. We have to take into account that the memory transformation introduces correlations into the message.

6.1 Process estimation in memory settings

To our best knowledge, there is no general way how to do parameter estimation in memory settings. There has been some interesting studies of hamiltonian estimations with restricted access [7, 8, 9, 19, 30, 67]. In these works the task was to estimate coupling strengths of interaction between a set of particles. This set was divided into two subsets, one of which had the experimenter full control and the rest, which was inaccessible to experimenter, and served as the memory system. If the subset under control of experimenter possessed a certain simple property, called infectivity, the experimenter was able to manipulate the inaccessible part, in order to supplement its relaxation to a desired state, and then obtain the coupling strengths of the particular model. The work [19] didn't require the relaxation of inaccessible part, due to the symmetry of interaction.

Other studies [13] focused on discrimination of combs, where combs are in principle memory channels with finite length input. They assumed that the experimenter is able to replicate the comb perfectly, thus this task is equivalent to discrimination of memoryless channels with causal structure. This structure gives additional resources to the experimenter. One can vary the input states according to outputs of previous inputs, and thus introducing a different distance measure on such processes. As a result this distance measure allows a larger set of channels to be perfectly distinguishable than the usual cb-norm.

An estimation scheme applied to a memory channel is depicted on figure 6.1. The experimenter can prepare each input in a state from a set $\{\omega(k)\}$. Then he submits this input into channel and measures the output with some measurement, which can also depend on $\omega(k)$. The data from estimation are then collected in the string of pairs (preparation, outcome).

Such estimation scheme can be interpreted as a single measurement of the memory system. Thus the string of observed events cannot contain more information about the initial state of the memory than one can obtain from a single measurement. This is in contrast with memoryless channels where, depending on the interaction, some nontrivial information can be obtained about the initial state of memory. Thus the observed string of events only contains information about the inputs, measurements taken and the interaction.

Let us have look on following two examples. They show what happens when we use such estimation scheme naively, without taking into account the effects of memory.

Example 6.1 (Shift channel)**.** As was noted in previous chapters (see examples 4.3, 4.4), a simple shift channel σ_1^* on a qubit chain is modeled by a concatenated swap collision with a qubit

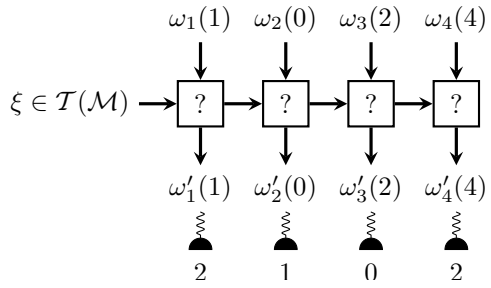

Fig. 6.1. Estimation scheme of a memory channel. The inputs ω_i are drawn from a finite set of preparations, $\omega_i(k)$ means i-th input is prepared with preparation k. Then the output $\omega'_i(k)$ is produced and a measurement is performed with some outcome l. The result of estimation is then a string of events, where event is a pair (k, l), preparation - outcome.

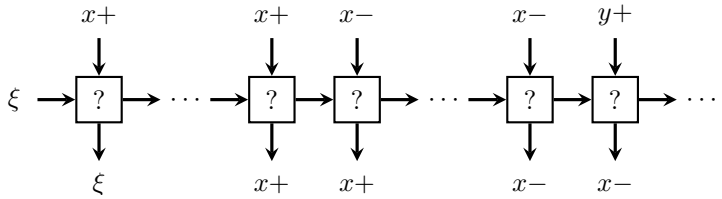

Fig. 6.2. Inputs are grouped to sequences of equal preparations. Independent of measurement the result of estimation will be an almost perfect channel.

memory. We will use an estimation procedure illustrated on figure 6.1. Let us have 6 preparations $k \in \{x+, x-, y+, y-, z+, z-, \}$ producing the eigenstates of respective Pauli matrices with ± 1 eigenvalues. Assume that we order the inputs so that first n inputs will be $\omega_i(x+), 0 < i \leq n$ then $\omega_i(x-), n < i \leq 2 * n$ and so on, see figure 6.2. Since $\omega'_i = \omega_{i-1}$ for all $i > 0$ and fix $\omega_0 \equiv \xi$, the resulting estimation will converge to an ideal channel. Because most of the time $\omega'_i = \omega_i$, only when $i = jn$ for some integer j it does not hold. As the statistics grows with n these cases become insignificant very fast. The result of estimation is independent of the details of measurement because of swap interaction.

We could choose a different strategy. We could first alternate $x+, x-$ inputs then alternate $y+, y-$ and so on. This strategy is depicted on figure 6.3. Surprisingly we will find that any input state goes to its orthogonal state. The result of estimation would be a non-completely positive mapping, a perfect NOT gate. We can see that the ordering of inputs can have significant impact on the result of estimation.

Yet, using a third strategy, we can choose the input states randomly, according to some discrete distribution with probabilities p_k of input being $\omega_i(k)$. This models a situation when the memory channel is used for communication, different preparations correspond to distinct "letters" and are distributed more or less randomly. In this case we find out that on average the output state of *any* input ω_i is $\overline{\rho} = \sum p_k \omega_{i-1}(k)$, the same as the average state of input. Thus the result

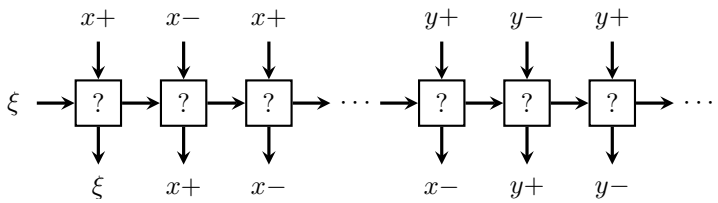

Fig. 6.3. Inputs are grouped to sequences of alternating orthogonal preparations. Independent of measurement the result of estimation will be a perfect NOT, a not completely positive mapping, hence unphysical.

of estimation would be a contraction to single state $\overline{\rho}$. This is a completely positive mapping.

For memoryless channels the ordering is not important, and you cannot possibly run into such problems. As we will see in next example not only ordering of inputs causes trouble.

Example 6.2 (Control not channel)**.** Let us have a memory channel with control not interaction $U_{\rm cnot}$, initial qubit memory state ξ and first input $\omega_1 = 1/2(\mathbb{I} + Z)$ in the positive eigenstate of Pauli matrix Z. Let the first measurement by also a perfect S-G measurement along the z-axis. Let $p_\pm$ denote the probabilities of measuring the qubit aligned or anti-aligned with the z-axis. Then

$$\begin{aligned} p_+ &= \mathrm{Tr}(U_{\rm cnot}(\xi \otimes \omega_1)U_{\rm cnot}^\dagger(\mathbb{I}_{\mathcal{M}} \otimes E_+)) = \langle 0|\xi|0\rangle \\ p_- &= \mathrm{Tr}(U_{\rm cnot}(\xi \otimes \omega_1)U_{\rm cnot}^\dagger(\mathbb{I}_{\mathcal{M}} \otimes E_-)) = \langle 1|\xi|1\rangle \end{aligned} \tag{6.1}$$

where $E_\pm = 1/2(\mathbb{I} \pm Z)$ are the effects of measurement observable. If the positive outcome was measured then the state of memory after collision is $\xi = 1/2(\mathbb{I} + Z)$ and if the negative outcome occurred the state would be $\xi = 1/2(\mathbb{I} - Z)$. However if memory is in one of those two states, the channel will behave exactly as unitary one. For the positive result it will be ideal channel and for the negative result the inputs will experience rotation by Pauli matrix X. The state of memory will remain untouched after the first collision. The first outcome will decide how the channel will behave afterwards. The result of estimation will be either an ideal channel or unitary rotation by X with probabilities equal to the diagonal elements of the initial state of memory.

In the examples we have observed the impact of ordering on the result of estimation. In the first case we used sequences of equal inputs. By the ordering we introduced a correlation. The state of memory before collision was in most cases the same as the input. In this sense input and state of memory were correlated. This happened also in the case of second ordering, when the state of memory before interaction was perpendicular to input. This relation of memory and colliding input was there put by hand, by means of the ordering. Intuitively it would be best to put the least "information" into the ordering. This can be achieved by the third option, when we have chosen the ordering to be random. In fact, if we choose random ordering the result of estimation will be always a completely positive map with a direct connection to interaction.

To scrutinize this let us have a memory channel whose input-output relation on first N inputs is described by channel $T^*(\omega_{[1,N]}) = \mathrm{Tr}_{\mathcal{M}}[S^*_{[1,N]}(\xi \otimes \omega_{[1,N]})]$. We can measure the N outputs of the channel only once, and obtain a single (collective) outcome. However we are free to

choose N, the number and state of inputs we use and the measurement performed for our estimation procedure. We also make the natural assumption that this memory channel is translationally invariant and thus fully described by the double $\{S^*, \xi\}$ where S^* describes the primitive collision of input with memory, $S^* : \mathcal{M} \otimes \mathcal{H} \mapsto \mathcal{H} \otimes \mathcal{M}$ and $\xi \in \mathcal{S}(\mathcal{H})$ is the initial state of the memory.

Remark 6.3. A memory channel specified by the double $\{S^*, \xi\}$ has the same input-output relation as another memory channel specified by $\{(\mathbb{I}_{\mathcal{H}} \otimes V_{\mathcal{M}})S^*(V_{\mathcal{M}}^\dagger \otimes \mathbb{I}_{\mathcal{M}}), V_{\mathcal{M}}^\dagger \xi V_{\mathcal{M}}\}$, for arbitrary unitary on memory space $V_{\mathcal{M}} : \mathcal{M} \mapsto \mathcal{M}$. Thus S^* can be estimated at most up to the unitary conjugation of the memory system.

Let us fix a set of distinctive preparations described by density matrices $\{\rho_k\}_{k=1}^{n_p}$, $\rho_k \in \mathcal{S}(\mathcal{H})$. For each preparation we also have a single measurement M_k described by set of effects $M_k = \{E_{kl}\}_{l=1}^{n_k}$, such that $E_{kl} > O$ and $\sum_{l=1}^{n_k} E_{kl} = \mathbb{I}_{\mathcal{H}}$. Let us employ following estimation procedure.

Estimation procedure:

1. choose randomly a preparation procedure from the set $\{\rho_k\}_{k=1}^{n_p}$ with probability distribution $\{p_k\}_{k=1}^{n_p}$, $\sum_{k=1}^{n_p} p_k = 1$, where this probability distribution is fixed and known for the whole estimation procedure. Let the chosen preparation be ρ_A.

2. submit ρ_A into memory channel.

3. perform measurement M_A on the output. Let the outcome be E_{AB}. Record the event $(\rho_A, E_{A,B})$ in a time-ordered list of events.

4. repeat steps 1. - 3. N times.

5. calculate $\tilde{p}(l|k) = \frac{N_{kl}}{N_k}$, where N_{kl} is the number of events (ρ_k, E_{kl}) and N_k is the number of times ρ_k was chosen as an input.

To complement this cooking recipe we also need a way how to interpret the data acquired by such estimation procedure. Thus we need following theorem:

Theorem 6.4 (Estimation interpretation)**.** *There exists such state of memory $\overline{\xi} \in \mathcal{S}(\mathcal{M})$ that*

$$\lim_{N\to\infty} \tilde{p}(l|k) = p(l|k), \tag{6.2}$$

where

$$p(l|k) = \mathrm{Tr}[U(\overline{\xi} \otimes \rho_k)U^\dagger(E_{kl} \otimes \mathbb{I}_{\mathcal{M}})] \tag{6.3}$$

is the probability of getting outcome l when preparation k was input.

Proof. Let $\xi \in \mathcal{S}(\mathcal{M})$ be the initial state of memory. Then

$$q_{kl} = p_k * \mathrm{Tr}[U(\xi \otimes \rho_k)U^\dagger(E_{kl} \otimes \mathbb{I}_{\mathcal{M}})] \tag{6.4}$$

is the probability of memory to jump to state

$$\xi' = \frac{\mathrm{Tr}_{\mathcal{H}}[U(\xi \otimes \rho_k)U^\dagger(E_{kl} \otimes \mathbb{I}_{\mathcal{M}})]}{\mathrm{Tr}[U(\xi \otimes \rho_k)U^\dagger(E_{kl} \otimes \mathbb{I}_{\mathcal{M}})]} \tag{6.5}$$

after collision with the input ρ_k and measured outcome E_{kl}. Since the inputs occur randomly, the memory system conducts a classical random walk on the memory space with probabilities of jump defined by (6.4) and ending positions of the jump defined by the equation (6.5) when the starting position was ξ. Let ξ_i be the state of memory entering the i-th collision, so that $\xi_1 = \xi$.

Let $\mathfrak{S}$ be the set of all ξ_i and $\mathfrak{S}_k$ the set of all ξ_i entering collision with ρ_k. Since input ρ_k occur randomly, $\mathfrak{S}_k$ is a random sample of $\mathfrak{S}$ for all k.

Precisely speaking, for every N we are able to calculate

$$N(\mathcal{X}) = |\mathfrak{S} \cap \mathcal{X}| \tag{6.6}$$

$$N_k(\mathcal{X}) = |\mathfrak{S}_k \cap \mathcal{X}| \tag{6.7}$$

for arbitrary $\mathcal{X} \subset \mathcal{S}(\mathcal{M})$.

Since the inputs occur independently at random with probability $p(k)$ we know that for large N

$$|N_k(\mathcal{X}) - p(k)N(\mathcal{X})| \approx O(\sqrt{N}). \tag{6.8}$$

This is because all of the states $\xi_i \in \mathcal{X}$ were divided between the sets $\mathfrak{S}_k$ at random with probability $p(k)$.

We can partition $\mathcal{S}(\mathcal{M})$ into mutually exclusive partitions $\{\mathcal{X}_\mu\}_\mu$ such that

$$\mathcal{X}_\mu \cap \mathcal{X}_\nu = 0 \quad \forall \mu \neq \nu \tag{6.9}$$

$$\mathcal{S}(\mathcal{M}) = \bigcup_\mu \mathcal{X}_\mu, \tag{6.10}$$

and define

$$p(\mathcal{X}_\mu) = \frac{N(\mathcal{X}_\mu)}{N} \tag{6.11}$$

$$p_k(\mathcal{X}_\mu) = \frac{N_k(\mathcal{X}_\mu)}{N_k} \approx \frac{p(k)N(\mathcal{X}_\mu)}{p(k)N} = p(\mathcal{X}_\mu). \tag{6.12}$$

Since the last equation is true for arbitrary partitioning in the limit of large N, we can conclude that the distribution of states in each $\mathfrak{S}_k$ is the same.

We would like to illustrate the fact that the distribution in each $\mathfrak{S}_k$ is independent of the outcomes of measurements. Imagine a game with two players. Player 1 has two boxes and infinite number of black and white cubes. Player 2 decides in which box will player 1 put the next cube. Player 1 decides the color of the cube.

Player 1 decides the color of the cubes in random fashion, with probability $p(w)$ he selects a white and with probability $p(b) = 1 - p(w)$ a black cube. Then, the distribution of white cubes among the boxes will be the same as the distribution of black cubes in the limit of many trials.

Let N_i be the number of cubes in i-th box and W_i, B_i the number of white, black cubes in i-th box. Then

$$\begin{aligned} W_i &\approx p(w)N_i \pm \sqrt{N} \\ B_i &\approx p(b)N_i \pm \sqrt{N} \end{aligned} \tag{6.13}$$

Player 2 can only influence the number of cubes in each box whatever strategy he uses. This argument can be used for arbitrary partitioning.

In our case player 1 is the person who chooses the inputs which collide with memory system. Player 2 chooses the outcome of the measurement and equation (6.13) is the analogue of equation (6.8).

The average state of each set $\mathfrak{S}_k$ is then also the same for all k. Lets denote the average state as $\overline{\xi}$. The probability of obtaining outcome E_{kl} when ρ_k was input is then

$$\mathrm{Tr}_{\mathcal{M}}[U(\overline{\xi} \otimes \rho_k)U^\dagger(E_{kl} \otimes \mathbb{I}_{\mathcal{M}})]. \tag{6.14}$$

□

These probabilities simply describe a single quantum channel T_1^*,

$$T_1^*(\omega) = \mathrm{Tr}_{\mathcal{M}}[U(\overline{\xi} \otimes \omega)U^\dagger]. \tag{6.15}$$

Similarly, by grouping δ subsequent events, the measured probabilities will correspond to a channel on δ subsequent inputs $T^*_{[1,\delta]}$

$$T^*_{[1,\delta]}(\omega_{[1,\delta]}) = \mathrm{Tr}_{\mathcal{M}}[U_{[1,\delta]}(\overline{\xi} \otimes \omega_{[1,\delta]})U^\dagger_{[1,\delta]}], \tag{6.16}$$

with the same $\overline{\xi}$ as T_1^*, where

$$\begin{aligned} U_{[1,\delta]} \quad = \quad & (\mathbb{I}_{[1,\delta-1]} \otimes U)(\mathbb{I}_{1,\delta-2} \otimes U \otimes \mathbb{I}_{\delta-1}) \\ & \cdots (U \otimes \mathbb{I}_{[2,\delta]}). \end{aligned} \tag{6.17}$$

The theorem however does not state what $\overline{\xi}$ is, and how it is dependent on the preparations ρ_k, their distribution p_k nor the measurements. To partially improve this situation we provide following lemma.

Lemma 6.5 (Fixed point lemma)**.** *Let C_k^* be the so called* concurrent *mapping*

$$C_k^*(\xi) = \mathrm{Tr}_{\mathcal{H}}[U(\xi \otimes \rho_k)U^\dagger]. \tag{6.18}$$

The average point $\overline{\xi}$ is then a fixed point of an average concurrent mapping, i.e.:

$$\sum_k p_k C_k^*(\overline{\xi}) = \overline{\xi}, \tag{6.19}$$

where

$$\overline{C}^* = \sum_k p_k C_k^* \tag{6.20}$$

is the average concurrent mapping.

Proof. Let $\widetilde{\mathfrak{S}_k}$ be the set of all ξ_i which *exit* a collision with ρ_k irrespective of the measurement outcome. We know that the average state of $\mathfrak{S}_k$ is $\overline{\xi}$. Since we ignore the measurement outcomes (sum over them), the average state of $\widetilde{\mathfrak{S}_k}$ is $C_k^*(\overline{\xi}) =: \overline{\xi}^k$.

The average state of $\mathfrak{S}$ can be also calculated as the average of all exiting states, because the set of all exiting states of memory and all entering states of memory are identical. Thus

$$\sum_k p_k \overline{\xi}^k = \overline{\xi} = \sum_k p_k C_k^*(\overline{\xi}) =: \overline{C}^*(\overline{\xi}), \tag{6.21}$$

wher we calculated the average exiting state as the weighted sum of average exiting states after fixed input ρ_k. □

If the average concurrent mapping is contractive, it has unique fixed point. The average state of memory is then this point.

6.2 Control unitary interaction

In example 6.2 we have seen that the result of estimation was a unitary channel, in spite of that the interaction of memory and input was not factorized. In this section we will prove that this will be true whenever the interaction is a control unitary interaction, i.e. whenever

$$U = \sum_{i=0}^{d_{\mathcal{M}}-1} |i\rangle\langle i| \otimes U_i, \tag{6.22}$$

where U_i are unitaries on $\mathcal{A}$. And this result will be irrespective of details of the estimation procedure. The proof is as follows. Fix any estimation procedure. Given a memory channel, we can also define a probability distribution on all possible outcomes that could come out of the estimation procedure. If we would repeat the same experiment, same initial memory state and same input sequence, infinitely many times the outcomes of the estimation will be distributed according to this probability distribution.

The only relevant parameters of memory are the diagonal elements of memory state ξ in the $\{|i\rangle\}$ basis as was already observed in (4.13). If, say $\langle 0|\xi|0\rangle = 1$ and all others are zero the probability distribution over possible outcomes, P_0, is as if the box was memoryless unitary evolution U_0 and similarly for other diagonal elements with distributions P_i. Due to linearity of the whole procedure, for a general state ξ the probability distribution over possible outcomes, P_ξ, will be the convex combination of P_i

$$P_\xi = \sum_{i=0}^{d_{\mathcal{M}}-1} \langle i|\xi|i\rangle P_i. \tag{6.23}$$

Thus only outcomes belonging to unitary channels U_i have solid probabilities. If the estimation procedure was informationally complete and statistics infinite, the result of estimation will be a memoryless unitary channel U_i with probability $\langle i|\xi|i\rangle$. For any finite statistics such estimation cannot disprove that the channel is not one of the U_i thus asymptotically will converge to the anticipated result.

In terms of the one shot measurement of the memory, the estimation is equal to measurement of the memory in the basis $\{|i\rangle\}$, assuming all U_i are different. If the input is an eigenstate of

U_i, call it $|\psi\rangle$ and you measure on the output a perpendicular state $|\psi_\perp\rangle$. Then the result of estimation cannot be U_i, because it prohibits such events. The diagonal term ξ_{ii} will be set to zero after such event occurs. This has an interesting application. Given imperfect measurements on $\mathcal{A}$ one can attain asymptotically a projective measurement on $\mathcal{M}$ in aforementioned basis, by doing tomography of the memory channel. The off-diagonal terms of initial memory state go to zero exponentially and the diagonal terms ξ_{ii} are proportional to the probability of measuring the particular chain of events in unitary channel U_i. In the limit of many uses, the memory will be projected to one pure diagonal state.

Note that a special case of a control unitary interaction is factorized interaction

$$U = W \otimes V = \sum_{i=0}^{d_{\mathcal{M}}-1} |\psi_i\rangle\langle\psi_i| \otimes e^{i\phi_i} V, \tag{6.24}$$

where $|\psi_i\rangle$ and $e^{i\phi_i}$ are the eigenvectors and eigenvalues of W. For these interactions the result is completely obvious and natural.

We can use similar argumentation for a larger class of interactions of following form

$$U = \sum_{i=0}^{d_{\mathcal{M}}-1} |\pi(i)\rangle\langle i| \otimes U_i, \tag{6.25}$$

where $\pi(i)$ gives another member of the basis of memory. As in the previous case, the only relevant parameters of memory are the diagonal elements in the $\{|i\rangle\}$ basis. Now the pure diagonal state of memory is not stationary but constantly cycles through the basis states. Thus the sequence of events will correspond to a cyclical change of unitaries. In fact this is again a control unitary channel if we group the inputs to larger sequences, this is when $\pi \cdots \pi(i) = i$ again, i.e. when the memory makes full cycle and returns to the original state. Then

$$U_{[1,n]} = \sum_{i=0}^{d_{\mathcal{M}}-1} |i\rangle\langle i| \otimes U_i \otimes U_{\pi(i)} \otimes U_{\pi(\pi(i))} \otimes \ldots, \tag{6.26}$$

where n is the length of the cycle.

6.3 2D case study

We are going to examine the simplest example when we have a two dimensional memory system and sequence of two dimensional inputs combined with a special unitary interaction of form

$$U = e^{i\frac{1}{2}(\alpha_x X\otimes X + \alpha_y Y\otimes Y + \alpha_z Z\otimes Z)}, \tag{6.27}$$

with $-\pi \leq \alpha_i \leq \pi$.

The most general 2-qubit unitary interaction differs from this one only by applying local unitaries on both sides. We could also study such case however it does not provide much more insight than this basic example.

Given this U, our task is to estimate the angles α_i. Due to the symmetries of the problem, not all information can be obtained. Since we restrict ourselves to have access only to the input

- output part, and have no access to memory degrees of freedom, the class of all interaction that will yield the same I-O relation is (see 5.6):

$$U = e^{i\frac{1}{2}(\alpha_x X'\otimes X + \alpha_y Y'\otimes Y + \alpha_z Z'\otimes Z)}, \tag{6.28}$$

where $X' = \gamma X \gamma^\dagger$ and others are just unitarily conjugated Pauli matrices with arbitrary unitary γ. This implies that any two signs of the angles α_i can be flipped simultaneously by choosing one of the Pauli matrices as the unitary γ. Thus we can only estimate the sign of the product of all three angles, $\alpha_x\alpha_y\alpha_z$, which is invariant under such conjugation. Furthermore the shift by π of all three angles simultaneously introduces only a global phase on the unitary U, which is undetectable.

We will use estimation scheme described in previous section 6.1, randomizing set of preparations and measurements. We have a set of testing preparations which prepare testing states $\{\rho_k\}$ that we input randomly with probabilities p_k and a set of measurement observables. Fix the average state of memory as

$$\overline{\xi} = \frac{1}{2}(\mathbb{I} + \overline{m}_x X + \overline{m}_y Y + \overline{m}_z Z), \tag{6.29}$$

for suitable $\overline{m}_x, \overline{m}_y, \overline{m}_z$. Then any probability $p(k,l)$ of some event (k,l) is consistent with a channel in vector representation

$$A(\mathcal{E}_1^*) = A^1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \overline{m}_x s_y s_z & c_y c_z & \overline{m}_z c_y s_z & -\overline{m}_y s_y c_z \\ \overline{m}_y s_x s_z & -\overline{m}_z c_x s_z & c_x c_z & \overline{m}_x s_x c_z \\ \overline{m}_z s_x s_y & \overline{m}_y c_x s_y & -\overline{m}_x s_x c_y & c_x c_y \end{pmatrix}, \tag{6.30}$$

where the matrix is written with respect to the traceless operator basis of Pauli matrices and $c_i = \cos\alpha_i$ and $s_i = \sin\alpha_i$.

The diagonal elements of the vector representation of channel $\mathcal{E}_1^*$ are independent of the average state of memory. Thus we can prepare the two testing states $\rho_{i,\pm} = 1/2(\mathbb{I} \pm \sigma_i)$ with random probability $p_{i,+} = p_{i,-} = 1/2$ and a S-G experiment along the same axis with effects $E_{i,\pm} = 1/2(\mathbb{I} \pm \sigma_i)$ where σ_i is one of the Pauli matrices. The appropriate diagonal element is then

$$\begin{aligned} A^1_{ii} &= 1/2\mathrm{Tr}(\mathcal{E}_1^*(\sigma_i)\sigma_i) = 1/4\big(\mathrm{Tr}(\mathcal{E}_1^*(\rho_{i,+})E_{i,+}) + \mathrm{Tr}(\mathcal{E}_1^*(\rho_{i,-})E_{i,-}) \\ &\quad -\mathrm{Tr}(\mathcal{E}_1^*(\rho_{i,+})E_{i,-}) - \mathrm{Tr}(\mathcal{E}_1^*(\rho_{i,-})E_{i,+})\big) \\ &= 1/2(p_{(i+,i+)} + p_{(i-,i-)} - p_{(i+,i-)} - p_{(i-,i+)}), \end{aligned} \tag{6.31}$$

where $p_{(\cdot,\cdot)}$ is the probability of measured event. If all A^1_{ii}, then

$$\begin{aligned} \cos\alpha_x &= \sqrt{\frac{A^1_{yy}A^1_{zz}}{A^1_{xx}}} \\ \cos\alpha_y &= \mathrm{sgn}(A^1_{zz})\sqrt{\frac{A^1_{xx}A^1_{zz}}{A^1_{yy}}} \\ \cos\alpha_z &= \mathrm{sgn}(A^1_{yy})\sqrt{\frac{A^1_{xx}A^1_{yy}}{A^1_{zz}}}, \end{aligned} \tag{6.32}$$

where $\cos\alpha_x$ can be always taken positive because of the π-shift symmetry mentioned earlier, The $\mathrm{sgn}(A^1_{ii})$ is the sign of the diagonal element A^1_{ii}. The last missing piece of information, the sign of $\alpha_x\alpha_y\alpha_z$ cannot be read out from the local channel. It can be obtained from the map on two subsequent inputs, $\mathcal{E}^*_{[1,2]}$. We do not need the whole $\mathcal{E}^*_{[1,2]}$. It is enough to look at an *posterior mapping*, the channel after fixed input $\rho_{i,\pm}$:

$$\mathcal{E}^*_2(\omega|\rho_{i,\pm}) = \mathrm{Tr}_1\big(\mathcal{E}^*_{[1,2]}(\rho_{i,\pm}\otimes\omega)\big). \tag{6.33}$$

For example the vector representation of channel after the input $\rho_{x,+}$ is

$$\begin{aligned} A\big(\mathcal{E}^*_2(\cdot|\rho_{x,+})\big) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ \overline{m}'_x s_y s_z & c_y c_z & \overline{m}'_z c_y s_z & -\overline{m}'_y s_y c_z \\ \overline{m}'_y s_x s_z & -\overline{m}'_z c_x s_z & c_x c_z & \overline{m}'_x s_x c_z \\ \overline{m}'_z s_x s_y & \overline{m}'_y c_x s_y & -\overline{m}'_x s_x c_y & c_x c_y \end{pmatrix} \\ &=: A^{2|x+}, \end{aligned} \tag{6.34}$$

where $\overline{m}'$ describes the posterior average state of memory after the input ρ_{x+}:

$$\begin{aligned} \overline{m}'_x &= c_y c_z \overline{m}_x + s_y s_z \\ \overline{m}'_y &= c_x c_z \overline{m}_x + s_x c_z \\ \overline{m}'_z &= c_x c_y \overline{m}_x + s_x c_y. \end{aligned} \tag{6.35}$$

If we also calculate $A^{2|x-}$ and subtract it from $A^{2|x+}$ we will get

$$\frac{1}{2}(A^{2|x+} - A^{2|x-}) =: A^{2|X}, \tag{6.36}$$

where

$$A^{2|X} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ s_y^2 s_z^2 & 0 & -c_y^2 s_z s_x \overline{m}_y & -c_z^2 s_x s_y \overline{m}_z \\ s_x^2 c_z s_z \overline{m}_z & c_x c_y s_z s_x \overline{m}_y & 0 & c_z s_x s_y s_z \\ -c_y s_x^2 s_y \overline{m}_y) & c_x c_z s_y s_x \overline{m}_z & c_y s_x s_y s_z & 0 \end{pmatrix}. \tag{6.37}$$

The sign of $\alpha_x\alpha_y\alpha_z$ is then obtained from the sign of $A^{2|X}_{yz}$ or $A^{2|X}_{zy}$ irrespective of the average state of memory. The probabilities we need to measure are

$$\begin{aligned} A^{2|x+}_{yz} &= 1/2\mathrm{Tr}(\mathcal{E}^*_2(Z|\rho_{x,+})Y) = 1/4\big(\mathrm{Tr}(\mathcal{E}^*_2(\rho_{z,+}|\rho_{x,+})E_{y,+}) + \mathrm{Tr}(\mathcal{E}^*_2(\rho_{z,-}|\rho_{x,+})E_{y,-}) \\ &\quad -\mathrm{Tr}(\mathcal{E}^*_2(\rho_{z,+}|\rho_{x,+})E_{y,-}) - \mathrm{Tr}(\mathcal{E}^*_2(\rho_{z,-}|\rho_{x,+})E_{y,+})\big) \\ &= 1/2(p_{(z+,y+|\rho_{x,+})} + p_{(z-,y-|\rho_{x,+})} - p_{(z+,y-|\rho_{x,+})} - p_{(z-,y+|\rho_{x,+})}), \end{aligned} \tag{6.38}$$

and analogously for

$$A^{2|x-}_{yz} = 1/2(p_{(z+,y+|\rho_{x,-})} + p_{(z-,y-|\rho_{x,-})} - p_{(z+,y-|\rho_{x,-})} - p_{(z-,y+|\rho_{x,-})}), \tag{6.39}$$

where $p(a,b|c)$ is the posterior probability of event (a,b) right after c occurred. This requires that the average state of memory has to be the same for both maps during the estimation.

If A^1 has zero elements on diagonal then at least two of them have to be zero. This is when at least one of the $\alpha_i = \pm\pi/2$. One can get from (5.62) the only remaining nonzero product $A^1_{ii} = c_k c_l$. To obtain c_k and c_l we have to look in the posterior part of $\mathcal{E}^*_{[1,2]}$:

$$A^{2|l\pm}_{l,0} = \frac{1}{2}\Big(\mathrm{Tr}\big(\mathcal{E}^*_2(1/2\mathbb{I}|\rho_{l,\pm})\rho_{l,+}\big) - \mathrm{Tr}\big(\mathcal{E}^*_2(1/2\mathbb{I}|\rho_{l,\pm})\rho_{l,-}\big)\Big) =$$
$$p(l+, l+|\rho_{l,\pm}) + p(l-, l+|\rho_{l,\pm}) - p(l+, l-|\rho_{l,\pm}) - p(l-, l-|\rho_{l,\pm})$$
$$= s_i s_k (c_i c_k \overline{m}_l \pm s_i s_k) \tag{6.40}$$
$$= \pm s_k^2, \tag{6.41}$$

where the last equality is because $c_i = 0$ and $s_i^2 = 1$. In case $c_i = c_k = 0$ we will get a swap of classical information, a memory channel with depth $\delta = 2$ similar to the one in example **??**, where the l-component is swapped to the l-component of the subsequent input. And finally if all $c_i = 0$ we have a memory channel with depth $\delta = 1$, what is a swap-like interaction, which can be easily checked on the posterior maps. In this case also the sign of $\alpha_x\alpha_y\alpha_z$ cannot be measured, because it only introduces a global phase shift on the swap interaction.

Thus, a good strategy is following. For every i, input randomly states $\rho_{i,\pm}$ with equal probabilities $1/2$[7] and do a measurement along the same axis with effects $E_{i,\pm}$, in order to obtain the diagonal elements of A^1 from (6.31) and the elements of posterior map using (6.40):

$$\frac{1}{2}(A^{2|l+}_{l,0} - A^{2|l-}_{l,0}) = s_i^2 s_k^2. \tag{6.42}$$

Then, when lets say $c_z \neq 0$, we randomly input four 2-qubit sequences $\rho_{x,+}\otimes\rho_{z,+}$, $\rho_{x,-}\otimes\rho_{z,+}$, $\rho_{x,+}\otimes\rho_{z,-}$ and $\rho_{x,-}\otimes\rho_{z,-}$ with equal probabilities $1/4$ and measure the second output along the y- axis to measure the probabilities in (6.38) and (6.39) for estimating the $c_z s_x s_y s_z$ and subsequently the sign of $\alpha_x\alpha_y\alpha_z$. Note that this sign is unobservable if at least one of the angles is a multiple of π.

[7] The probabilities can be arbitrary, however it is good to keep them equal to have comparable statistics of events.

Appendices

A Hilbert space refresher

A.1 Hilbert space

Let $\mathcal{H}$ be a complex vector space.

Definition A.1 (Inner Product)**.** A complex valued function $\langle\cdot|\cdot\rangle$ on $\mathcal{H} \times \mathcal{H}$ is called an *inner product* on $\mathcal{H}$ if it satisfies following three conditions for all vectors $\phi, \psi, \theta \in \mathcal{H}$ and $c \in \mathbb{C}$:

- $\langle\phi|\phi\rangle \geq 0$ if $\phi \neq 0$ - positive definiteness
- $\langle\phi|\psi + c\theta\rangle = \langle\phi|\psi\rangle + c\langle\phi|\theta\rangle$ - linearity in second argument
- $\overline{\langle\phi|\psi\rangle} = \langle\psi|\phi\rangle$ - conjugate symetricity

We say that two vectors $\phi, \psi \neq 0$ of an inner product space $\mathcal{H}$ are orthogonal iff $\langle\phi|\psi\rangle = 0$. A set $X \subset \mathcal{H}$ is orthogonal set if any pair of vectors from X is orthogonal.

Definition A.2 (Finite Dimensional space)**.** The inner product space $\mathcal{H}$ is of dimension $d \in \mathbb{N}, d < \infty$ if there does not exist a orthogonal set of k vectors such that $k > d$ but a orthogonal set of d vectors exists. If no such d exists, the space $\mathcal{H}$ is *infinitely dimensional*. Any set of d orthogonal vectors forms a (unnormed) basis in d-dimensional Hilbert space.

A complex vector space $\mathcal{H}$ equipped with inner product is a normed space with a norm defined as:

Definition A.3 (Canonical Norm)**.** The canonical norm of a vector $\psi \in \mathcal{H}$ with respect to inner product on $\mathcal{H}$ is

$$\| \psi \| \equiv \langle\psi|\psi\rangle^{\frac{1}{2}}. \tag{A.1}$$

Normed space is *complete* if every Cauchy sequence is convergent and *separable* if it has a countable dense subset.

Definition A.4 (Hilbert Space)**.** Any complete separable inner product space with respect to the norm A.1 is a *Hilbert space*.

Every finite dimensional inner product space is separable and complete and thus a Hilbert space. In this work all Hilbert spaces will be mostly finite dimensional unless explicitly stated.

Definition A.5 (Basis)**.** Every set of d orthonormal vectors $\{\psi_i\}$ in Hilbert space $\mathcal{H}$, where d is the dimension of this Hilbert space is a basis of this Hilbert space. Any vector ϕ can be then expressed as

$$\phi = \sum_{i=1}^{d} c_i \psi_i, \tag{A.2}$$

where $c_i = \langle\psi_i|\phi\rangle$.

Remark A.6 (Cauchy-Schwartz Inequality)**.** For every inner product space following inequality holds: if $\psi, \phi \in \mathcal{H}$ then

$$|\langle\phi|\psi\rangle|^2 \leq \langle\phi|\phi\rangle\langle\psi|\psi\rangle. \tag{A.3}$$

The equality happens only if ϕ and ψ are linearly dependent, ie. $\phi = c\psi$ for some $c \in \mathbb{C}$.

Definition A.7 (Linear Functional, Dual Space)**.** A linear mapping f from a complex vector space V to the field of complex numbers is called a linear functional. If this V has norm defined we can construct the set of all continuous linear functionals on V called the dual space V^*. This is also a vector space where the linear structure can be defined pointwise: $(f_1 + cf_2)(v) = f_1(v) + cf_2(v)$ for all $v \in V$ and normed with norm

$$\| f \| := \sup_v \frac{|f(v)|}{\| v \|}. \tag{A.4}$$

In Hilbert space $\mathcal{H}$ every vector ϕ defines such linear functional by the formula

$$f_\phi = \langle\phi|\psi\rangle \tag{A.5}$$

for every $\psi \in \mathcal{H}$.

Lemma A.8 (Riesz)**.** *Let $f \in \mathcal{H}^*$, then there exists a unique vector $\phi \in \mathcal{H}$ such that*

$$f(\psi) = \langle\phi|\psi\rangle =: f_\phi(\psi) \tag{A.6}$$

for every $\psi \in \mathcal{H}$. Moreover $\| f_\phi \| = \| \phi \|$.

Remark A.9 (Dirac Notation)**.** A single vector $\psi \in \mathcal{H}$ can be written as $|\psi\rangle$ and is called as *ket vector*. Symbol $\langle\phi|$ will denote a linear functional:

$$\psi \mapsto \langle\phi|\psi\rangle, \tag{A.7}$$

and is called *bra vector*. The inner product $\langle\phi|\psi\rangle$ will be then called a *bra(c)ket*.

A.2 Linear operators on Hilbert spaces

Definition A.10 (Linear operator)**.** We call a mapping $A : \mathcal{H} \mapsto \mathcal{H}$ *linear* if

$$A(p\psi + q\phi) = pA\psi + qA\phi, \tag{A.8}$$

for every $\psi, \phi \in \mathcal{H}$ and every $p, q \in \mathbb{C}$.

Definition A.11 (Bounded operator)**.** We call a linear mapping $A : \mathcal{H} \mapsto \mathcal{H}$ an *operator*. An operator A is *bounded* if there exists such number $t < \infty$ that

$$\| A\psi \| \leq t\| \psi \| \tag{A.9}$$

for all $\psi \in \mathcal{H}$. The set of all bounded operators on $\mathcal{H}$ is $\mathcal{L}(\mathcal{H})$. This set has a structure of a complex vector space. In finite dimensional Hilbert space the elements of $\mathcal{L}(\mathcal{H})$ can be represented by square matrices with finite matrix elements $\mathcal{M}_d(\mathbb{C})$.

Definition A.12 (Operator Norm)**.** The norm of bounded operator A is

$$\| A \|_\infty = \sup_\psi \frac{\| A\psi \|}{\| \psi \|}. \tag{A.10}$$

Definition A.13 (Adjoint operator)**.** For every operator $A \in \mathcal{L}(\mathcal{H})$ we can define the *adjoint operator* $A^\dagger$ as

$$\langle \phi | A^\dagger \psi \rangle = \langle A\phi | \psi \rangle \tag{A.11}$$

for all $\psi, \phi \in \mathcal{H}$. For every A, $(A^\dagger)^\dagger = A$.

Definition A.14 (C^*-algebra)**.** A C^*-algebra $\mathfrak{A}$ is an associative algebra over complete normed complex vector space, equipped with a $\dagger$ involution such that

$$(A + cB)^\dagger = A^\dagger + \bar{c}B^\dagger \tag{A.12}$$
$$(AB)^\dagger = B^\dagger A^\dagger \tag{A.13}$$
$$(A^\dagger)^\dagger = A \tag{A.14}$$
$$\| A^\dagger A \| = \| A \| \| A^\dagger \|, \tag{A.15}$$

for every $c \in \mathbb{C}$ and $A, B \in \mathfrak{A}$.

The algebra of bounded operators $\mathcal{L}(\mathcal{H})$ over Hilbert space $\mathcal{H}$ together with the $\dagger$ operation and operator norm $\| \cdot \|_\infty$ is a C^*-algebra and conversely every C^*-algebra can be viewed as subalgebra of operators over some suitable Hilbert space. We call a C^*-algebra $\mathfrak{A}$ *unital* if it has identity.

Definition A.15 (Self-adjoint operator)**.** Let $A \in \mathcal{L}(\mathcal{H})$. If $A^\dagger = A$ we call such operator *self-adjoint*. The set of all self-adjoint operators is $\mathcal{L}_S(\mathcal{H})$.

For every self-adjoint operator A we know that $\langle \psi | A\psi \rangle$ is real:

$$\overline{\langle \psi | A\psi \rangle} = \langle A\psi | \psi \rangle = \langle \psi | A^\dagger \psi \rangle = \langle \psi | A\psi \rangle. \tag{A.16}$$

Definition A.16 (Positive operator)**.** We call an operator $A \in \mathcal{L}(\mathcal{H})$ positive if for every $|\psi\rangle \in \mathcal{H}$

$$\langle \psi | A\psi \rangle \geq 0. \tag{A.17}$$

If A is an positive operator, we write $A \geq O$.

It directly follows that positive operators are self-adjoint. It also follows that $A^\dagger A$ is a positive operator since

$$\langle \psi | A^\dagger A\psi \rangle = \langle A\psi | A\psi \rangle = \| A\psi \|^2 \geq 0 \tag{A.18}$$

Lemma A.17 (Square root lemma)**.** *Let $A \in \mathcal{L}_S(\mathcal{H})$ such that $A \geq O$. Then there is a unique positive operator $A^{1/2}$ such that $A^{1/2}A^{1/2} = A$. The operator $A^{1/2}$ is called a square root of A.*

Definition A.18 (Absolute value)**.** Absolute value of an operator $A \in \mathcal{L}(\mathcal{H})$ is $|A| := (A^\dagger A)^{1/2}$.

Definition A.19 (Trace of a bounded operator)**.** We define a *trace* of a bounded operator $A \in \mathcal{L}(\mathcal{H})$ as

$$\mathrm{Tr}(A) := \sum_i \langle \psi_i | A \psi_i \rangle, \tag{A.19}$$

where $\{|\psi_i\rangle\}$ forms an orthonormal basis on $\mathcal{H}$.

Definition A.20. The set of all operators $A \in \mathcal{L}(\mathcal{H})$ for which $\mathrm{Tr}(|A|) < \infty$ is called a *trace class* and is denoted as $\mathcal{T}(H)$.

The reason why we insist that the trace of *absolute value* of operator exists is that in infinite dimensions this ensures that the trace is unitarily invariant.

Definition A.21 (Trace norm)**.** Trace norm of an operator $A \in \mathcal{T}(\mathcal{H})$ is

$$\| A \|_{\mathrm{tr}} \equiv \mathrm{Tr}|A| = \mathrm{Tr}\sqrt{AA^\dagger}. \tag{A.20}$$

Definition A.22 (Hilbert-Schmidt product)**.** We can define an inner product on $\mathcal{T}(\mathcal{H})$:

$$\mathrm{Tr}[A^\dagger B] =: \langle A|B\rangle_{\mathrm{HS}}. \tag{A.21}$$

Hence $\mathcal{T}(\mathcal{H})$ is also a Hilbert space in finite dimension. Given that $\{|\psi_i\rangle\}$ is an orthonormal basis in d-dimensional $\mathcal{H}$ we can define a $d \times d$ dimensional orthonormal basis on $\mathcal{T}(\mathcal{H})$ as $\{|\psi_i\rangle\langle\psi_j|\}$ and every trace class operator can be expresed as

$$A = \sum_{ij} a_{ij} |\psi_i\rangle\langle\psi_j|, \tag{A.22}$$

where $a_{ij} = \mathrm{Tr}[(|\psi_i\rangle\langle\psi_j|)^\dagger A] = \langle\psi_i|A\psi_j\rangle$. Thus $\mathcal{T}(\mathcal{H})$ can be identified with the set of $d \times d$ complex matrices $\mathcal{M}_d(\mathbb{C})$. Let a_{jk} be the matrix entries of an operator $A \in \mathcal{T}(\mathcal{H})$. Arbitrary linear functional $f : \mathcal{T}(\mathcal{H}) \mapsto \mathbb{C}$ can be written as

$$\begin{aligned} f(A) &= f(\sum_{ij} a_{ij}|\psi_i\rangle\langle\psi_j|) = \sum_{ij} a_{ij} f(|\psi_i\rangle\langle\psi_j|) \\ &= \sum_{ij} a_{ij} s_{ij} = \mathrm{Tr}(AS). \end{aligned} \tag{A.23}$$

Every linear functional on $\mathcal{T}(\mathcal{H})$ is thus represented by matrix $S \in \mathcal{M}_d(\mathbb{C})$ via the trace formula A.23. This is consistent with the Riesz lemma A.8 if we equip $\mathcal{M}_d(\mathbb{C})$ with the Hilbert-Schmidt inner product. Let us note that in finite dimensions $\mathcal{L}(\mathcal{H}) \equiv \mathcal{L}(\mathcal{H})^* \equiv \mathcal{T}(\mathcal{H}) \equiv \mathcal{T}(\mathcal{H})^*$.

When we move into infinite dimensional case $\mathcal{L}(\mathcal{H})$ and $\mathcal{T}(\mathcal{H})$ are no longer Hilbert spaces. However $\mathcal{T}(\mathcal{H})$ is still a normed vector space and is an ideal in $\mathcal{L}(\mathcal{H})$ so that $\mathrm{Tr}(BA) < \infty$ for every $A \in \mathcal{T}(\mathcal{H})$ and $B \in \mathcal{L}(\mathcal{H})$. For each $B \in \mathcal{L}(\mathcal{H})$ we define a linear functional f_B on $\mathcal{T}(\mathcal{H})$ with

$$f_B(A) = \mathrm{Tr}(BA), \tag{A.24}$$

for every $A \in \mathcal{T}(\mathcal{H})$. Thus $\mathcal{T}(\mathcal{H})^* = \mathcal{L}(\mathcal{H})$ for infinitely dimensional $\mathcal{H}$.

The reason why $\mathcal{T}(\mathcal{H})$ is not a Hilbert space in infinite dimensions is that in fact it is too small. It turns out that the Hilbert-Schmidt inner product makes sense for a larger class of operators than trace class operators. Operators which satisfy the Hilbert-Schmidt norm $\| A \|_{\mathrm{HS}} := \sqrt{\mathrm{Tr}(A^\dagger A)} \leq \infty$ form a Hilbert space.

Definition A.23 (Unitary operator)**.** We call an operator U unitary if

$$UU^\dagger = U^\dagger U = \mathbb{I}. \tag{A.25}$$

The inner product in $\mathcal{H}$ is invariant under unitary change of vectors:

$$\langle U\psi|U\phi\rangle = \langle U^\dagger U\psi|\phi\rangle = \langle\psi|\phi\rangle. \tag{A.26}$$

The trace of an operator is invariant under unitary change of basis in finite dimensions, due to rotational symmetry of trace on matrices.

$$\begin{aligned}&\mathrm{Tr}(A) = \sum_i \langle U\psi_i|AU\psi_i\rangle = \mathrm{Tr}(U^\dagger AU)\\ &= \mathrm{Tr}(UU^\dagger A) = \mathrm{Tr}(A)\end{aligned} \tag{A.27}$$

Definition A.24 (Eigenvalues and eigenvectors)**.** A complex number $\lambda \in \mathbb{C}$ is an *eigenvalue* of a bounded operator $A \in \mathcal{L}(\mathcal{H})$ if there exists a vector $|\lambda\rangle \in \mathcal{H}, |\lambda\rangle \neq 0$ such that $A|\lambda\rangle = \lambda|\lambda\rangle$. The vector $|\lambda\rangle$ is then the *eigenvector* of A associated with the eigenvalue λ.

Definition A.25 (Spectrum of bounded operators)**.** *Spectrum* of an operator $A \in \mathcal{L}(\mathcal{H})$ is the set of all $\lambda \in \mathbb{C}$ such that the operator

$$R(\lambda) = (A - \lambda\mathbb{I})^{-1} \tag{A.28}$$

is not a bounded operator in $\mathcal{L}(\mathcal{H})$.

In finite dimensional case, multiplicity of eigenvalue λ is the dimension of subspace spanned by all eigenvectors associated with this eigenvalue. All eigenvalues of A are in spectrum of A. For finite dimensional $\mathcal{H}$ also the converse holds. If multiplicity of eigenvalue λ is greater than 1, we call this eigenvalue degenerate.

Unitary conjugation preserves the eigenvalues of operator $A \in \mathcal{L}(\mathcal{H})$. Let λ be an eigenvalue of A associated with vector $|\lambda\rangle$. Then $U|\lambda\rangle$ is eigenvector of $UAU^\dagger$ associated with λ:

$$UAU^\dagger U|\lambda\rangle = UA|\lambda\rangle = \lambda U|\lambda\rangle. \tag{A.29}$$

Let $|\lambda_1\rangle, |\lambda_2\rangle$ be eigenvectors of A with associated eigenvalues. Then $\langle\lambda_2|\lambda_1\rangle = c$ and we can write

$$|\lambda_1\rangle = c|\lambda_2\rangle + |\lambda_{2\perp}\rangle, \tag{A.30}$$

where $|\lambda_{2\perp}\rangle$ is orthogonal to $|\lambda_2\rangle$. Since A is self-adjoint we have

$$\langle\lambda_2|A\lambda_{2\perp}\rangle = \lambda_2\langle\lambda_2|\lambda_{2\perp}\rangle = 0. \tag{A.31}$$

Thus $|A\lambda_{2\perp}\rangle = |\lambda'_{2\perp}\rangle$ is orthogonal to $|\lambda_2\rangle$. This leads to

$$\begin{aligned}A|\lambda_1\rangle &= cA|\lambda_2\rangle + A|\lambda_{2\perp}\rangle = c\lambda_2|\lambda_2\rangle + |\lambda'_{2\perp}\rangle,\\ A|\lambda_1\rangle &= \lambda_1|\lambda_1\rangle = c\lambda_1|\lambda_2\rangle + \lambda_1|\lambda_{2\perp}\rangle.\end{aligned} \tag{A.32}$$

Both equalities can be true at the same time only if $c = 0$, states $|\lambda_1\rangle$ and $|\lambda_2\rangle$ are orthogonal, or $\lambda_1 = \lambda_2$. In finite dimensional case eigenvectors of self-adjoint operators associated with

the same eigenvalue λ span a linear subspace whose dimension is equal to multiplicity of λ. You can then find a basis in this subspace and members of this basis will be again eigenvectors. Every self-adjoint operator on finite d-dimensional Hilbert space has d real eigenvalues (counting multiplicity) and therefore also d linearly independent eigenvectors which define some basis $\{|\lambda_i\rangle\}$ on this Hilbert space. Every self-adjoint operator then can be written in this basis as

$$A = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|. \tag{A.33}$$

The trace of a self-adjoint operator is the only the sum of its eigenvalues counting multiplicities.

Example A.26. Let us have operator A

$$A = |0\rangle\langle 0| + |0\rangle\langle 1| + 2|1\rangle\langle 1|. \tag{A.34}$$

This operator is not self-adjoint and has two eigenvectors $|\lambda_1\rangle = |0\rangle$, $|\lambda_2\rangle = |0\rangle + |1\rangle$ associated with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. We see that eigenvectors corresponding to different eigenvalues are not orthogonal for operator which is not self-adjoint.

Example A.27 (Identity)**.** Identity operator is self-adjoint and has only one eigenvalue 1 with multiplicity d, where d is the dimension of underlying Hilbert space $\mathcal{H}$. Naturally we can choose any basis $\{|\psi_i\rangle\}$ in this space and it will be automaticaly basis of eigenvectors and thus

$$\mathbb{I} = \sum_i |\psi_i\rangle\langle\psi_i| \tag{A.35}$$

for any basis in $\mathcal{H}$.

Definition A.28 (Projectors)**.** A projector P is a self-adjoint operator for which $P^2 = P$. Such operator can have only eigenvalues 0 or 1 and any self-adjoint operator with such eigenvalues is a projector. If the multiplicity of eigenvalue 1 is 1 then we call such projector 1-dimensional. Every one dimensional projector can then be written as $|\psi\rangle\langle\psi| =: P_\psi$ for some normed vector $\psi, \| \psi \| = 1$. And every projector is a sum of one dimensional projectors. We call projectors P_1, P_2 orthogonal if $\langle P_1|P_2\rangle_{\mathrm{HS}} = 0 \Leftrightarrow P_1P_2 = O$

Unitary operators can be also decomposed in a nice way. Every unitary operator U can be written as

$$U = \sum_k e^{i\alpha_k} |\psi_k\rangle\langle\psi_k|, \tag{A.36}$$

where $\{|\psi_k\rangle\}$ form orthonormal basis in $\mathcal{H}$ and $e^{i\alpha_k}$ are eigenvalues of U.

B Various

B.1 Monoticity of von Neumann entropy under unital channels

Lemma B.1. *If G^* is a unital channel, then $S\big(G^*(\varrho)\big) \geq S(\varrho)$ for all states ϱ.*

Proof. The proof of entropy monoticity for unital channels is a consequence of the monoticity of the relative entropy [51]. In particular, for arbitrary quantum channel G^*

$$S\big(G^*(\varrho)||G^*(\omega)\big) \leq S(\varrho||\omega)\,, \tag{B.1}$$

where $S(\varrho||\omega) = \mathrm{Tr}\big(\varrho(\log\varrho - \log\omega)\big)$ is the quantum relative entropy. Setting $\omega = \frac{1}{d}\mathbb{I}$ we get $S(\varrho||1/d\mathbb{I}) = -S(\varrho) + \log d$. Using this fact and assuming that G^* is unital the above inequality can be rewritten as

$$\begin{aligned} S\big(G^*(\varrho)||1/d\mathbb{I})\big) &\leq S(\varrho||1/d\mathbb{I}) \\ -S\big(G^*(\varrho)\big) &\leq -S(\varrho)\,, \end{aligned}$$

from which the lemma follows. □

Acknowledgments

I am very grateful for intense and fruitful discussions with Prof. RNDr. Vladimír Bužek, DrSc., Prof. Reinhard F. Werner and doc. Mgr. Mário Ziman, PhD. I would also like to acknowledge the generous financial support of the European Union projects 2010-248095 (Q-ESSENCE), HIP FP7-ICT-2007-C-221889, QAP and also the projects COST action, MP1006, CE SAS QUTE, APVV DO7RP-0002-10, VEGA 2/0127/11 (TEQUDE), OP CE QUTE ITMS NFP 262401022, APVV project QIAM and especially the APVV LPP-0264-07 (QWOSSI).

References

[1] A. Ahlbrecht, F. Richter, and R. F. Werner, *How long can it take for a quantum channel to forget everything?*, International Journal of Quantum Information **10** (2012), no. 05, 1250057.

[2] P. Arrighi, V. Nesme, and R. F. Werner, *One-dimensional quantum cellular automata over finite, unbounded configurations*, Language and Automata Theory and Applications (Carlos Martn-Vide, Friedrich Otto, and Henning Fernau, eds.), Lecture Notes in Computer Science, vol. 5196, Springer Berlin / Heidelberg, 2008, pp. 64–75.

[3] J. Ball, A. Dragan, and K. Banaszek, *Exploiting entanglement in communication channels with correlated noise*, Phys. Rev. A (2004).

[4] A. Bayat, D. Burgarth, S. Mancini, and S. Bose, *Memory effects in spin chain channels for information transmission*, Phys. Rev. A (2008).

[5] G. Bowen, I. Devetak, and S. Mancini, *Bounds on classical information capacities for a class of quantum memory*, Phys. Rev. A (2005).

[6] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics. i+ii.*, Springer, 1979, 1997.

[7] D. Burgarth and K. Maruyama, *Indirect hamiltonian identification through a small gateway*, New Journal of Physics **11** (2009), no. 10, 103019.

[8] D. Burgarth, K. Maruyama, and F. Nori, *Coupling strength estimation for spin chains despite restricted access*, Phys. Rev. A **79** (2009), 020305.

[9] D. Burgarth, K. Maruyama, and F. Nori, *Indirect quantum tomography of quadratic hamiltonians*, New Journal of Physics **13** (2011), no. 1, 013019.

[10] P. Busch, P. Lahti, and P. Mittelstaedt, *The quantum theory of measurement*, Springer, Berlin; New York, 1996.

[11] F. Caruso, V. Giovannetti, C. Macchiavello, and M. B. Ruskai, *Qubit channels with small correlations*, Phys. Rev. A (2008).

[12] N. J. Cerf, J. Clavareau, C. Macchiavello, and J. Roland, *Quantum entanglement enhances the capacity of bosonic channels with memory*, Phys. Rev. A (2005).

[13] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Memory effects in quantum channel discrimination*, Phys. Rev. Lett. **101** (2008), 180501.

[14] D. Daems, *Entanglement-enhanced classical capacity of two-qubit quantum channels with memory: the exact solution*, Phys. Rev. A (2007).

[15] G. M. D'Ariano, *Physics as quantum information processing: quantum fields as quantum automata*, arxiv:1110.6725v4.

[16] A. D'Arrigo, G. Benenti, and G. Falci, *Quantum capacity of a dephasing channel with memory*, New J. Phys. 9 (2007).

[17] N. Datta and T. Dorlas, *The coding theorem for a class of quantum channels with long-term memory*, J. Phys. A: Math. Theor. (2007).

[18] E. B. Davies, *Quantum theory of open sysytems*, Academic Press Inc, London, 1977.

[19] C. Di Franco, M. Paternostro, and M. S. Kim, *Bypassing state initialization in hamiltonian tomography on spin-chains*, SUPP. **1** (2011), 181.

[20] T. Dorlas and C. Morgan, *The classical capacity of quantum channels with memory*, Phys. Rev. A (2009).

[21] Ch. Durr and M. Santha, *A decision procedure for unitary linear quantum cellular automata*, Proceeding of the 37th IEEE Symposium on Foundations of Computer Science, 1996, pp. 38–45.

[22] T. Eggeling, D. Schlingemann, and R. F. Werner, *Semicausal operations are semilocalizable*, Europhys. Lett. **57** (2002), no. 6, 782–788.

[23] R. P. Feynman, *Simulating physics with computers*, Int. J. Theor. Phys. (1982), 467–488, Reprinted in A.J.G. Hey (ed.), Feynman and Computation, Perseus Books 1999.

[24] V. Giovannetti, *A dynamical model for quantum memory channels*, J. Phys. A: Math. Gen. (2005).

[25] V. Giovannetti and S. Mancini, *Bosonic memory channels*, Phys. Rev. A (2005).

[26] D. Gross, V. Nesme, H. Vogts, and R. F. Werner, *Index theory of one dimensional quantum walks and cellular automata*, Communications in Mathematical Physics (2009), 38.

[27] J. Gruska, *Quantum computing*, John Wiley & Sons, Inc., 2007, Section 4.3 treates QCA.

[28] J. Gütschow, T. Rybár, and R. F. Werner, *Memory requirements for general reversible qubit stream processors*, in preparation.

[29] T. Heinosaari and M. Ziman, *The mathematical language of quantum theory: From uncertainty to entanglement*, Cambridge University Press, 2011.

[30] E. L. Hosseini, K. Maruyama, D. Burgarth, T. Takui, Y. Kondo, and M. Nakahara, *Estimation of coupling constants of a three-spin chain: a case study of hamiltonian tomography with nuclear magnetic resonance*, New Journal of Physics **14** (2012), no. 1, 013043.

[31] M. Houshmand, S. Hosseini-Khayat, and M. M. Wilde, *Minimal memory requirements for pearl-necklace encoders of quantum convolutional codes*, IEEE Transactions on Computers (2010), 17–20.

[32] V. Karimipour and L. Memarzadeh, *Entanglement and optimal strings of qubits for memory channels*, Phys. Rev. A (2006).

[33] E. Karpov, D. Daems, and N. J. Cerf, *Entanglement enhanced classical capacity of quantum communication channels with correlated noise in arbitrary dimensions*, Phys. Rev. A (2006).

[34] A. Kossakowski, *On quantum statistical mechanics of non-hamiltonian systems*, Reports on Mathematical Physics **3** (1972), no. 4, 247 – 274.

[35] B. Kraus and J. I. Cirac, *Optimal creation of entanglement using a two-qubit gate*, Phys. Rev. A **63** (2001), 062309.

[36] K. Kraus, *General state changes in quantum theory*, Annals of Physics **64** (1971), no. 2, 311 – 335.

[37] K. Kraus, A. Bohm, J. D. Dollard, and W. H. Wootters, *States, effects, and operations : fundamental notions of quantum theory : lectures in mathematical physics at the university of texas at austin / karl kraus ; edited by a. bohm, j.d. dollard, and w.h. wootters*, Springer-Verlag, Berlin ; New York :, 1983 (English).

[38] D. Kretschmann and R. F. Werner, *Quantum channels with memory*, Phys. Rev. A **72** (2005), no. 6, 062323.

[39] G. Lindblad, *On the generators of quantum dynamical semigroups*, Communications in Mathematical Physics **48** (1976), 119–130, 10.1007/BF01608499.

[40] C. Lupo, V. Giovannetti, and S. Mancini, *Capacities of lossy bosonic memory channels*, Phys. Rev. Lett. (2010).

[41] C. Lupo, L. Memarzadeh, and S. Mancini, *Forgetfulness of continuous markovian quantum channels*, Phys. Rev. A (2009).

[42] C. Lupo, O. V. Pilyavets, and S. Mancini, *Capacities of lossy bosonic channel with correlated noise*, New J. Phys. (2009).

[43] C. Macchiavello and G. M. Palma, *Entanglement-enhanced information transmission over a quantum channel with correlated noise*, Phys. Rev. A (2002).

[44] C. Macchiavello, G. M. Palma, and S. Virmani, *Transition behavior in the channel capacity of two-qubit channels with memory*, Phys. Rev. A (2004).

[45] M. Nielsen and I. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2000.

[46] M. Ozawa, *Conditional expectation and repeated measurements of continuous quantum observables*, Probability Theory and Mathematical Statistics, Lecture Notes in Mathematics, vol. 1021, Springer Berlin / Heidelberg, 1983, 10.1007/BFb0072947, pp. 518–525.

[47] M. Ozawa, *Quantum measuring processes of continuous observables*, Journal of Mathematical Physics **25** (1984), no. 1, 79–87.

[48] M. Paris and J. Řeháček, *Quantum state estimation*, Lecture Notes in Physics, Springer, 2004.

[49] V. I. Paulsen, *Completely bounded maps and dilations*, Cambridge University Press, Cambridge, 2002.

[50] A. Peres, *Quantum theory: Concepts and methods*, Kluwer Academic Publishers, 1995.

[51] D. Petz, *Monotonicity of quantum relative entropy revisited*, Reviews in Mathematical Physics **15** (2003), no. 01, 79–91.

[52] O. V. Pilyavets, V. G. Zborovskii, and S. Mancini, *A lossy bosonic quantum channel with non-markovian memory*, Phys. Rev. A (2008).

[53] M. B. Plenio and S. Virmani, *Spin chains and channels with memory*, Phys. Rev. Lett. (2007).

[54] M. B. Plenio and S. Virmani, *Many body physics and the capacity of quantum channels with memory*, New J. Phys. (2008).

[55] E. Prugovečki, *Information-theoretical aspects of quantum measurement*, International Journal of Theoretical Physics **16** (1977), 321–331, 10.1007/BF01807146.

[56] S. Richter and R. Werner, *Ergodicity of quantum cellular automata*, Journal of Statistical Physics **82** (1996), 963–998, 10.1007/BF02179798.

[57] D. Rossini, V. Giovannetti, and S. Montangero, *Spin chain model for correlated quantum channels*, New Journal of Physics **10** (2008), no. 11, 115009.

[58] G. Ruggeri and S. Mancini, *Privacy of a lossy bosonic memory channel*, Phys. Lett. A (2007).

[59] G. Ruggeri, G. Soliani, V. Giovannetti, and S. Mancini, *Information transmission through lossy bosonic memory channels*, Europhys. Lett. (2005).

[60] V. Scarani, S. Iblisdir, N. Gisin, and A. Acin, *Quantum cloning*, REV.MOD.PHYS. **77** (2005), 1225.

[61] J. Schäfer, D. Daems, E. Karpov, and N. J. Cerf, *Capacity of a bosonic memory channel with gauss-markov noise*, Phys. Rev. A (2009).

[62] B. Schumacher and R. F. Werner, *Reversible quantum cellular automata*, quant-ph/0405174, 2005.

[63] A. J. Scott and M. Grassl, *Symmetric informationally complete positive-operator-valued measures: A new computer study*, Journal of Mathematical Physics **51** (2010), no. 4, 042203.

[64] W. F. Stinespring, *Positive functions on c^* algebras*, Proc. Amer. Math. Soc. **6** (1955), no. 2, 211–216.

[65] W. van Dam, *Quantum cellular automata*, 1996.

[66] J. Watrous, *On one-dimensional quantum cellular automata*, Proceedings of the 36th Annual Symposium on Foundations of Computer Science (Washington, DC, USA), FOCS '95, IEEE Computer Society, 1995, pp. 528–.

[67] M. Wieśniak and M. Markiewicz, *Finding traps in nonlinear spin arrays*, Phys. Rev. A **81** (2010), 032340.

[68] M. Wolf and J. Cirac, *Dividing quantum channels*, Communications in Mathematical Physics **279** (2008), 147–168, 10.1007/s00220-008-0411-y.

[69] J. Wouters, M. Fannes, I. Akhalwaya, and F. Petruccione, *Classical capacity of a qubit depolarizing channel with memory*, Phys. Rev. A (2009).

[70] M. Ziman and T. Rybár, *Quantum finite-depth memory channels: Case study*, Phys. Rev. A **80** (2009), no. 4, 7.

[71] youtu.be/C55F4FKysh0.

Mgr. Tomáš Rybár, PhD. recieved his PhD in physics in 2012 from the Comenius university in Bratislava under supervision of prof. RNDr. Vladimír Bužek, DrSc. and doc. Mário Ziman, PhD. He is currently a postdoctoral research associate at the Institute of theoretical physics on Leibniz University Hannover. His research interests are in foundations of quantum information, open systems and lattice systems.