PHYSICAL REVIEW A **92**, 042315 (2015)

Process estimation in the presence of time-invariant memory effects

Tomáš Rybár[1] and Mário Ziman[2,3]
[1]Institut für Theoretische Physik, Leibniz Universität Hannover, Appelstrasse 2, 30167 Hannover, Germany
[2]Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11 Bratislava, Slovakia
[3]Faculty of Informatics, Masaryk University, Botanická 68a, 60200 Brno, Czech Republic

(Received 31 July 2015; published 13 October 2015)

Any repeated use of a fixed experimental instrument is subject to memory effects. We design an estimation method uncovering the details of the underlying interaction between the system and the internal memory without having any experimental access to memory degrees of freedom. In such case, by definition, any memoryless quantum process tomography (QPT) fails because the observed data sequences do not satisfy the elementary condition of statistical independence. However, we show that the randomness implemented in certain QPT schemes is sufficient to guarantee the emergence of observable "statistical" patterns containing complete information on the memory channels. We demonstrate the algorithm in detail for the case of qubit memory channels with two-dimensional memory. Interestingly, we find that for the arbitrary estimation method, the memory channels generated by controlled unitary interactions are indistinguishable from memoryless unitary channels.

DOI: 10.1103/PhysRevA.92.042315 PACS number(s): 03.67.−a, 06.20.Dk, 02.50.Cw

I. INTRODUCTION

Repeatability of experiments is one of the main conceptual paradigms of modern science, although its meaning has evolved over time. In particular, the quantum experiments are not repeatable in a strict sense of individual observations (e.g., no one knows whether or not a given photon passes the polarizer); however, the repeated runs of such experiments exhibit repeatable statistical patterns (e.g., the fraction of photons passing the polarizer is fixed). In other words, quantum theory does not give a clear conceptual meaning (in the sense of repeatability) to individual outcomes, but rather to numbers represented by averages and probabilities.

Therefore, the interpretation of quantum experiments is intimately related to our understanding of probabilities, especially with the question of whether the observed frequencies are really the probabilities occurring in theoretical models of the experiments. In any case, the repeatability of statistical features assumes that individual runs of the experiment are independent. In theory, this means that each run of the experiment is performed with "fresh" apparatuses (under exactly the same conditions); however, in practice, we do not really employ a new apparatus every time the experiment is run. Instead, it is implicitly assumed that the internal relaxation processes are sufficiently fast to refresh the whole experimental setup. But is such assumption justified?

Consider an experiment in which a quantum particle is sent through a quantum channel. While the particle is transferred, it interacts with the degrees of freedom of the channel. According to quantum theory, these interactions are described by the Schrödinger equation and result in a unitary transformation of the joint particle-channel system. As a result, both the particle and the channel are disturbed by this interaction and the disturbances depend on their original characteristics. Consequently, the repeated use of the same channel device is not independent of the previous uses; thus, the induced particle transformation will be typically different. If this is the case, we say the channel exhibits memory effects. Let us stress that all of the relaxation processes can be incorporated into this unitary model by extending the size of the memory.

Indeed, suppose the channel is just "delaying" the transfer of the particles, i.e., its nth output equals the $(n-1)$th input (first output is set to be in some fixed state). In this case, the uses are clearly not independent. This can be demonstrated if one's goal is to estimate the parameters of the quantum process assuming the channel devices are memoryless. Then different (equivalent in the memoryless case) estimation procedures could lead to different conclusions. In particular, if the channel action is tested in an "ordered" fashion, i.e., we first analyze how the state ϱ_1 is transformed to ϱ_1', then ϱ_2 to ϱ_2', etc., then any delay vanishes in the statistical analysis and we must conclude the transformation is noiseless, i.e., $\varrho \mapsto \varrho' = \varrho$. However, if the channel is tested in a "random" fashion, i.e., in each run a random test state is used, then for each fixed input ϱ, the output state ϱ' is a fixed state ϱ_0 being the average input test state; thus, the channel is recognized as the maximal noise and therefore not very useful for the transfer *per se*.

In the described case, the action of the memory is quite simple and, when cleverly used, this memory device can be used to transfer information in a noiseless way [1]. But how can one find out the action if the interaction is not known in advance? How can one proceed in order to detect such memory behavior and, finally, exploit the memory for our purposes? Exactly these questions will be addressed in this work. It is organized as follows. We start with introducing all of the necessary concepts and tools in Secs. II and III. In Sec. IV, we state the problem. In Sec. V, we solve this problem in the special case of control unitary interactions, and in Sec. VI we formulate theorems allowing us to design the estimation algorithm. Finally, in Sec. VII, we illustrate in detail the algorithm for the simplest possible case of qubit channels with a two-dimensional memory.

II. PRELIMINARIES

States ϱ of quantum systems are identified with the set of density operators $\mathcal{S}(\mathcal{H})$ being positive linear operators on a

Hilbert space $\mathcal{H}$ of a unit trace. Measurement apparatuses M are described by positive operator valued measures (POVM) being a set of positive operators $E_1,\ldots,E_m$ such that $O \leqslant E_j \leqslant I$ (called *effects*) and $\sum_j E_j = I$. Each measurement outcome is associated with exactly one effect and we write $E_k \in M$ if E_k is an effect associated with one of the outcomes M. Quantum channels $\mathcal{E}$ describing the memoryless processes are identified with completely positive trace-preserving linear maps defined on the set of trace-class operators $\mathcal{T}(\mathcal{H})$. In particular, $\mathcal{E} : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$. If $\mathcal{E}(\varrho) = U\varrho U^\dagger$ for some unitary operator U, then we say the channel is unitary and we denote it by $\mathcal{U}$. Due to the Stinespring theorem, any channel can be understood as a result of a unitary interaction between the system and some (initially factorized) memory, i.e., $\mathcal{E}(\varrho) = \mathrm{tr}_{\mathcal{M}}[\mathcal{U}(\xi \otimes \varrho)]$, where ξ is the initial state of the memory and $\mathcal{U} : \mathcal{S}(\mathcal{M} \otimes \mathcal{H}) \to \mathcal{S}(\mathcal{H} \otimes \mathcal{M})$ (for more details, see, for instance, Ref. [2]).

When modeling (see Refs. [1,3]) the experiment with the memory process device (used repeatedly), we will assume that its action is described by a fixed unitary channel and includes all of the relaxation processes of the memory. Also we assume that we do not have any access to memory degrees of freedom; thus, when we want to learn something about the underlying process $\mathcal{U}$, we can only manipulate the system, eventually employing some ancillary systems and devices. The experiment gives rise to a sequence of channels $\mathcal{E}^1,\ldots,\mathcal{E}^n$ for n uses of the device, defined as follows:

$$\mathcal{E}^j(\varrho^j) = \mathrm{tr}_{\mathcal{M}}[\mathcal{U}^j(\xi \otimes \varrho^j)],$$

where $\mathcal{U}^j = \mathcal{U}_j \cdots \mathcal{U}_1$ is the j-fold concatenation of $\mathcal{U}_k$ and ϱ^j is the joint input state describing first j uses of the memory device. The unitary channel $\mathcal{U}_k$ acts as $\mathcal{U}$ on the memory and the kth system, and trivially elsewhere.

Let us stress that by construction, the sequence of processes $\mathcal{E}^1,\ldots,\mathcal{E}^n$ is causal (jth output does not depend on kth input for $k > j$), and thus, $\mathcal{E}^j(\varrho^j) = \mathrm{tr}_{j+1}[\mathcal{E}^{j+1}(\varrho^{j+1})]$. Indeed, due to the seminal paper by Kretschmann and Werner [1], every causal memory process can be represented as a concatenation of unitary channels describing the sequence of interactions between the memory and the processed systems. In our case, the memory channel is also time invariant (see Fig. 1), i.e., the unitary channels applied in concatenations coincide.

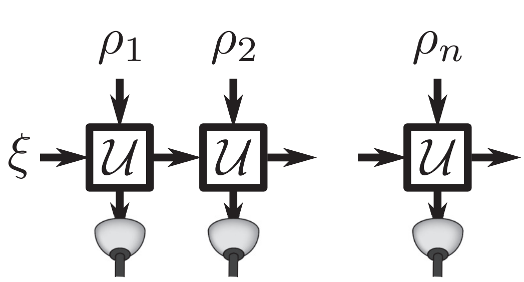

FIG. 1. Repeated uses of the time-invariant memory process identified with the unitary channel $\mathcal{U}$ describing the interaction between the device inputs ϱ_j and experimentally inaccessible memory degrees of freedom initially in an unknown state ξ.

III. QUANTUM PROCESS TOMOGRAPHY

Quantum process tomography (QPT) is any processing of experimental data uniquely identifying an unknown memoryless quantum channel [4,5]. It is known to be a complex task; however, under certain assumptions, it can also be efficiently applied to large systems [6–10] and even the accuracy can be assessed [11–14]. QPT deals with a scenario where an experimenter is given an unknown input-output black box $\mathcal{E}$. In each run of the experiment, he prepares some test state ϱ and performs a measurement M; thus, he chooses the setting $x = (\varrho,M)$ and records the outcome E_k, where $E_k \in M$. Let us denote by $X = \{(\varrho_x,M_x)\}_x$ the set of all possible settings. The measurement M_x is described by effects E_{xk}, and N_x labels the total number of times the setting x was chosen, i.e., $\sum_x N_x = n$. In each run of the experiment, we observe an event $x_k = (\varrho_x, E_{xk})$ indicating that the setting $x = (\varrho_x,M_x)$ is used and the outcome E_{xk} is recorded. The conditional probability of observing the event x_k is given by the formula $p(x_k|\mathcal{E}) = q_x \mathrm{tr} E_{xk}\mathcal{E}[\varrho_x]$, where $q_x = N_x/n$ describes the frequency of the setting x. For suitable choice of X, this probability distribution $p(x_k|\mathcal{E})$ enables us to reveal the identity of the channel $\mathcal{E}$. Conceptually, the simplest example consists of a linearly independent collection of test states $\{\varrho_x\}_x$ and a fixed state tomography measurement M (the same for each x).

Clearly, the ordering of the events $x_1,\ldots,x_n$ is irrelevant for QPT and only their fraction is needed. However, this is true only if the condition of memoryless channel is met, i.e., when a fresh copy of the channel is used each time the experiment is made. Otherwise, the QPT procedure may lead to wrong conclusions. Suppose we have tested a communication channel (the delaying channel from Sec. I) using the well-ordered sequence of settings and find out the transfer is just perfect; thus, we use it to built a noiseless worldwide communication network. However, the communication itself is quite far from a well-ordered sequence of symbols. It is much closer to a random one and, for such, the considered communication device does not work at all; hence, the seemingly "perfect" network fails dramatically. On the other side, the usage of a random sequence of settings leads to a conclusion that the communication device is of no use. But this is also not true because shifting the outputs by one results in perfect transmission.

IV. FORMULATION OF THE PROBLEM

Our task is to capture the dynamics underlying the memory process, i.e., the interaction $\mathcal{U}$ and the memory state ξ. However, as we have only a single copy of the state ξ, learning any nontrivial information on ξ is forbidden by the no-cloning theorem [15]. Moreover, not all of the parameters of $\mathcal{U}$ are accessible within our model. In particular, the output of the memory channel given by $(\mathcal{U},\xi)$ is the same as that of $[(\mathcal{I} \otimes \mathcal{V}_M)\mathcal{U}(\mathcal{V}_{\mathcal{M}}^{-1} \otimes \mathcal{I}), \mathcal{V}_{\mathcal{M}}\xi\mathcal{V}_{\mathcal{M}}^{-1}]$, for some unitary $\mathcal{V}_{\mathcal{M}} : \mathcal{M} \mapsto \mathcal{M}$. In conclusion, our goal is to estimate $\mathcal{U}$ modulo this freedom under the condition that the initial state of the memory is unknown and the memory is experimentally inaccessible.

Before we proceed, let us stress that (just like in the memoryless case) we are able to predict probabilities, however, by construction, our experiments cannot be repeated in the statistical sense; hence, the standard tools and methods of statistical analysis are simply inapplicable. In full generality of the problem, we are free to choose the input state for a given number n of uses of the device, we can choose the output measurements, and we may also employ some ancillary systems.

V. CONTROLLED UNITARY INTERACTIONS

In this example, we will show a family of memory channels for which the freedom in the estimation of the interaction $\mathcal{U}$ is much larger. We say the interaction is controlled unitary if it can be written in the following form [16]: $\mathcal{U}^{\rm ctrl} = \sum_l |l\rangle_{\mathcal{M}}\langle l| \otimes \mathcal{V}_l$, where $\mathcal{V}_l$ are arbitrary unitary channels defined on the system and vectors $|l\rangle$ form an orthonormal basis of the memory Hilbert space.

Theorem 1. The memory device induced by a controlled unitary interaction $\mathcal{U}^{\rm ctrl}$ is indistinguishable from a memoryless unitary device.

Proof. Suppose ϱ^n is the joint state of n inputs and let $\xi_{\mathcal{M}}$ be the initial state of the memory. Then, $\varrho^{n\prime} = \sum_l q_l \mathcal{V}_l^{\otimes n}(\varrho^n)$ with $q_l = \langle l|\xi|l\rangle$. Suppose E is an effect on n outputs such that $p_E(\mathcal{U}^{\rm ctrl}) = {\rm tr}[E\mathcal{U}_n^{\rm ctrl}(\varrho^n)] > 0$. Then, for the same input state ϱ^n, also $p_E(\mathcal{V}_l) = {\rm tr}[E\mathcal{V}_l^{\otimes n}(\varrho^n)] > 0$ for some l, and thus for any test state the observation of the individual outcome E cannot be used to distinguish $\mathcal{U}^{\rm ctrl}$ from $\mathcal{V}_l$ (for a suitable l). ■

In other words, any estimation procedure for this class of channels results randomly (with probability q_l) in one of the unitaries $\mathcal{V}_l$.

VI. ESTIMATION ALGORITHM

The algorithm we are going to explain is based on the QPT method with randomly chosen settings (see Fig. 2). In particular, in each run of the experiment, the setting $x = (\varrho_x, M_x)$ is selected independently with the probability q_x. Let us remind the reader that among n uses of the channel, approximately $N_x \approx q_x n$ times the setting x is selected. Denote by N_{xk} the number of occurrences of the event (ϱ_x, E_{xk}) (with E_{xk} being the effect observed in the measurement M_x) and define a number $\tilde{p}(k|x) = N_{xk}/N_x$ (playing the role of conditional probabilities in the case of QPT). The following theorem provides the basis for the statistical interpretation of this number.

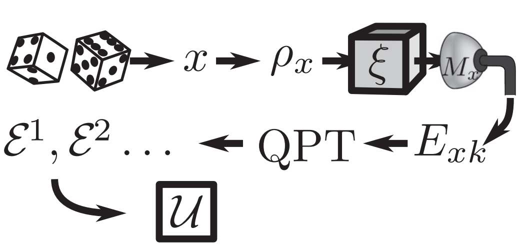

FIG. 2. Schematic illustration of the estimation method. Setting x is chosen at random and outcome E_{xk} is observed. Collecting this data and performing process tomography (QPT) yields a family of channels $\mathcal{E}^n$ on n subsequent inputs. From these channels, the interaction $\mathcal{U}$ is determined up to local unitary rotation of the memory system.

Theorem 2. If QPT is implemented with randomly chosen settings, then for all settings x there exists a state of the memory $\overline{\xi} \in \mathcal{S}(\mathcal{M})$ such that

$$\lim_{n\to\infty} \tilde{p}(k|x) = p(x_k|x) \equiv {\rm tr}[\mathcal{U}(\overline{\xi} \otimes \rho_x)(E_{xk} \otimes I_{\mathcal{M}})].$$

Consequently, we may treat $p(x_k|x)$ as the conditional probability $p(x_k|\mathcal{E})$ for some *average* channel $\mathcal{E}(\varrho) = {\rm tr}_{\mathcal{M}}[\mathcal{U}(\overline{\xi} \otimes \varrho)]$ induced by the state $\overline{\xi}$; hence, QPT reconstruction results in the memoryless channel $\mathcal{E}$.

Proof. Let us denote by ξ_j the state of the memory before the jth run of the experiment leading to observation of some effect $E_{x_j k}$. During the algorithm, the memory system undergoes a sequence of transformations,

$$\xi \equiv \xi_1 \mapsto \xi_2 \mapsto \cdots \mapsto \xi_n. \quad (1)$$

Denote by $\mathfrak{S}$ the set of all states $\{\xi_j\}_j$ occurring in the sequence and by $\mathfrak{S}_x$ a subset of $\mathfrak{S}$ for which the setting x was used. Consider a partitioning of $\mathcal{S}(\mathcal{M})$ into mutually exclusive subsets $\{\mathcal{X}_\mu\}_\mu$, i.e., $\mathcal{X}_\mu \cap \mathcal{X}_\nu = \emptyset$ and $\mathcal{S}(\mathcal{M}) = \bigcup_\mu \mathcal{X}_\mu$. Define $p(\mathcal{X}_\mu) = |\mathfrak{S} \cap \mathcal{X}_\mu|/n$ and $p_x(\mathcal{X}_\mu) = |\mathfrak{S}_x \cap \mathcal{X}_\mu|/|\mathfrak{S}_x|$ determining the frequency of the memory state being from the subset $\mathcal{X}_\mu$ and the frequency being from $\mathcal{X}_\mu$ conditioned on the settings x, respectively. As the choice of the setting x is random, the states $\xi_j \in \mathcal{X}_\mu$ are distributed between the sets $\mathfrak{S}_x$ at random with probability q_x; hence, the subset $\mathfrak{S}_x$ is a random sample of $\mathfrak{S}$. Formally, $|\mathfrak{S}_x \cap \mathcal{X}_\mu| \approx q_x|\mathfrak{S} \cap \mathcal{X}_\mu|$ for large n. Consequently, for all x, we obtain the relation $p_x(\mathcal{X}_\mu) \approx p(X_\mu)$, i.e., for any partitioning the conditional distribution $p_x(\mathcal{X}_\mu)$ is (in the limit of large n) independent of the initial settings x. In other words, whatever initial setting is used, the average memory state $\overline{\xi}_x$ is fixed and $\overline{\xi}_x = \overline{\xi}$. Therefore, for each x, the observed transformation is $\varrho_x \mapsto \varrho_x' = (1/n)\sum_{\mathfrak{S}_x} {\rm tr}_{\mathcal{M}}\mathcal{U}(\xi_l \otimes \varrho_x) \equiv \mathcal{E}(\varrho_x)$ with $\mathcal{E}(\varrho_x) = {\rm tr}_{\mathcal{M}}[\mathcal{U}(\overline{\xi} \otimes \varrho_x)]$. ■

Note that a trivial implication of this result is that the average channel on n subsequent inputs reads $\mathcal{E}^n(\rho^n) = {\rm tr}_{\mathcal{M}}[\mathcal{U}^n(\overline{\xi} \otimes \rho^n)]$ and corresponds to the probabilities of n joint events. This theorem enables us to interpret the result of any QPT method with randomly chosen settings; however, it does not tell us what the generating state $\overline{\xi}$ is. When the channel $\mathcal{E}$ is reconstructed, we know that the interaction $\mathcal{U}$ is one of its dilations. The following theorem characterizes the average state $\overline{\xi}$.

Theorem 3. The average state $\overline{\xi}$ is a fixed point of the so-called (average) concurrent channel $\mathcal{C}(\xi) = {\rm tr}_{\mathcal{H}}[\mathcal{U}(\xi \otimes \overline{\varrho})]$, where $\overline{\varrho} = \sum_x q_x \varrho_x$ is the average test state, i.e., $\mathcal{C}(\overline{\xi}) = \overline{\xi}$.

Proof. Let f be the measure on $\mathcal{S}(\mathcal{M})$ characterizing the distribution of states in the set $\mathfrak{S}$ for large n, i.e., $\int_{\mathcal{X}_\mu} df(\xi) \approx |\mathcal{X}_\mu|$ and $\overline{\xi} \approx \int_{\mathcal{S}(\mathcal{M})} df(\xi)\xi$. Given that state ξ enters a collision with state ρ_x and the measured output is E_{xk}, the exiting state of memory is $\xi_{\rm out} = \mathcal{I}_{xk}[\xi]/{\rm tr}(\mathcal{I}_{xk}[\xi])$ where $\mathcal{I}_{xk}[\xi] = {\rm tr}_{\mathcal{H}}[\mathcal{U}(\xi \otimes \rho_x)(E_{xk} \otimes I_{\mathcal{M}})]$. The probability of the event (ρ_x, E_{xk}) is $q_x{\rm tr}(\mathcal{I}_{xk}[\xi])$, where q_x is the probability of setting (ρ_x, M_x). The average over $\mathfrak{S}$ can be expressed as the average over exiting states $\xi_{\rm out}$ (for inputs ξ distributed according to μ), and thus,

$$\overline{\xi} = \sum_{xk} \int_{\mathcal{S}(\mathcal{M})} df(\xi) \frac{q_x \mathrm{tr}(\mathcal{I}_{xk}[\xi])\mathcal{I}_{xk}[\xi]}{\mathrm{tr}(\mathcal{I}_{xk}[\xi])} = \mathcal{C}(\overline{\xi}).$$

■

It follows that in order to identify the memory process, it remains to characterize the mapping $\mathcal{C}$, in particular, its fixed points. Moreover, if the fixed point ξ_0 is unique, then $\overline{\xi} = \xi_0$ (for large n) and the reconstruction procedure is clear. However, in general, we do not know how to test the uniqueness of a fixed point. The proposed estimation procedure identifies jointly both the memory process $\mathcal{U}$ and the state $\overline{\xi}$. In what follows, we will illustrate in detail the essence of the proposed memory estimation procedure.

To set up our experiment, let us select a set of linearly independent test states $\{\rho_x\}_{x=1}^r$ and measurements $\{M_x\}_{x=1}^r$ occurring with probabilities $\{q_x\}_{x=1}^r$, respectively. By performing the QPT with inputs randomly selected according to probabilities q_x, we obtain a sequence of events $x^n = \{(\rho_{x_j}, E_{x_j k_j})\}_{j=1}^n$. Let us denote by $N(y^m)$ the number of occurrences (counting overlaps) of a subsequence of events $y^m = \{(\rho_{y_j}, E_{y_j l_j})\}_{j=1}^m$ of a given length m. Assuming $m \ll n$ and $n \to \infty$, Theorem 2 (see the implication below) guarantees the existence of $\overline{\xi}$ such that

$$\frac{N(y^m)}{n-m+1} \approx p(y^m), \tag{2}$$

where

$$p(y^m) \equiv \mathrm{tr}\big[\mathcal{E}^m\big(Y_\rho^m\big)Y_E^m\big]p\big(Y_\rho^m\big) \tag{3}$$

with

$$\mathcal{E}^m\big(Y_\rho^m\big) = \mathrm{tr}_{\mathcal{M}}\big[\mathcal{U}^m\big(\overline{\xi} \otimes Y_\rho^m\big)\big], \tag{4}$$

and $Y_\rho^m = \bigotimes_{j=1}^m \rho_{y_j}$ is the concatenation of inputs in the subsequence y^m. Similarly, $Y_E^m = \bigotimes_{j=1}^m E_{y_j l_j}$ is the concatenation of observed effects in the subsequence y^m and $p(Y_\rho^m) = \prod_{j=1}^m q_{y_j}$ is the probability of Y_ρ^m to occur.

A family of subsequences $\{y_1^{m_1}, \ldots, y_2^{m_l}\}$ is called tomographically complete if the observed family of probability distributions $\{p(y_1^{m_1}), \ldots, p(y_l^{m_l})\}$ faithfully identifies a unique memory process (associated with U). Any tomographically complete family $\{y_1^{m_1}, \ldots, y_2^{m_l}\}$ constitutes a valid estimation procedure. Clearly, for practical purposes, it is better to keep the maximal length $m = \max\{m_1, \ldots, m_l\}$ as small as possible and the question on minimal value of m is of interest. Also let us stress that in general, it is an open problem how to verify whether a given family of subsequences is tomographically complete.

Further, we will consider a special class of random QPT methods for which we can make an "educated guess" for $\overline{\xi}$. In particular, suppose that QPT consist of test states ϱ_x satisfying $\overline{\varrho} = \sum_x q_x \varrho_x = \frac{1}{d} I$ and of a fixed informationally complete measurement (i.e., $M_x = M$ for all x). In such case, the map $\mathcal{C}(\xi) = \frac{1}{d}\mathrm{tr}_{\mathcal{H}}[\mathcal{U}(\xi \otimes I)]$ is unital $[\mathcal{C}(I) = I]$. It follows that for any QPT with the average input state being the complete mixture, the state $\overline{\xi} = \frac{1}{d} I$ is a fixed point of $\mathcal{C}$. It remains to analyze whether there are some other fixed states. If it is unique, then the QPT reconstruction is straightforward. In the following section, we will explicitly exploit this class of QPT methods and design an estimation procedure for the case of qubit memory channels with two-dimensional memory.

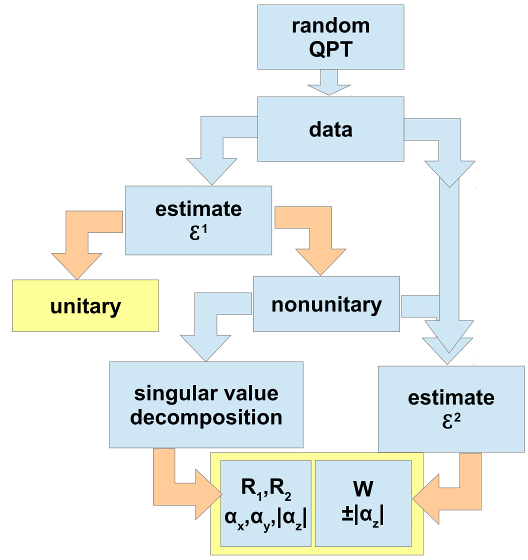

FIG. 3. (Color online) Estimation algorithm for the qubit case.

VII. QUBIT MEMORY CHANNEL WITH TWO-DIMENSIONAL MEMORY

In this section, we will illustrate an example of an estimation procedure for the simplest case of a single-qubit memory process with a two-dimensional memory. Luckily, this case can be treated analytically (see Fig. 3 for a schematic view of the algorithm). As we said before, we will consider test states $\{\rho_x\}_{x=1}^r$ such that $\overline{\varrho} = \frac{1}{2} I$. The most general unitary operator U in the considered qubit-qubit case is parametrized as follows [17]:

$$U = (W_2 \otimes V_2) D(\vec{\alpha})(W_1 \otimes V_1)\,, \tag{5}$$

where V_1, V_2, W_1, W_2 are single-qubit unitaries and

$$D(\vec{\alpha}) = \exp \frac{1}{2} \sum_j \alpha_j \sigma_j \otimes \sigma_j, \tag{6}$$

where σ_j are Pauli operators and $0 \leqslant |\alpha_z| \leqslant \alpha_y \leqslant \alpha_x \leqslant \pi/2$. Note that remark [16] also applies here. Taking into account the equivalence class of unitaries due to the unitary conjugation on the memory system, it follows that memory channels can be parametrized as

$$U = (W \otimes V_2) D(\vec{\alpha})(I \otimes V_1), \tag{7}$$

where W is arbitrary single-qubit unitary operator.

Further, let us analyze fixed points (uniqueness of $\frac{1}{2} I$ as a fixed point) of the concurrent map induced by such unitary interactions. The average concurrent map $\mathcal{C}(\xi) = \mathrm{tr}_{\mathcal{H}}[U(\xi \otimes \frac{1}{2} I)]$ is independent of V_1 and V_2 and can be written as a composition $\mathcal{C} = \mathcal{W}\mathcal{D}(\vec{\alpha})$, where $\mathcal{D}(\vec{\alpha}) = \mathrm{tr}_{\mathcal{H}}[D(\vec{\alpha})(\xi \otimes \frac{1}{2} I)]$. Clearly, if the state $\frac{1}{2} I$ is the unique fixed point of $\mathcal{D}(\vec{\alpha})$, then it is also the only fixed point of the concurrent channel $\mathcal{C}$. Therefore, it is sufficient to start our analysis with fixed points of $\mathcal{D}(\vec{\alpha})$.

Representing ξ in the basis of Pauli operators $I,\sigma_x,\sigma_y,\sigma_z$, i.e., $\xi = \frac{1}{2}(I + \vec{s}\cdot\vec{\sigma})$, the action of $\mathcal{D}(\vec{\alpha})$ can be written in the form [18]

$$\vec{s} \mapsto D\vec{s} + \vec{t}, \tag{8}$$

with

$$D = \begin{pmatrix} c_{2y}c_{2z} & 0 & 0 \\ 0 & c_{2z}c_{2x} & 0 \\ 0 & 0 & c_{2x}c_{2y} \end{pmatrix}, \tag{9}$$

where, due to unitality, $\vec{t} = \vec{0}$. We used a shorthand notation, $c_{nj} = \cos(n\alpha_j)$. Clearly, $\vec{s} = \vec{0}$ is always the fixed point of D. It is not unique if at least two of the parameters c_{2x},c_{2y},c_{2z} are equal to one. For our choice of parametrization, it means $c_{2z} = c_{2y} = 1$, and hence, $\alpha_z = \alpha_y = 0$. In such case, $D = \mathrm{diag}\{1,c_{2x},c_{2x}\}$ and any state of the form $\xi = \frac{1}{2}(I + x\sigma_x)$ is preserved. This corresponds to a family of so-called pure-decoherence channels [19] preserving the diagonal entries of density operators expressed in the decoherence basis (eigenbasis of σ_x in our case); thus, U is a controlled-unitary interaction (with the memory playing the role of the control system), which we have shown in Sec. V to be indistinguishable from factorized unitary interactions. We made the same conclusion when $c_{2x} = c_{2y} = c_{2z} = 1$, and hence, D describes the identity map. In other words, the interaction U can be concluded to be factorized whenever the channel $\mathcal{C}$ has more fixed points.

Suppose now that at least two of the angles α_j (of D) are nonvanishing; thus, $\frac{1}{2}I$ is the unique fixed point of the channel $\mathcal{C}$. From the frequency of single events $N(y^1)$, we can estimate the (unital) channel $\mathcal{E}^1$ transforming the initial state $\varrho = \frac{1}{2}(I + \vec{r}\cdot\vec{\sigma})$ into the state $\varrho' = \frac{1}{2}(I + \vec{r}'\cdot\vec{\sigma})$, with $\vec{r}' = A\vec{r}$ and

$$A = R_2 B(\vec{\alpha}) R_1 \tag{10}$$

is the singular value decomposition of A with R_j being orthogonal rotations induced by unitaries V_j and, due to the symmetry of $U(\vec{\alpha})$ with respect to change of the system and memory, the matrix $B(\vec{\alpha}) = \mathrm{diag}\{c_{2y}c_{2z},c_{2z}c_{2x},c_{2x}c_{2y}\}$ coincides with D given in Eq. (9). Let us stress that $\mathcal{E}^1$ is independent of W; thus, the reconstructed channel $\mathcal{E}^1$ contains no information on W. The estimation of W will be treated later. If the singular values are nondegenerate, then V_1,V_2 are uniquely determined. However, in the case of degeneracy, not all parameters of R_1,R_2 can be accessed and there is an ambiguity in their specification. In summary, the estimated transformation A contains complete information on local unitaries V_1,V_2 and the singular values forming $B(\vec{\alpha})$ allow us to specify the angles α_x,α_y, but only the absolute value of $|\alpha_z|$ (see our choice of parametrization).

In order to determine the sign of $|\alpha_z|$ and the unitary W, we need to reveal some of the properties of $\mathcal{E}^2$, i.e., of the action of the memory process on two inputs. We can reconstruct this channel from the frequencies of double events $N(y^2)$; thus the action of $\mathcal{E}^2$ on arbitrary input is known to us.

Define operators $S_j = V_1\sigma_j V_1^\dagger$ and $T_j = V_2\sigma_j V_2^\dagger$ and let us parametrize W as

$$W = \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix}, \tag{11}$$

where $z = re^{i\phi}$, $w = \sqrt{1-r^2}e^{i\psi}$, and $r \in [0,1]$. Then the parameters of W and sign of α_z can be extracted out of the following equations:

$$\begin{aligned}
a &= \mathrm{tr}[\mathcal{E}^2(S_z\otimes I)(I\otimes T_z)] = 4(2r^2-1)s_{2x}^2 s_{2y}^2,\\
b &= \mathrm{tr}[\mathcal{E}^2(S_x\otimes I)(T_y\otimes T_z)] = 2(1-2r^2)s_{2z}s_{2x}s_{4y},\\
c &= \mathrm{tr}[\mathcal{E}^2(S_z\otimes S_x)(I\otimes T_z)] = 4r\sqrt{1-r^2}s_+ s_{4x}s_{2y}^2,\\
d &= \mathrm{tr}[\mathcal{E}^2(S_z\otimes S_y)(I\otimes T_z)] = 4r\sqrt{1-r^2}c_+ s_{4y}s_{2x}^2,\\
e &= \mathrm{tr}[\mathcal{E}^2(S_z\otimes I)(T_x\otimes T_z)] = 4r\sqrt{1-r^2}s_- s_{4x}s_{2y}^2,\\
f &= \mathrm{tr}[\mathcal{E}^2(S_z\otimes I)(T_y\otimes T_z)] = 4r\sqrt{1-r^2}c_- s_{4y}s_{2x}^2,
\end{aligned} \tag{12}$$

where $s_\pm = \sin(\psi \pm \phi)$ and $c_\pm = \cos(\psi \pm \phi)$. The left-hand sides of these equations can be computed from the data because V_1,V_2 are known from the analysis of $\mathcal{E}^1$. So far we have silently assumed that the singular value decomposition was not degenerate in (10). In the case of degenerate singular values, at least two angles in $\vec{\alpha}$ are equal in absolute value. For example, let $\alpha_x = \alpha_y$. Then, $\mathcal{D}(\vec{\alpha}) = (e^{i\beta\sigma_z}\otimes e^{i\beta\sigma_z})\mathcal{D}(\vec{\alpha})(e^{-i\beta\sigma_z}\otimes e^{-i\beta\sigma_z})$. Hence the unitary $(\mathcal{W}\otimes\mathcal{V}_2)\mathcal{D}(\vec{\alpha})(I\otimes\mathcal{V}_1)$ is equivalent to $[(e^{i\beta\sigma_z}\mathcal{W}e^{-i\beta\sigma_z})\otimes(\mathcal{V}_2e^{-i\beta\sigma_z})]\mathcal{D}(\vec{\alpha})[I\otimes(e^{i\beta\sigma_z}\mathcal{V}_1)]$. Therefore, we can fix the resulting freedom by choosing some V_i in Eq. (10) and then the unitary W is computed with respect to this choice. The only remaining case is when $A = 0$ in (10). In this case, the memory is "essentially" induced by the swap gate (see the Appendix for the detailed analysis of this case).

VIII. SUMMARY

We have proposed an estimation method for estimating the underlying system-memory interaction U generating the memory channel assuming that this interaction is time invariant. The algorithm is based on a random implementation of the arbitrary memoryless quantum process tomography (QPT) procedure. We proved that arbitrary memoryless QPT (implemented with random settings) results in some memoryless channel $\mathcal{E}$ with a dilation being the system-memory interaction U. Moreover, when the average state of the memory (during QPT) is known, the correct identification of the interaction (among the unitary dilations of $\mathcal{E}$) is possible. We proved this happens when the average testing state is chosen to be the complete mixture and the average concurrent channel is strictly contractive. In particular, in this case, the average memory channel is the complete mixture as well; hence, the reconstructed channel $\mathcal{E}$ is necessarily unital. The reconstruction method is illustrated for qubit memory channels with two-dimensional memories (see Fig. 3). The proposed algorithm was successfully implemented and tested numerically.

The reconstruction procedure can be extended for systems and memories of arbitrary size; however, it is an open question as to what size of concatenation $\mathcal{E}^n$ is sufficient for completing the estimation of $\mathcal{U}$. Let us note at the end that the unitary evolution is not a requirement; one only needs to assume *linearity*. Hence, for that matter, $\mathcal{U}$ can be an arbitrary channel without any change in the arguments made. However, this enlarges the complexity of the task enormously, even in the memoryless case, because of the large number of free parameters.

The presented estimation method is universal; however, it is neither the most general one and likely nor the optimal one. We believe that our abilities to treat the concept of memory channels in experiments provide us with better understanding and control of quantum apparatuses and, therefore, they are not only of a deep foundational interest but have direct application. Conceptually, this work is challenging our understanding and interpretation of elementary scientific tools: the repeatability and the probability. Practically, the problem is intimately related to a single-copy estimation of matrix product states and the results may be applied for the characterization of Hamiltonians [20,21] of a single-copy many-body system. In particular, our scenario and results cover the case of a sequence of repeated measurements on the subsystem followed by the system's evolution (for a fixed time interval) as considered in [22,23]. In their case, they develop central limit theorems for the non–independently-and-identically-distributed distribution of outcomes under the condition that the concurrent channel $\mathcal{C}$ is mixing.

ACKNOWLEDGMENTS

This work has been supported by Projects FP7-ICT 2013-600645 (SIQS), APVV-0808-12 (QIMABOS), and VEGA 2/0125/13 (QUICOST). T.R. acknowledges the support of FP7-IDEAS-ERC 291401 (DQSIM). M.Z. acknowledges the support of FP7-ICT 2013-323970 (RAQUEL) and GACR Project No. P202/12/1142.

APPENDIX: DELAYING CHANNEL EXAMPLE

In this section, we are going to thoroughly examine a memory channel generated by a special kind of unitary—the swap operation $\mathcal{S}(a \otimes b) = b \otimes a$; see Fig. 4. In this case, the ith output is the $(i-1)$th input. To analyze this channel, we are going to use a finite set of test states $\rho(x)$, $x \in \{1,\dots,k\}$, and a single informationally complete POVM with effects E_y, $y \in 1,\dots,l$. We use the notation ρ_i to refer to the time ordering of the inputs. The state of the ith input could, in principle, depend on time; hence, $\rho_i = \rho(x_i)$. Due to the delaying nature of the channel, the probability of observing outcome y at the ith turn is $p[y|\rho(x_i)] = \mathrm{tr}[E_y \rho(x_{i-1})]$ and is completely independent of ρ_i. In this notation, we understand that $\rho_0 = \xi$

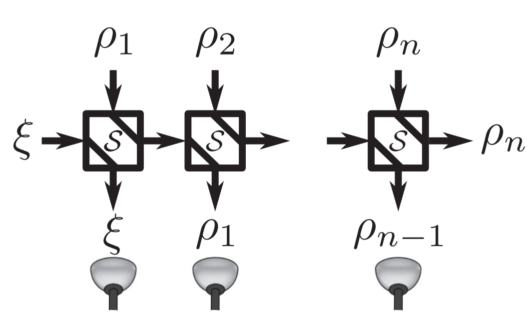

FIG. 4. The interaction $\mathcal{S}$ swaps the memory with the input, effectively making the ith output the $(i-1)$th input. Therefore, this memory channel is also called the delaying channel.

is the initial state of the memory. Let us first analyze the effect of ordering on the result of the estimation procedure.

1. Sequential ordering

Let $x_i = 1$ for $1 \geqslant i \geqslant N$, $x_i = 2$ for $N+1 \geqslant i \geqslant 2N$, and so on, for some fixed integer N. Thus, for the first N uses, we input $\rho(1)$ into the channel; then, for the next N uses, we input $\rho(2)$, and so on. Then the probability of observing outcome E_y when $\rho(x)$ is present at the input is

$$p[y|\rho(x)] = \frac{\mathrm{tr}\{[\rho(x-1) + (N-1)\rho(x)]E_y\}}{N} \approx \mathrm{tr}[\rho(x)E_y] \quad \text{for } N \gg 1. \tag{A1}$$

Since the output of the channel is identical to the input most of the time, for large N the experimenter has to conclude that he has an ideal channel $\mathcal{E}^1(\rho) = \rho$ for all $\rho \in \mathcal{S}(\mathcal{H})$. Similarly, if the experimenter looked on the concatenation of two uses of the channel, he would find that $\mathcal{E}^2(\rho_1 \otimes \rho_2) = \rho_1 \otimes \rho_2$.

2. Random ordering

Let x_i be randomly drawn according to discrete probability distribution q_i. Then the average output when $\rho(x)$ is at the input is the average input $\overline{\rho} = \sum_{x=1}^k q_x \rho(x)$. Hence,

$$p[y|\rho(x)] \approx \mathrm{tr}\left[\sum_{x=1}^{k} q_x \rho(x) E_y\right] \quad \text{for } N \gg 1, \\ = \mathrm{tr}[\overline{\rho} E_y]. \tag{A2}$$

Thus, the experimenter concludes that $\mathcal{E}^1(\rho) = \overline{\rho}$ for all $\rho \in \mathcal{S}(\mathcal{H})$. However, if the experimenter looks at the concatenation of two uses, he will find that $\mathcal{E}^2(\rho_1 \otimes \rho_2) = \overline{\rho} \otimes \rho_1$ for all $\rho_1,\rho_2 \in \mathcal{S}(\mathcal{H})$.

Note that as a result of this work, when using randomly ordered inputs, the estimated transformation is going to converge towards a channel. For sequential ordering, this is not true, even if in this particular example the result was indeed a channel. However, one can find examples already in the 2×2-qubit example, where the sequential ordering will converge towards a non-cp map.

3. Estimation

The unitary for swap $\mathcal{S}$ can be expressed using (7) as $\mathcal{S} = \mathcal{D}(\pi/4,\pi/4,\pi/4)$. Therefore, one can directly see that the singular value decomposition in (10) will be $A = 0$. However, from observing $A = 0$, we can only conclude that $\alpha_x = \alpha_y = \pi/4$ and we cannot determine α_z or the $\mathcal{V}_i$ or $\mathcal{W}$.

In order to see whether we deal with the swap interaction or its modified version, we will thus look on the vector representation of the delayed channel from the first input to the second input $\mathcal{E}^{1\to 2}$,

$$\mathcal{E}^{1\to 2}(\rho) = \mathrm{tr}_{\mathcal{H}_1}[\mathcal{E}^2(\rho \otimes 1/2\mathbb{I})], \tag{A3}$$

where we take the partial trace over the first input. Let us recall that for the swap interaction, this would lead to the identity channel, i.e., $\mathcal{E}^{1\to 2} = \mathcal{I}$.

In general, the vector representation can be written as

$$A^{1\to 2} = R_2 S R_{\mathcal{W}} S R_1, \tag{A4}$$

where R_i are rotations due to $\mathcal{V}_i$, $R_{\mathcal{W}}$ is the rotation corresponding to $\mathcal{W}$, and $S = \mathrm{diag}(s_{2z}, s_{2z}, 1)$. Further, we can expand $R_{\mathcal{W}}$ using the parametrization from (11),

$$R_{\mathcal{W}} = Z(\phi + \psi) Y(r) Z(\phi - \psi), \tag{A5}$$

where

$$Z(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{A6}$$

$$Y(r) = \begin{pmatrix} 2r^2 - 1 & 0 & -2r\sqrt{1-r^2} \\ 0 & 0 & 0 \\ 2r\sqrt{1-r^2} & 0 & 2r^2 - 1 \end{pmatrix}. \tag{A7}$$

Since $Z(\theta)$ commutes with S, the singular values of $A^{1\to 2}$ are the same as of $SY(r)S$ and take the following form:

$$\lambda_1 = \frac{1}{4}\sqrt{x^2 + 16y + |x|\sqrt{32y + x^2}}, \tag{A8}$$

$$\lambda_2 = \frac{1}{4}\sqrt{x^2 + 16y - |x|\sqrt{32y + x^2}}, \tag{A9}$$

$$\lambda_3 = s_{2z}^2 \tag{A10}$$

where $x^2 = 8(1 - 2r^2)^2 c_{2z}^4$, $y = s_{2z}^2$, and $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$. The value of α_z can be computed from the smallest singular value λ_3. If $\alpha_z = \pi/4$, then all singular values will be 1; thus, the delayed channel will be unitary, $A^{1\to 2} = R_2 R_{\mathcal{W}} R_1$, and, consequently, we reconstruct a unitary rotated (nonentangling) swap gate. If $\alpha_z \neq \pi/4$, then the reconstruction results in a nontrivial interaction correlating the system with the memory degrees of freedom.

[1] D. Kretschmann and R. F. Werner, Quantum channels with memory, Phys. Rev. A **72**, 062323 (2005).
[2] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory* (Cambridge University Press, Cambridge, 2013).
[3] T. Rybár, Quantum channels with memory, Acta Phys. Slov. **62**, 275 (2012).
[4] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).
[5] M. Paris and J. Řeháček, *Quantum State Estimation*, Lecture Notes in Physics (Springer, New York, 2004).
[6] C. T. Schmiegelow, A. Bendersky, M. A. Larotonda, and J. P. Paz, Selective and Efficient Quantum Process Tomography without Ancilla, Phys. Rev. Lett. **107**, 100502 (2011).
[7] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y. K. Liu, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).
[8] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices without Tomography, Phys. Rev. Lett. **107**, 210404 (2011).
[9] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert, Quantum State Tomography via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).
[10] D. H. Mahler, Lee A. Rozema, Ardavan Darabi, Christopher Ferrie, Robin Blume-Kohout, and A. M. Steinberg, Adaptive Quantum State Tomography Improves Accuracy Quadratically, Phys. Rev. Lett. **111**, 183601 (2013).
[11] R. Blume-Kohout, Hedged Maximum Likelihood Quantum State Estimation, Phys. Rev. Lett. **105**, 200504 (2010).
[12] Robin Blume-Kohout, Robust error bars for quantum tomography, arXiv:1202.5270.
[13] M. Christandl and R. Renner, Reliable Quantum State Tomography, Phys. Rev. Lett. **109**, 120403 (2012).
[14] Takanori Sugiyama, Peter S. Turner, and Mio Murao, Precision-Guaranteed Quantum Tomography, Phys. Rev. Lett. **111**, 160406 (2013).
[15] V. Scarani, S. Iblisdir, N. Gisin, and A. Acin, Quantum cloning, Rev. Mod. Phys. **77**, 1225 (2005).
[16] Notice that in this form, we omit the swapping of memory and system after the interaction for the sake of clarity. Hence, as written in the text, the interaction is $\mathcal{U}: \mathcal{M}\otimes\mathcal{H} \mapsto \mathcal{M}\otimes\mathcal{H}$ and not $\mathcal{M}\otimes\mathcal{H} \mapsto \mathcal{H}\otimes\mathcal{M}$, which is assumed when we need it for nice concatenation properties.
[17] B. Kraus and J. I. Cirac, Optimal creation of entanglement using a two-qubit gate, Phys. Rev. A **63**, 062309 (2001).
[18] I. L. Chuang and M. A. Nielsen, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2004).
[19] Mário Ziman and Vladimír Bužek, All (qubit) decoherences: Complete characterization and physical implementation, Phys. Rev. A **72**, 022110 (2005).
[20] Daniel Burgarth, Koji Maruyama, and Franco Nori, Indirect quantum tomography of quadratic Hamiltonians, New J. Phys. **13**, 013019 (2011).
[21] Daniel Burgarth and Kazuya Yuasa, Identifiability of open quantum systems, Phys. Rev. A **89**, 030302(R) (2014).
[22] Daniel Burgarth, Vittorio Giovannetti, Airi N. Kato, and Kazuya Yuasa, Quantum estimation via sequential measurements, arXiv:1507.07634.
[23] Merlijn van Horssen and Madalin Guta, Sanov and central limit theorems for output statistics of quantum Markov chains, J. Math. Phys. **56**, 022109 (2015).