Information-disturbance tradeoff in estimating a unitary transformation

Alessandro Bisio,[1] Giulio Chiribella,[2] Giacomo Mauro D'Ariano,[1] and Paolo Perinotti[1]
[1]*QUIT Group, Dipartimento di Fisica "A. Volta", INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy*
[2]*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, Ontario N2L 2Y5, Canada*
(Received 29 June 2010; published 7 December 2010)

We address the problem of the information-disturbance tradeoff associated to the estimation of a quantum transformation and show how the extraction of information about a black box causes a perturbation of the corresponding input-output evolution. In the case of a black box performing a unitary transformation, randomly distributed according to the invariant measure, we give a complete solution of the problem, deriving the optimal tradeoff curve and presenting an explicit construction of the optimal quantum network.

DOI: 10.1103/PhysRevA.82.062305 PACS number(s): 03.67.Lx, 03.65.−w

I. INTRODUCTION

One of the key features of quantum theory is the impossibility of extracting information from a system without producing a disturbance in its state; the only exception to this rule is the trivial case when the state belongs to a set of orthogonal states. A canonical illustration of the unavoidable disturbance caused by quantum measurements is Heisenberg's γ-ray microscope thought experiment [1]. The impossibility of a nondisturbing extraction of information is the working principle of quantum cryptography, whose security relies on the fact that any amount of information extracted by the eavesdropper causes a corresponding amount of disturbance that can be detected by the communicating parties. A quantitative expression of such an information-disturbance tradeoff is a nontrivial issue because there are many different ways to quantify "information" and "disturbance" which have been put forward in the literature [2–16].

All the scenarios analyzed in the past have one point in common: they concern the disturbance produced by measurements on *quantum states*. However, one can consider other scenarios where the measurements produce a disturbance on *quantum transformations*. For example, we may have a black box implementing an unknown transformation belonging to a set $\{\mathcal{E}_i\}$, with the restriction that the black box can be used only one time. On the one hand, we may try to identify the unknown transformation (that is, to find out the index i). On the other hand, we may want to use the black box on a variable input state. Clearly, in general the two tasks are incompatible: In this case there is a tradeoff between the amount of information that can be extracted about a black box and the disturbance caused on its action. In other words, we cannot estimate an unknown quantum dynamics without perturbing it. Therefore, it is important to find a quantitative formulation of the information-disturbance tradeoff and to find the optimal scheme that introduces the minimum amount of disturbance for any given amount of extracted information. Like the tradeoff for states, the tradeoff for transformations is relevant to the discussion of quantum cryptographic protocols where the secret key is encoded in a set of transformations, as it happens in the two-way protocols of Refs. [17–19] for finite-dimensional systems, and in the protocol of Ref. [20] for continuous variables. Here for simplicity we restrict our attention to the case of unitary transformations on finite-dimensional quantum systems. As in Refs. [7,11], we quantify the information gain and the disturbance with suitable fidelities, and we derive the minimum amount of disturbance associated to any possible value of the information gain.

The paper is structured as follows. In Sec. II we introduce the problem and the notation. Section III then provides a brief review of the formalism of quantum combs and generalized instruments [21–23], which is crucial in our paper. The complete analysis of the information-disturbance tradeoff for arbitrary unitary transformations is presented in Sec. IV; in particular, we first give the rigorous mathematical formulation of the problem (Sec. IV A), the analysis of its symmetries (Sec. IV B), the derivation of the optimal tradeoff curve (Sec. IV C), and, finally, the construction of the optimal network (Sec. IV D). We conclude the paper with a discussion of the results in Sec. V.

II. PRELIMINARIES AND NOTATION

In the case of states, the mathematical tool to analyze the information-disturbance tradeoff is the *quantum instrument*. In the discrete-outcome case, a quantum instrument is a set of *quantum operations* (trace-decreasing completely positive maps) $\{\mathcal{T}_i\}$ transforming operators on the input system Hilbert space $\mathcal{H}_0$ to operators on the output system Hilbert space $\mathcal{H}_1$, with the normalization condition that $\mathcal{T} := \sum_i \mathcal{T}_i$ is trace preserving (that is, it is a *quantum channel*). A quantum instrument describes a measurement process that outputs the classical outcome i and the quantum state $\mathcal{T}_i(\rho)/\mathrm{Tr}[\mathcal{T}_i(\rho)]$ with probability $p_i = \mathrm{Tr}[\mathcal{T}_i(\rho)]$. To derive our results, we use the generalization of the notion of instrument to measurement processes on quantum transformations, rather than on quantum states [21–23]. This extension is presented in Sec. III.

In the following we denote the linear operators on a Hilbert space $\mathcal{H}$ by $L(\mathcal{H})$. We make extensive use of the isomorphism between linear operators in $L(\mathcal{H})$ and vectors in $\mathcal{H}\otimes\mathcal{H}$ given by

$$A = \sum_{nm} \langle n|A|m\rangle |n\rangle\langle m| \to |A\rangle\rangle = \sum_{nm} \langle n|A|m\rangle |n\rangle|m\rangle, \quad (1)$$

where $\{|n\rangle\}$ is a fixed orthonormal basis for $\mathcal{H}$. The isomorphism satisfies the property

$$|A\rangle\rangle = (A\otimes I)|I\rangle\rangle = (I\otimes A^T)|I\rangle\rangle, \quad (2)$$

where T denotes transposition with respect to the fixed basis. For the sake of clarity, we often use the notation $\mathcal{H}_{a,b}$ to denote the tensor product $\mathcal{H}_a \otimes \mathcal{H}_b$, $A_{a,b}$ to stress that A belongs to $(\mathcal{H}_{a,b})$, and, similarly, $|\psi\rangle_a$ and $|A\rangle\rangle_{a,b}$ to stress that $|\psi\rangle$ belongs to $\mathcal{H}_a$ and $|A\rangle\rangle$ belongs to $\mathcal{H}_{a,b}$.

Using the Choi isomorphism [24], we can associate each completely positive map $\mathcal{T}_i : L(\mathcal{H}_0) \to L(\mathcal{H}_1)$ with a positive operator $T_i \in L(\mathcal{H}_1 \otimes \mathcal{H}_0)$ given by

$$T_i = (\mathcal{T}_i \otimes \mathcal{I}_0)(|I\rangle\rangle\langle\langle I|_{0,0}), \tag{3}$$

where $\mathcal{I}_0$ is the identity map on $\mathcal{H}_0$.

In terms of the Choi operator, the condition that $\mathcal{T}$ is trace preserving (resp. trace decreasing) becomes

$$\mathrm{Tr}_1[T] = I_0 \quad (\mathrm{Tr}_1[T] \leqslant I_0), \tag{4}$$

where Tr_1 denotes a partial trace over $\mathcal{H}_1$.

We now introduce the tradeoff problem for quantum transformations. Consider a quantum network $\mathcal{R}$ with an empty slot that can be linked with a variable quantum device, the input-output action of the latter being described by a channel in the set $\{\mathcal{E}_i\}$. Ideally, we would like the network $\mathcal{R}$ to give us some information about the channel $\mathcal{E}_i$ without affecting the output state $\mathcal{E}_i(\rho)$ that the channel should produce when an input state ρ is fed in the corresponding device (see Fig. 1). However, as already mentioned, this is not possible in general.

As we already mentioned, there are two extreme situations. At one extreme, if we are only interested in extracting information, the best strategy is to apply the channel on one side of a suitable bipartite state $\sigma \in L(\mathcal{H}_o \otimes \mathcal{H}_0)$, thus getting the output state $(\mathcal{E}_i \otimes \mathcal{I}_0)(\sigma)$, and then to perform a suitable measurement $\{P_j\}$. In this case the available use of the channel is consumed for estimation; after this step, the best we can do to produce an output state close to $\mathcal{E}_i(\rho)$ is to apply to the input state ρ some channel $\tilde{\mathcal{E}}_j$ that depends on the outcome j of our measurement. At the opposite extreme, if we do not tolerate any disturbance, the only possibility is to apply the black box to the input state ρ. In this case we correctly obtain the output state $\mathcal{E}_i(\rho)$, but we have no measurement data to infer the identity of the unknown device. In the intermediate cases, it is important to assess the maximum amount of information that can be gathered without trespassing a given disturbance threshold.

Since the tradeoff problem involves optimization of quantum networks, we use the approach of *quantum combs* developed in Refs. [21–23]. This approach is based on the characterization of the most general transformations that quantum

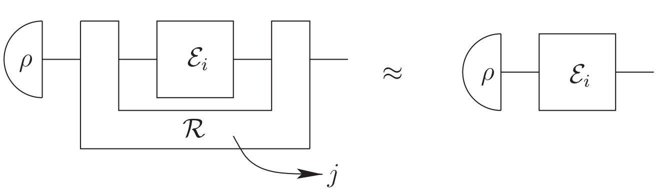


FIG. 1. Given a black box implementing an unknown channel $\mathcal{E}_i$ drawn from a set $\{\mathcal{E}_i\}$, we want to link it with a quantum network $\mathcal{R}$ that both gives an estimate j of the parameter i and affects the output state $\mathcal{E}_i(\rho)$ as little as possible. (The symbol ≈ means that the network $\mathcal{R}$ is optimized in such a way that the output of the two circuits on the left and right is as close as possible.)

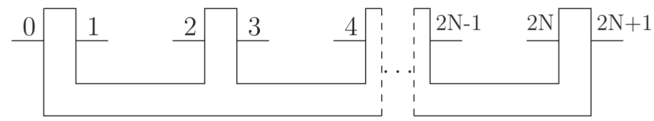


FIG. 2. A quantum comb with $N+1$ teeth. The flow of quantum information is from left to right. The input wires of the network are labeled with even numbers from 0 to $2N$, the output wires with odd numbers from 1 to $2N+1$.

channels can undergo and on realization theorems proving that all these abstract transformations can be implemented by quantum networks. Since our theorems are constructive, this approach also provides the explicit form of the optimal quantum network.

III. QUANTUM COMBS AND GENERALIZED INSTRUMENTS

By stretching and rearranging the internal wires, we can give to every quantum network the shape of a comb. The empty slots of the network become the empty spaces between the teeth of the comb.

Referring to Fig. 2, each wire is labeled with a natural number, which is even for the input wires and odd for the output ones; the corresponding Hilbert spaces are labeled accordingly.

If our network consists of a sequence of N quantum channels (trace-preserving maps), then we call it *deterministic*. To every deterministic network, we can associate a positive operator $R^{(N)} \geqslant 0$, called a *quantum comb*, satisfying the normalization condition [21,23,25]

$$\mathrm{Tr}_{2k-1}[R^{(k)}] = I_{2k-2} \otimes R^{(k-1)}, \quad k = 1, \ldots, N, \tag{5}$$

where $R^{(0)} = 1$, $R^{(k)} \in L(\mathcal{H}_{\mathrm{out_k}} \otimes \mathcal{H}_{\mathrm{in_k}})$ with $\mathcal{H}_{\mathrm{in_k}} = \bigotimes_{n=0}^{k-1} \mathcal{H}_{2n}$ and $\mathcal{H}_{\mathrm{out_k}} = \bigotimes_{n=0}^{k-1} \mathcal{H}_{2n+1}$, is the comb of the reduced circuit obtained by discarding the last $N-k$ teeth.

The normalization condition of Eq. (5) reflects the causal ordering in the deterministic network. We call a comb satisfying Eq. (5) *deterministic*, and we denote by $\mathsf{DetComb}(\bigotimes_{i=0}^{2N-1} \mathcal{H}_i)$ the set of all deterministic combs with the given ordering of the input and output spaces. A deterministic quantum comb with $N = 1$ is simply the Choi operator of a quantum channel: in this case the condition of Eq. (5) is equivalent to the normalization of the channel given in Eq. (4). Accordingly, $\mathsf{DetComb}(\mathcal{H}_b \otimes \mathcal{H}_a)$ is the set of (Choi operators of) quantum channels from $L(\mathcal{H}_a)$ to $L(\mathcal{H}_b)$.

This framework of quantum combs can be easily extended to the case of networks consisting of quantum operations (trace-decreasing maps). We call a *probabilistic comb* a positive operator $S^{(N)} \geqslant 0$ that is bounded by some deterministic comb, that is, an operator $S^{(N)}$ with the property

$$\exists R^{(N)} \in \mathsf{DetComb}\left(\bigotimes_{k=0}^{2N-1} \mathcal{H}_k\right) \quad \text{such that } S^{(N)} \leqslant R^{(N)}. \tag{6}$$

We denote by $\mathsf{ProbComb}(\bigotimes_{k=0}^{2N-1} \mathcal{H}_k)$ the set of all probabilistic combs with given ordering of the input and output spaces. A probabilistic comb with $N = 1$ is simply the Choi operator of a quantum operation: in this case the condition of Eq. (6) is equivalent to the bound in Eq. (4). Accordingly, $\mathsf{ProbComb}(\mathcal{H}_b \otimes \mathcal{H}_a)$ is the set of (Choi operators of) quantum operations from $L(\mathcal{H}_a)$ to $L(\mathcal{H}_b)$.

Two quantum networks $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(M)}$ can be linked together by connecting some wires of $\mathcal{R}^{(N)}$ with some wires of $\mathcal{S}^{(M)}$. Let us denote by J the set of wires that are connected and by K (L) the set of wires of $\mathcal{R}^{(N)}$ ($\mathcal{S}^{(M)}$) that are not. The circuit resulting from the connection, denoted by $\mathcal{R}^{(N)} * \mathcal{S}^{(M)}$, has a Choi operator given by the *link product*

$$R^{(N)} * S^{(M)} = \mathrm{Tr}_{\mathsf{J}}\big[\big(R^{(N)}_{\mathsf{K},\mathsf{J}} \otimes I_{\mathsf{L}}\big)\big(I_{\mathsf{K}} \otimes S^{(M)T_{\mathsf{J}}}_{\mathsf{J},\mathsf{L}}\big)\big], \qquad (7)$$

where $R^{(N)}, S^{(M)}$ are the Choi operators of $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(M)}$, respectively, and T_{J} denotes the partial transposition with respect to the fixed orthonormal basis of Eq. (1). For example, let $\mathcal{E}$ be a channel from $\mathcal{H}_0$ to $\mathcal{H}_1$ and $\mathcal{F}$ be a channel from $\mathcal{H}_1$ to $\mathcal{H}_2$; then the Choi operator of the composition $\mathcal{F}\mathcal{E}$ is

$$F * E = \mathrm{Tr}_1\big[(F_{2,1} \otimes I_0)\big(I_2 \otimes E^{T_1}_{1,0}\big)\big]. \qquad (8)$$

As a particular case for $\mathcal{H}_0 \simeq \mathbb{C}$, if ρ is a state on $\mathcal{H}_1$ and $\mathcal{E}$ is a channel from $\mathcal{H}_1$ to $\mathcal{H}_2$, one has

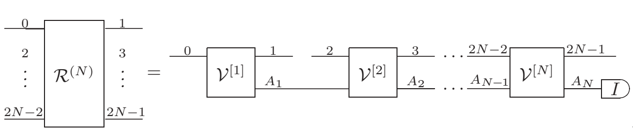
$$\mathcal{E}(\rho) = E_{21} * \rho_1 = \mathrm{Tr}_1\big[E_{21}\big(I_2 \otimes \rho_1^T\big)\big]. \qquad (9)$$

A deterministic (probabilistic) quantum comb, besides representing a quantum network with some empty slots, can also represent a quantum channel (operation) $\mathcal{R}^{(N)}$ from $\mathcal{H}_{\mathrm{even}} := \bigotimes_{k=0}^{N-1} \mathcal{H}_{2k}$ to $\mathcal{H}_{\mathrm{odd}} := \bigotimes_{k=0}^{N-1} \mathcal{H}_{2k+1}$. Due to the Choi isomorphism, the channel $\mathcal{R}^{(N)}$ is in one-to-one correspondence with the comb $R^{(N)}$. In the following two subsections, we exploit this correspondence to discuss the physical realization of quantum combs, both in the deterministic case (Sec. III A) and in the probabilistic case (Sec. III B).

A. Realization of deterministic combs

The following theorem, proved in Ref. [26], gives an explicit construction for the realization of every deterministic quantum comb as a sequence of isometric channels.

Theorem 1. Realization of deterministic combs. Every deterministic comb can be realized as a concatenation of isometric channels in the following way:



, (10)

where A_k is an ancilla with Hilbert space $\mathcal{H}_{A_k}$, $\mathcal{V}^{[k]} : L(\mathcal{H}_{2k-2} \otimes \mathcal{H}_{A_{k-1}})$ is the channel defined by $\mathcal{V}^{[k]}(\rho) = V^{[k]}\rho V^{[k]\dagger}, \forall \rho \in L(\mathcal{H}_{2k-2} \otimes \mathcal{H}_{A_{k-1}})$ for a suitable isometry $V^{[k]} : \mathcal{H}_{2k-2} \otimes \mathcal{H}_{A_{k-1}} \to \mathcal{H}_{2k-1} \otimes \mathcal{H}_{A_k}$, and $\overset{A_N}{—}\!\boxed{I}$ represents the partial trace on $\mathcal{H}_{A_N}$. Precisely, the ancillary Hilbert spaces $\mathcal{H}_{A_{k-1}}, \mathcal{H}_{A_k}$ are defined by $\mathcal{H}_{A_0} := \mathbb{C}$ and $\mathcal{H}_{A_k} := \mathsf{Supp}(R^{(k)*}) \subseteq \bigotimes_{n=0}^{2k-1} \mathcal{H}_n$ for $k \geqslant 1$, where $\mathsf{Supp}(R^{(k)*})$ denotes the support of the complex conjugate of $R^{(k)}$ in the fixed basis. The isometry $V^{[k]}$ is given by the expression

$$\begin{aligned} V^{[k]} := \big\{ I_{2k-1} \otimes \big[\big(R^{(k)*}_{(2k-1)',\dots,0'}\big)^{\frac{1}{2}}\big(I_{(2k-1)',(2k-2)'} \\ \otimes R^{(k-1)*}_{(2k-3)',\dots,0'}\big)^{-\frac{1}{2}}\big]\big\} (|I\rangle\rangle_{(2k-1),(2k-1)'} \\ \otimes T_{(2k-2)' \leftarrow (2k-2)} \otimes I_{(2k-3)'} \otimes \cdots \otimes I_{0'}\big), \end{aligned} \qquad (11)$$

where $\mathcal{H}_{k'} \simeq \mathcal{H}_k$ and $T_{m \leftarrow n}$ is the teleportation operator from $\mathcal{H}_n$ to $\mathcal{H}_m$, given by $T_{m\leftarrow n} = \sum_k |k\rangle_m \langle k|_n$.

Note that the isometry $V^{[k]}$ defined in Eq. (11) has the correct input and output Hilbert spaces. Indeed, for $k = 1$, one has $R^{(0)} = 1$ and the isometry $V^{[1]}$, given by $V^{[1]} = [I_1 \otimes (R^{(1)*}_{1',0'})^{\frac{1}{2}}](|I\rangle\rangle_{1,1'} \otimes T_{0\to 0'})$, sends vectors in $\mathcal{H}_0$ to vectors in $\mathcal{H}_1 \otimes \mathcal{H}_{A_1}$, where $\mathcal{H}_{A_1} = \mathsf{Supp}(R^{(1)*}_{1',0'})$ is a subspace of $\mathcal{H}_{1'} \otimes \mathcal{H}_{0'}$. For $k > 1$, since $\mathcal{H}_{A_{k-1}} = \mathsf{Supp}(R^{(k-1)*}_{(2i-3)',\dots,0'})$ is a subspace of $\mathcal{H}_{(2k-3)'} \otimes \cdots \otimes \mathcal{H}_{0'}$, the isometry $V^{[k]}$ sends vectors in $\mathcal{H}_{2k-2} \otimes \mathcal{H}_{A_{k-1}}$ to vectors in $\mathcal{H}_{2k-1} \otimes \mathcal{H}_{A_k}$, as stated by the thesis.

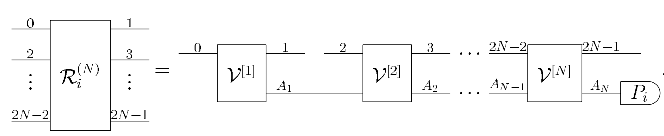
B. Generalized N-instruments

In the discrete-outcome case, a *generalized N-instrument* is a set of probabilistic combs $\{R_i\} \subset \mathsf{ProbComb}(\bigotimes_{k=0}^{2N-1} \mathcal{H}_k)$ satisfying the normalization condition

$$R^{(N)} := \sum_i R_i \in \mathsf{DetComb}\left(\bigotimes_{k=0}^{2N-1} \mathcal{H}_k\right). \qquad (12)$$

When $N = 1$ the notion of a generalized instrument coincides with the usual notion of a quantum instrument. Every N-instrument can be realized as a quantum network, due to an analog of Ozawa's dilation theorem [27]. The proof of the dilation theorem for generalized N-instruments was originally presented in Ref. [23] and is combined here with Theorem 1.

Theorem 2. Realization of N-instruments. Every generalized N-instrument $\{R_i^{(N)}\} \subset \mathsf{ProbComb}(\bigotimes_{k=0}^{2N-1} \mathcal{H}_k)$ can be realized as a quantum network of isometric channels followed by a measurement of the last ancilla, as follows:



,

Here $A_k, \mathcal{H}_{A_k}$, and $V^{[k]}$ are the ancillas, the Hilbert spaces, and the isometries providing the realization of the deterministic comb $R^{(N)} = \sum_i R_i^{(N)}$, as given by Theorem 1, $\overset{A_N}{—}\!\boxed{P_i}$ represents the partial trace on $\mathcal{H}_{A_N}$ with the operator P_i. $\{P_i\}$ is a quantum measurement on the ancilla A_N, described by the positive operator-valued measure (POVM)

$$P_i = (R^{(N)*})^{-\frac{1}{2}} R_i^{(N)*} (R^{(N)*})^{-\frac{1}{2}}. \qquad (13)$$

In our study of the information-disturbance tradeoff, we use generalized 2-instruments, which can be graphically represented by combs with two teeth and one empty slot where the unknown black box can be inserted, as in Fig. 1. Since the value of N is fixed to $N = 2$, in the following we drop the index (N) in $R_i^{(N)}$ and $R^{(N)}$ and simply write R_i and R.

IV. INFORMATION-DISTURBANCE TRADEOFF FOR UNITARY CHANNELS

A. Formulation of the problem

Suppose that a black box performs an unknown unitary channel $\mathcal{U}(\rho) = U\rho U^\dagger$, where the unitary $U \in \mathrm{SU}(d)$ is randomly drawn according to the normalized Haar measure dU. Let $\mathcal{H}_1 \simeq \mathcal{H}_2 \simeq \mathbb{C}^d$ be the input and output Hilbert spaces for the unknown channel, respectively. In order to extract information, we then use a quantum network like that of Eq. (12) with $N = 2$. The network is then described by a generalized 2-instrument $\{R_{\hat{U}}\}$, where the outcome $\hat{U} \in \mathrm{SU}(d)$ is the estimate of the unknown parameter U. For every possible outcome $\hat{U}$, $R_{\hat{U}}$ is an element of $\mathsf{ProbComb}(\bigotimes_{k=0}^3 \mathcal{H}_k)\}$, with $\mathcal{H}_0 \simeq \mathcal{H}_1 \simeq \mathcal{H}_2 \simeq \mathcal{H}_3$, and one has the normalization condition

$$R := \int_{\mathrm{SU}(d)} d\hat{U} R_{\hat{U}} \in \mathsf{DetComb}\left(\bigotimes_{k=0}^{3} \mathcal{H}_k\right), \tag{14}$$

which is the continuous version of Eq. (12). Since there is no ambiguity, from now on we omit the domain of integration SU(d).

When the unknown black box with channel $\mathcal{U}$ is connected to the quantum network with generalized instrument $\{\mathcal{R}_{\hat{U}}\}$, one obtains a set of quantum operations $\mathcal{R}_{\hat{U}} * \mathcal{U}$, each corresponding to a possible result of the measurement. However, to speak about the "probability of the outcome $\hat{U}$," we need to know what input state ρ is fed in the circuit: we cannot speak of the "probability of a quantum operation" without specifying its input state. If the input state is $\rho \in L(\mathcal{H}_0)$, then the probability is given by the trace of the output state $(\mathcal{R}_{\hat{U}} * \mathcal{U})(\rho)$:

$$\begin{aligned} p(\hat{U}|U,\rho) &= \mathrm{Tr}[(\mathcal{R}_{\hat{U}} * \mathcal{U})(\rho)] \\ &= \mathrm{Tr}[R_{\hat{U}}(I_3 \otimes |U^*\rangle\!\rangle\langle\!\langle U^*|_{2,1} \otimes \rho_0^*)], \end{aligned} \tag{15}$$

where we used the link product of Eqs. (7) and (9) to compute the Choi operator of $\mathcal{R}_{\hat{U}} * \mathcal{U}$ and the action of the channel $\mathcal{R}_{\hat{U}} * \mathcal{U}$ on the state ρ, respectively. We also used the fact that $A^T = A^*$ for every self-adjoint operator.

To quantify the information gain and the disturbance, we now introduce two suitable fidelities. Suppose that the black box performs the unitary channel $\mathcal{U}$ and that the measurement outcome is $\hat{U}$. In this case we quantify information gain with the fidelity

$$g(\hat{U},U) = \frac{1}{d^2}|\,\mathrm{Tr}[\hat{U}U^\dagger]|^2. \tag{16}$$

Note that the maximum value of the fidelity is 1, and it is achieved if and only if $\hat{U} = \omega U$ for some phase $|\omega| = 1$, that is, if and only if the two unitary channels $\hat{\mathcal{U}}$ and $\mathcal{U}$ coincide. The fidelity $g(\hat{U},U)$ enjoys the invariance property

$$g(\hat{U},U) = g(V\hat{U}W, VUW) \quad \forall V,W \in \mathrm{SU}(d). \tag{17}$$

Averaging the fidelity with the probability of the estimate $\hat{U}$ given the true value U and the input state ρ, we then obtain the *average information gain*

$$G_\rho := \int dU \int d\hat{U}\, p(\hat{U}|U,\rho) g(\hat{U},U). \tag{18}$$

In our analysis we always assume that the input state is given by $\rho = I/d$. The reason for this choice is that the condition $\rho = I/d$ arises in two relevant scenarios:

(1) when the input system (wire 0) of the circuit is prepared in a maximally entangled state with some reference system $0'$. This is the case in the cryptographic protocols of Refs. [17,19] (and, in the infinite energy limit, also in the continuous-variable scenario of Ref. [20]).

(2) when the input system is prepared at random in one of the states of an ensemble $\{\rho_i, p_i\}$, with the property that $\sum_i p_i \rho_i = I/d$. This is the case of the cryptographic protocol of Ref. [18].

Since we are setting $\rho = I/d$, we drop the subscript ρ from G_ρ. Using Eqs. (15) and (16), the expression for the information gain G is

$$G = \frac{1}{d^3}\int dU \int d\hat{U}\, \mathrm{Tr}_{3,0}[\langle\!\langle U^*|_{2,1} R_{\hat{U}} |U^*\rangle\!\rangle_{2,1}] |\langle\!\langle \hat{U}|U\rangle\!\rangle|^2. \tag{19}$$

We now introduce our figure of merit for the minimization of the disturbance. To this purpose, we consider the *channel fidelity* [28] between the overall quantum operation $\mathcal{R}_{\hat{U}} * \mathcal{U}$ performed by the network and the input channel $\mathcal{U}$. This is the fidelity between the two output states produced by the two operations $\mathcal{R}_{\hat{U}} * \mathcal{U}$ and $\mathcal{U}$ when applied on one side of the maximally entangled state $|\Phi\rangle\!\rangle = \frac{1}{\sqrt{d}}|I\rangle\!\rangle_{0,0'}$. In terms of Choi operators, the channel fidelity is given by

$$\begin{aligned} F(\mathcal{R}_{\hat{U}} * \mathcal{U},\mathcal{U}) &= \frac{1}{d^2}\langle\!\langle U|_{3,0}(R_{\hat{U}} * |U\rangle\!\rangle\langle\!\langle U|_{2,1})|U\rangle\!\rangle_{3,0} \\ &= \frac{1}{d^2}\langle\!\langle U|_{3,0}\langle\!\langle U^*|_{2,1} R_{\hat{U}}|U\rangle\!\rangle_{3,0}|U^*\rangle\!\rangle_{2,1}, \end{aligned}$$

where we used the fact that, by definition of the Choi operator [Eq. (3)], one has $(\mathcal{E} \otimes \mathcal{I})(|\Phi\rangle\!\rangle\langle\!\langle\Phi|) = E/d$ for every quantum operation $\mathcal{E}$.

Averaging over the outcomes and the true values, we then obtain the *average fidelity*

$$\begin{aligned} F &:= \frac{1}{d^2}\int dU \int d\hat{U} \langle\!\langle U|_{3,0}\langle\!\langle U^*|_{2,1} R_{\hat{U}}|U\rangle\!\rangle_{3,0}|U^*\rangle\!\rangle_{2,1} \\ &= \frac{1}{d^2}\int dU \langle\!\langle U|_{3,0}\langle\!\langle U^*|_{2,1} R|U\rangle\!\rangle_{3,0}|U^*\rangle\!\rangle_{2,1}. \end{aligned} \tag{20}$$

Note that the fidelity F naturally arises also in the case where the input state at the wire 0 is a pure state $\varphi = |\varphi\rangle\langle\varphi|$ chosen at random according to the uniform measure on pure states: in this case the fidelity between $\mathcal{R}_{\hat{U}} * \mathcal{U}(\varphi)$ and $\mathcal{U}(\varphi)$, averaged over φ, U, and $\hat{U}$, is given by

$$\begin{aligned} F' &= \int d\varphi \int dU \int d\hat{U}\, \mathrm{Tr}[(U\varphi U^\dagger)(\mathcal{R}_{\hat{U}} * \mathcal{U})(\varphi)] \\ &= \int d\varphi \int dU \int d\hat{U}\, \mathrm{Tr}[(R_{\hat{U}} * |U\rangle\!\rangle\langle\!\langle U|_{2,1})(U\varphi U^\dagger \otimes \varphi^*)] \\ &= \int d\varphi \int dU \langle U\varphi|_3 \langle\!\langle U^*|_{2,1}\langle\varphi^*|_0 R|U\varphi\rangle_3|U^*\rangle\!\rangle_{2,1}|\varphi^*\rangle_0 \\ &= \int dU \bigg(\frac{d}{d+1}F(\mathcal{R} * \mathcal{U},\mathcal{U}) \\ &\quad + \frac{1}{d(d+1)}\mathrm{Tr}[\langle\!\langle U^*|_{2,1} R|U^*\rangle\!\rangle_{2,1}]\bigg) \\ &= \frac{d}{d+1}F + \frac{1}{d+1}, \end{aligned}$$

having used the relation

$$\int d\varphi\varphi \otimes \varphi^* = \frac{|I\rangle\rangle\langle\langle I| + I \otimes I}{d(d+1)} \tag{21}$$

and the normalization $\mathrm{Tr}[R] = d^2$, which follows directly from Eq. (5) in the $N = 2$ case with $\mathcal{H}_3 \simeq \mathcal{H}_2 \simeq \mathcal{H}_2 \simeq \mathcal{H}_0 \simeq \mathbb{C}^d$. Since there is a tradeoff, the information gain G and the fidelity F cannot achieve their maximum values at the same time. Therefore, we introduce a weight $0 \leqslant p \leqslant 1$ that quantifies how much we care about information extraction versus disturbance minimization, and our figure of merit is the convex combination $pG + (1-p)F$. The extreme case $p = 0$ (resp. $p = 1$) corresponds to the situation where we do not tolerate any disturbance (resp. where we are only interested in extracting maximum information). The tradeoff curve obtained by the maximization of $pG + (1-p)F$ for all possible values of $p \in [0,1]$ is the same curve that would be obtained by maximizing F for given G (i.e., by finding the minimum disturbance for a given amount of extracted information) or by maximizing G for given F (i.e., by finding the maximum amount of extractable information for a given disturbance threshold).

B. Symmetry of the estimating network

Here we exploit the symmetries of the figure of merit $pG + (1-p)F$ to simplify the optimization problem. The crucial simplification comes from the following theorem, which states the symmetry properties of the optimal generalized instrument.

Theorem 3. Symmetries of the optimal instrument. Let G and F be the information gain and the fidelity defined in Eqs. (19) and (20). For every $p \in [0,1]$, the generalized instrument that maximizes $pG + (1-p)F$ can be chosen to be *covariant*, that is, of the form

$$R_{\hat{U}} = (\hat{\mathcal{U}}_3 \otimes \hat{\mathcal{U}}_2^* \otimes \mathcal{I}_{1,0})(\Xi), \tag{22}$$

for some positive operator $\Xi \in L(\mathcal{H}_3 \otimes \mathcal{H}_2 \otimes \mathcal{H}_1 \otimes \mathcal{H}_0)$.

Moreover, the operator Ξ satisfies the commutation relation

$$[\Xi, V_3 \otimes V_2^* \otimes V_1 \otimes V_0^*] = 0, \quad \forall V \in \mathrm{SU}(d). \tag{23}$$

Proof. The proof is based on the same argument used for the proof of Lemma 2 in Ref. [29]. Consider an arbitrary generalized instrument $\{R_{\hat{U}}\}$. Using the invariance of the Haar measure and of the fidelity $g(\hat{U},U)$ [Eq. (17)], it is easy to check that the values of F and G in Eqs. (19) and (20) do not change if each $R_{\hat{U}}$ is replaced by the group average

$$R'_{\hat{U}} := \int dV dW (\mathcal{V}_3 \otimes \mathcal{V}_2^* \otimes \mathcal{W}_1 \otimes \mathcal{W}_0^*)(R_{V^\dagger \hat{U} W}), \tag{24}$$

where $\mathcal{V},\mathcal{V}^*,\mathcal{W},\mathcal{W}^*$ are the unitary channels corresponding to the unitaries V,V^*,W,W^*, respectively. Note that $\{R'_{\hat{U}}\}$ is still a generalized instrument, because it satisfies the normalization condition of Eq. (14). Moreover, from Eq. (24) it is clear that $\Xi := R'_I$ satisfies the commutation relation of Eq. (23). Finally, from Eq. (24) it is also clear that for every $\hat{U},V,W \in \mathrm{SU}(d)$ one has

$$R'_{V\hat{U}W^\dagger} = (\mathcal{V}_3 \otimes \mathcal{V}_2^* \otimes \mathcal{W}_1 \otimes \mathcal{W}_0^*)(R'_{\hat{U}}).$$

Taking $\hat{U} = W = I$, one then obtains $R'_V = (\hat{\mathcal{V}}_3 \otimes \hat{\mathcal{V}}_2^* \otimes \mathcal{I}_{1,0})(\Xi)$; namely, $R'_{\hat{U}}$ is of the form of Eq. (22). Since the substitution $\{R_{\hat{U}}\} \to \{R'_{\hat{U}}\}$ can be done for every generalized instrument, in particular it can be done for the optimal one. ■

Using Theorem 3, we can now express the normalization condition of Eq. (14) in a particularly simple way. Indeed, Eqs. (22) and (23) imply that the normalization operator $R = \int \mathrm{d}\hat{U} R_{\hat{U}}$ satisfies the commutation relation

$$[R, V_3 \otimes V_2^* \otimes W_1 \otimes W_0^*] = 0, \quad \forall V,W \in \mathrm{SU}(d).$$

The Schur lemma then implies $\mathrm{Tr}_3[R] = I_2 \otimes R^{(1)}$ for some positive operator $R^{(1)} \in L(\mathcal{H}_1 \otimes \mathcal{H}_0)$, and $\mathrm{Tr}_1[R^{(1)}] = \alpha I_0$ for some positive number $\alpha \in \mathbb{R}$. Therefore, the normalization condition for R to be a deterministic comb [Eq. (5) for $N = 2$] becomes trivially $\mathrm{Tr}[R] = d^2$, or, equivalently,

$$\mathrm{Tr}[\Xi] = d^2. \tag{25}$$

C. Optimal tradeoff curve

Exploiting Theorem 3 and the Schur lemmas, we can now rewrite the figure of merit as

$$pG + (1-p)F = \mathrm{Tr}\{[p\Lambda_G + (1-p)\Lambda_F]\Xi\}, \tag{26}$$

where Λ_G and Λ_F are the positive operators given by

$$\begin{aligned}\Lambda_F :=& \frac{1}{d^2(d^2-1)}[I_{3,2,1,0} + d^2 P_{3,2} \otimes P_{1,0} \\ &- P_{3,2} \otimes I_{1,0} - I_{3,2} \otimes P_{1,0}],\end{aligned} \tag{27}$$

$$\begin{aligned}\Lambda_G :=& \frac{1}{d}(I_3 \otimes \mathrm{Tr}_{3,0}[\Lambda_F] \otimes I_0) \\ =& \frac{1}{d^2(d^2-1)}\left[\left(1 - \frac{2}{d^2}\right) I_{3,2,1,0} + I_3 \otimes P_{2,1} \otimes I_0\right],\end{aligned} \tag{28}$$

where $P = d^{-1}|I\rangle\rangle\langle\langle I|$ is the projector on the one-dimensional invariant subspace of $V \otimes V^*$.

Since the only restrictions on Ξ are positivity and the normalization given by Eq. (25), the optimal choice is to take Ξ proportional to the projector on the eigenvector of $p\Lambda_G + (1-p)\Lambda_F$ corresponding to the maximum eigenvalue; up to normalization, this eigenvector the can be shown to be of the form [11]

$$|\chi\rangle = x|I\rangle\rangle_{3,0}|I\rangle\rangle_{2,1} + y|I\rangle\rangle_{3,2}|I\rangle\rangle_{1,0}, \quad x,y \in \mathbb{R}^+. \tag{29}$$

In order to satisfy Eq. (25), we then choose $\Xi = |\chi\rangle\langle\chi|$ with the normalization $\langle\chi|\chi\rangle = d^2$. Recalling Eq. (22), we get

$$\begin{aligned}R_{\hat{U}} &= |\chi_{\hat{U}}\rangle\langle\chi_{\hat{U}}|, \\ |\chi_{\hat{U}}\rangle &:= x|\hat{U}\rangle\rangle_{3,0}|\hat{U}^*\rangle\rangle_{2,1} + y|I\rangle\rangle_{3,2}|I\rangle\rangle_{1,0}.\end{aligned} \tag{30}$$

The normalization of χ implies that x and y obey the quadratic equation

$$x^2 + y^2 + \frac{2xy}{d} = 1. \tag{31}$$

Note that in the above equation there is just one free parameter (either x or y), which can be expressed, for example, as a function on the tradeoff ratio p. Fidelity and gain can be calculated in terms of the parameters x and y, thus getting the following expressions:

$$F = 1 - \frac{d^2-2}{d^2}x^2, \quad G = \frac{2-y^2}{d^2}. \qquad (32)$$

The extreme situation of minimum disturbance (resp. maximum extraction of information) can be retrieved in the extreme case $x=0$, $y=1$ (resp. $y=1$, $x=0$). Indeed, when $x=0$ and $y=1$, one has $R_{\hat{U}} = |I\rangle\rangle\langle\langle I|_{3,2} \otimes |I\rangle\rangle\langle\langle I|_{1,0}$ for all $\hat{U}$. In this case, there is no information extracted and one has $\mathcal{R}_{\hat{U}} * \mathcal{U} = \mathcal{U}$; that is, the instrument realizes the identity map. Accordingly, the fidelity F reaches its maximum $F=1$, while the information gain takes its minimum $G = \frac{1}{d^2}$, the value achieved by a random guess according to the Haar measure. In the opposite case ($x=1$, $y=0$), one has instead $R_{\hat{U}} = |\hat{U}\rangle\rangle\langle\langle \hat{U}|_{3,0} \otimes |\hat{U}^*\rangle\rangle\langle\langle \hat{U}^*|_{2,1}$, which implies $\mathcal{R}_{\hat{U}} * \mathcal{U} = |\mathrm{Tr}[\hat{U}U^\dagger]|^2 \hat{\mathcal{U}}$. This means that in this case our circuit performs the optimal estimation of U [29] and then executes the transformation $\hat{\mathcal{U}}$ on the input state. Accordingly, the fidelity drops to its minimum value, $F = \frac{2}{d^2}$, and the information gain reaches its maximum, $G = \frac{2}{d^2}$.

Following Ref. [11], we now introduce the information variable $0 \leqslant I \leqslant 1$ and the disturbance variable $0 \leqslant D \leqslant 1$, given by

$$I := \frac{G - G_{\min}}{G_{\max} - G_{\min}}, \quad D := \frac{F_{\max} - F}{F_{\max} - F_{\min}},$$

where $G_{\max} = 2/d^2$, $G_{\min} = 1/d^2$, $F_{\min} = 2/d^2$, and $F_{\max} = 1$. Note that $I=0$ ($D=0$) corresponds to no information (no disturbance) and $I=1$ ($D=1$) corresponds to maximum information (maximum disturbance).

Using Eq. (32) and the definitions of I and D, we immediately obtain the relation

$$x = \sqrt{D}, \quad y = \sqrt{1-I}. \qquad (33)$$

Substituting the above equations into the normalization condition of Eq. (31), we finally obtain the curve of the optimal tradeoff:

$$d^2(D-I)^2 - 4D(1-I) = 0. \qquad (34)$$

The corresponding plot is shown in Fig. 3.

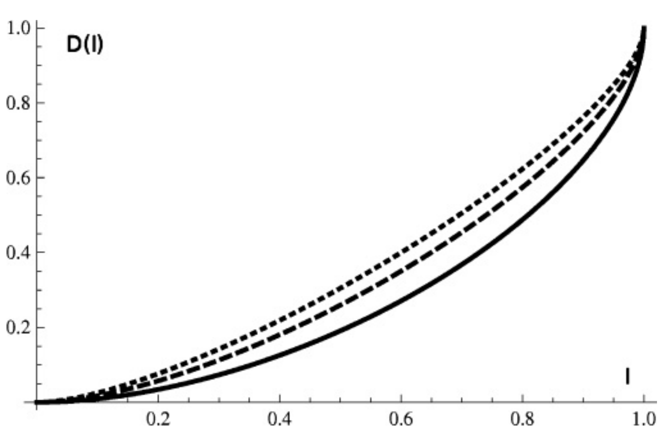


FIG. 3. Plot of the minimum disturbance $D(I)$ as a function of the information I [Eq. (34)], for various values of d: solid line, $d=2$; dashed line, $d=3$; dotted line, $d=4$.

D. Optimal quantum network

We now use Theorems 1 and 2 to construct explicitly the optimal network achieving the tradeoff of Fig. 3. The optimal network is derived for every possible value of the parameters x,y with $0 \leqslant x$, $y \leqslant 1$ belonging to the curve $x^2 + y^2 + 2xy/d = 1$. Since x and y can be easily expressed in terms of the information I and the disturbance D [Eq. (33)], our construction provides the optimal network for every point in the optimal tradeoff curve depicted in Fig. 3.

According to Theorem 2, the generalized instrument $\{R_{\hat{U}}\}$ is implemented as a sequence of $N=2$ isometries $V^{[1]}: \mathcal{H}_0 \to \mathcal{H}_1 \otimes \mathcal{H}_{A_1}$ and $V^{[2]}: \mathcal{H}_2 \otimes \mathcal{H}_{A_1} \to \mathcal{H}_3 \otimes \mathcal{H}_{A_2}$, followed by a measurement $\{P_{\hat{U}}\}$ on the ancilla A_2. The ancillary Hilbert space $\mathcal{H}_{A_1}$ ($\mathcal{H}_{A_2}$) is given by $\mathcal{H}_{A_1} = \mathsf{Supp}(R^{(1)*}_{1',0'}) \subseteq \mathcal{H}_{1'} \otimes \mathcal{H}_{0'}$ ($\mathcal{H}_{A_2} = \mathsf{Supp}(R^*_{3',2',1',0'}) \subseteq \mathcal{H}_{3'} \otimes \mathcal{H}_{2'} \otimes \mathcal{H}_{1'} \otimes \mathcal{H}_{0'}$), with $\mathcal{H}_{k'} \simeq \mathcal{H}_k$. The isometries $V^{[1]}, V^{[2]}$ are obtained from the realization of the deterministic comb $R = \int d\hat{U} R_{\hat{U}}$ (Theorem 1) and the ancilla measurement is given by the POVM

$$P_{\hat{U}} = (R^*)^{-\frac{1}{2}*} R^*_{\hat{U}} (R^*)^{-\frac{1}{2}}. \qquad (35)$$

Let us start from the construction of the isometries. By explicit calculation, we find

$$R = \left(\frac{x^2 d^2}{d^2-1} + y^2 d^2 + 2xyd\right)(P_{3,2} \otimes P_{1,0}) + \left(\frac{x^2}{d^2-1}\right)(I_{3,2,1,0} - P_{3,2} \otimes I_{1,0} - I_{3,2} \otimes P_{1,0}).$$

Taking the partial trace on $\mathcal{H}_3$ and using the condition $\mathrm{Tr}_3[R] = I_2 \otimes R^{(1)}$, we then obtain

$$R^{(1)} = \frac{(x+dy)^2}{d} P_{1,0} + \frac{x^2}{d}(I-P)_{1,0},$$

and, therefore,

$$(R^{(1)*})^{\frac{1}{2}} = \frac{1}{\sqrt{d}}(ydP_{1,0} + xI_{1,0}), \qquad (36)$$

$$(R^{(1)*})^{-\frac{1}{2}} = \sqrt{d}\left(\frac{-yd}{x(x+yd)}P_{1,0} + \frac{1}{x}I_{1,0}\right). \qquad (37)$$

According to Eq. (11), the isometry $V^{[1]}: \mathcal{H}_0 \to \mathcal{H}_1 \otimes \mathcal{H}_{A_1} \subseteq \mathcal{H}_{1,1',0'}$ is given by

$$V^{[1]} = \left[I_1 \otimes \left(R^{(1)*}_{1',0'}\right)^{\frac{1}{2}}\right](|I\rangle\rangle_{1,1'} \otimes T_{0'\leftarrow 0}) = \frac{1}{\sqrt{d}}(yT_{1\leftarrow 0} \otimes |I\rangle\rangle_{1',0'} + x|I\rangle\rangle_{1,1'} \otimes T_{0'\leftarrow 0}).$$

If we input a pure state $|\psi\rangle \in \mathcal{H}_0$, the output is then the superposition

$$V^{[1]}|\psi\rangle_0 = \frac{y}{\sqrt{d}}|\psi\rangle_1 |I\rangle\rangle_{1',0'} + \frac{x}{\sqrt{d}}|I\rangle\rangle_{1,1'}|\psi\rangle_{0'}.$$

Intuitively, we can understand the action of V_1 as a superposition of two different processes:

(1) With amplitude y, the quantum state $|\psi\rangle_0$ is propagated undisturbed from the input system 0 to the output system 1 so that the unknown unitary U can act on it. As we see in the following, the maximally entangled state $|\Phi\rangle\rangle_{1',0'} = \frac{1}{\sqrt{d}}|I\rangle\rangle_{1',0'}$

then serves as a resource to teleport the state $U|\psi\rangle_2$ to the output node 3.

(2) With amplitude x, the state $|\psi\rangle_0$ is transferred to the ancillary degree of freedom $0'$: in this case the unknown unitary does not act on it, but, instead, it acts on the maximally entangled state $|\Phi\rangle\rangle_{1,1'} = \frac{1}{\sqrt{d}}|I\rangle\rangle_{1,1'}$, thus producing the state $|\Phi_U\rangle\rangle_{2,1'} = \frac{1}{\sqrt{d}}|U\rangle_{2,1'}$. As we see in the following, the state $|\Phi_U\rangle\rangle_{2,1'}$ is used for the optimal estimation of U. Finally, the state $|\psi\rangle_{0'}$ is transferred to the output system 3, and, depending on the estimate, a transformation $\hat{U}$ is applied on it.

The isometry $V^{[2]}\mathcal{H}_{2,1',0'} \supseteq \mathcal{H}_2 \otimes \mathcal{H}_{A_1} \to \mathcal{H}_3 \otimes \mathcal{H}_{A_2} \subseteq \mathcal{H}_{3,3',2',1',0'}$ is given by

$$V^{[2]} = \left\{ I_3 \otimes \left[(R^*_{3',2',1',0'})^{\frac{1}{2}} \left(I_{3',2'} \otimes R^{(1)*}_{1',0'} \right)^{-\frac{1}{2}} \right] \right\} \times (|I\rangle\rangle_{3,3'} \otimes T_{2'\leftarrow 2} \otimes I_{1',0'}). \qquad (38)$$

On the other hand, using Eqs. (35) and (30), the POVM $\{P_{\hat{U}}\}$ on $\mathcal{H}_{A_2}$ can be written as

$$P_{\hat{U}} = |\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|, \quad |\eta_{\hat{U}}\rangle := (R^*)^{-\frac{1}{2}}|\chi^*_{\hat{U}}\rangle. \qquad (39)$$

Combining the isometry $V^{[2]}$ with the POVM $\{P_{\hat{U}}\}$, we then obtain the instrument $\{\mathcal{T}_{\hat{U}}\}$, with $\mathcal{T}_{\hat{U}} : L(\mathcal{H}_2 \otimes \mathcal{H}_{A_2}) \to L(\mathcal{H}_3)$ given by

$$\mathcal{T}_{\hat{U}}(\rho) = \mathrm{Tr}_{A_2}[V^{[2]}\rho V^{[2]\dagger}(I_3 \otimes P_{\hat{U}})], \quad \forall \rho \in L(\mathcal{H}_2 \otimes \mathcal{H}_{A_1}).$$

We now use Eqs. (38) and (39) to show that the instrument $\{\mathcal{T}_{\hat{U}}\}$ has a very simple form. To construct $\mathcal{T}_{\hat{U}}$ explicitly, we start from the Kraus form $\mathcal{T}_{\hat{U}}(\rho) = K_{\hat{U}}\rho K^\dagger_{\hat{U}}$, where the Kraus operator $K_{\hat{U}}$ is given by

$$\begin{aligned} K_{\hat{U}} &= (I_3 \otimes \langle \eta_{\hat{U}}|_{3',2',1',0'})V^{[2]} \\ &= (I_3 \otimes \langle \chi^*_{\hat{U}}|_{3',2',1',0'})\left(I_3 \otimes I_{3',2'} \otimes R^{(1)*}_{1',0'}\right)^{-\frac{1}{2}} \\ &\quad \times (|I\rangle\rangle_{3,3'} \otimes T_{2'\leftarrow 2} \otimes I_{1',0'}), \end{aligned}$$

having used Eqs. (38) and (39). Now, from Eq. (30), we have $\langle \chi^*_{\hat{U}}|_{3',2',1',0'} = \langle \chi|_{3',2',1',0'}(\hat{U}^T_{3'} \otimes \hat{U}^\dagger_{2'} \otimes I_{1',0'})$ and, therefore,

$$\begin{aligned} K_{\hat{U}} &= (I_3 \otimes \langle \chi|_{3',2',1',0'})\left(I_3 \otimes I_{3',2'} \otimes R^{(1)*}_{1',0'}\right)^{-\frac{1}{2}} \\ &\quad \times (|\hat{U}\rangle\rangle_{3,3'} \otimes \hat{U}^\dagger_{2'} T_{2'\leftarrow 2} \otimes I_{1',0'}). \end{aligned}$$

Finally, inserting Eqs. (29) and (37) in the above expression, we obtain

$$K_{\hat{U}} = \sqrt{d}(\langle\langle \hat{U}|_{2,1'} \otimes \hat{U}_3 T_{3\leftarrow 0'}),$$

which implies that the instrument $\{\mathcal{T}_{\hat{U}}\}$ is given by

$$\mathcal{T}_{\hat{U}}(\rho) = \hat{U}_3 T_{3\leftarrow 0'}(d\langle\langle \hat{U}|_{2,1'}\rho|\hat{U}\rangle\rangle_{2,1'})T_{0'\to 3}\hat{U}^\dagger_3, \qquad (40)$$

where we introduced the redundant notation $T_{0'\to 3} := T^\dagger_{3\leftarrow 0'} = \sum_{n=1}^d |n\rangle_{0'}\langle n|_3 \equiv T_{0'\leftarrow 3}$ to make the expression clearer.

The interpretation of Eq. (40) is straightforward. To implement the instrument $\{\mathcal{T}_{\hat{U}}\}$ we only have to perform on system 2 and on the ancila $1'$ the Bell measurement $\{B_{\hat{U}}\}$ with POVM $B_{\hat{U}} = d|\hat{U}\rangle\rangle\langle\langle\hat{U}|$ and, depending on the outcome, to perform the unitary $\hat{U}$ on output system 3, which is obtained from the ancilla $0'$ just by relabeling (represented here by the teleportation operator $T_{3\leftarrow 0'}$). In other words, the instrument $\{\mathcal{T}_{\hat{U}}\}$ is just obtained by a Bell measurement followed by unitary feed-forward. Remarkably, $\{\mathcal{T}_{\hat{U}}\}$ is independent of the tradeoff parameters x,y; this means that, after we perform the isometry $V^{[1]}$, the remaining part of the optimal network is independent of the particular value of the information-disturbance rate. The reason for this is that the combination of Bell measurement and feed-forward realized by the instrument $\{\mathcal{T}_{\hat{U}}\}$ can work both as a teleportation protocol (in the case of no disturbance) and as an estimate-and-reprepare strategy (in the case of maximal information extraction).

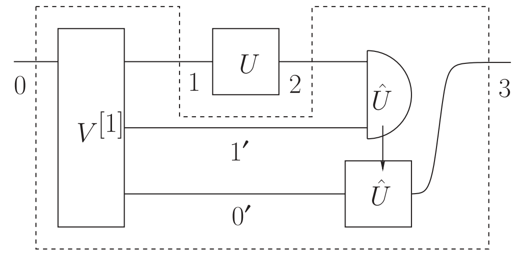


FIG. 4. Optimal quantum network for the information-disturbance tradeoff. The input state $|\psi\rangle$ enters from the wire 0. Then, the isometry $V^{[1]}$ prepares a coherent superposition $\frac{1}{\sqrt{d}}(y|\psi\rangle_1|I\rangle\rangle_{1',0'} + x|I\rangle\rangle_{1,1'}|\psi\rangle_{0'})$, which is tuned by the parameters $x = \sqrt{D}$ and $y = \sqrt{1-I}$, whose values depend on the information-disturbance rate. After that, the unknown unitary U is applied between nodes 1 and 2 of the network. Finally, a Bell measurement is performed and, depending on the result, the unitary transformation $\hat{U}$ is performed on output system 3.

To summarize the results of this section, we give the step-by-step evolution of a pure state $|\psi\rangle_0$ in the optimal quantum network:

$$\begin{aligned} |\psi\rangle_0 &\xrightarrow{V^{[1]}} \frac{1}{\sqrt{d}}(y|\psi\rangle_1|I\rangle\rangle_{1',0'} + x|I\rangle\rangle_{1,1'}|\psi\rangle_{0'}) \\ &\xrightarrow{U} \frac{1}{\sqrt{d}}(yU|\psi\rangle_2|I\rangle\rangle_{1',0'} + x|U\rangle\rangle_{2,1'}|\psi\rangle_{0'}) \\ &\xrightarrow{\sqrt{d}\langle\langle\hat{U}|} y\hat{U}^\dagger U|\psi\rangle_3 + x\mathrm{Tr}[\hat{U}U^\dagger]|\psi\rangle_3 \\ &\xrightarrow{\hat{U}} yU|\psi\rangle_3 + x\mathrm{Tr}[\hat{U}U^\dagger]\hat{U}|\psi\rangle_3. \end{aligned}$$

The action of the whole network is depicted in Fig. 4.

V. CONCLUSIONS

In this work we addressed the fundamental problem of the information-disturbance tradeoff in the estimation of an unknown quantum transformation. In particular, we completely solved the problem in the case of a unitary transformation, randomly distributed according to the Haar measure.

Interestingly, the analytical expression of the optimal tradeoff curve given in Eq. (34) happens to coincide with the tradeoff curve for the estimation of a maximally entangled state [11]. Note, however, that this is not a trivial consequence of the Choi isomorphism $U \to \frac{1}{\sqrt{d}}|U\rangle$: while this mathematical correspondence is one-to-one, operationally it cannot be inverted *with unit probability*. In other words, once the transformation U has been applied to the maximally entangled state $\frac{1}{\sqrt{d}}|I\rangle\rangle$, it is irreversibly degraded and can be retrieved only probabilistically. For this reason, there is no operational relation between the information-disturbance tradeoff for unitary transformations and that for maximally entangled states (none of them is a primitive for the other). Indeed, the optimal quantum network for unitary transformations depicted in Fig. 4 is quite different from the optimal network for maximally entangled states. In our case the optimal network consists of (i) a first interaction that produces a quantum superposition with amplitudes depending on the desired information-disturbance rate and (ii) a Bell measurement followed by unitary feed-forward.

Besides its fundamental relevance, the information-disturbance tradeoff for transformations is also interesting as a possible eavesdropping strategy in cryptographic protocols where the secret key is encoded in a set of unitary transformations, as it happens in the two-way protocols of Refs. [17–20]. Notice, however, that for protocols where the secret key is encoded in a set of *orthogonal* unitaries, like those of Refs. [17,18,20], the security of the protocol is not based on the information-disturbance tradeoff studied in this paper. Indeed, since the unitaries are orthogonal, they can be estimated and re-prepared without introducing any disturbance (or just introducing a vanishing disturbance, in the infinite-dimensional case). This is the reason why the protocols of Refs. [17,18,20] necessarily require random switching between a communication mode and a control mode. The present analysis is instead relevant for the analysis of the two-way protocol of Ref. [19], which uses two mutually unbiased bases of orthogonal qubit unitaries, given by $\mathcal{B}_1 = \{\sigma_\mu\}_{\mu=0}^3$ and $\mathcal{B}_2 = \{U\sigma_\mu\}_{\mu=0}^3$, where $\sigma_0 = I$, $\{\sigma_k\}_{k=1}^3$ are the three Pauli matrices, and $U = (I + i\sum_{k=1}^3 \sigma_k)/2$ is the rotation of $2\pi/3$ around the axis $\mathbf{n} = 1/\sqrt{3}(1,1,1)$. In this case, using the optimal network is a nontrivial eavesdropping attack. Of course, since the protocol does not involve all possible qubit unitaries, the optimal tradeoff curve for the restricted set $\mathcal{B}_1 \cup \mathcal{B}_2$ could possibly be more favorable to the eavesdropper than the universal tradeoff curve derived in this paper. The analysis of the tradeoff for nonuniversal sets of unitary transformations and the study of the relations between information-disturbance tradeoff and quantum cloning for unitary transformations [19] are interesting directions of future research.

ACKNOWLEDGMENTS

We acknowledge an anonymous referee for useful comments and suggestions. This work is supported by the Italian Ministry of Education through Grant No. PRIN 2008 and through the EC through the project COQUIT. Research at the Perimeter Institute for Theoretical Physics is supported in part by the government of Canada through NSERC and by the Province of Ontario through MRI.

[1] W. Heisenberg, Z. Phys. **43**, 172 (1927).
[2] M. O. Scully, B.-G. Englert, and H. Walther, Nature **351**, 111 (1991).
[3] S. Stenholm, Ann. Phys. (NY) **218**, 233 (1992).
[4] B.-G. Englert, Phys. Rev. Lett. **77**, 2154 (1996).
[5] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).
[6] H. Barnum, e-print arXiv:quant-ph/0205155.
[7] K Banaszek, Phys. Rev. Lett. **86**, 1366 (2001).
[8] G. M. D'Ariano, Fortschr. Phys. **51**, 318 (2003).
[9] M. Ozawa, Ann. Phys. (NY) **311**, 350 (2004).
[10] L. Mišta Jr., J. Fiuràšek, and R. Filip, Phys. Rev. A **72**, 012311 (2005).
[11] M. F. Sacchi, Phys. Rev. Lett. **96**, 220502 (2006).
[12] F. Sciarrino, M. Ricci, F. De Martini, R. Filip, and L. Mišta Jr., Phys. Rev. Lett. **96**, 020408 (2006).
[13] L. Maccone, Phys. Rev. A **73**, 042307 (2006).
[14] D. Kretschmann, D. Schlingemann, and R. F. Werner, IEEE Trans. Inf. Theory **54**, 1708 (2008).
[15] F. Buscemi, M. Hayashi, and M. Horodecki, Phys. Rev. Lett. **100**, 210504 (2008).
[16] G. M. D'Ariano, S. Facchini, P. Perinotti, and M. F. Sacchi, Phys. Lett. A **373**, 3011 (2009).
[17] K. Boström and T. Felbinger, Phys. Rev. Lett. **89**, 187902 (2002).
[18] M. Lucamarini and S. Mancini, Phys. Rev. Lett. **94**, 140501 (2005).
[19] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180504 (2008).
[20] S. Pirandola, S. Mancini, S. Lloyd, and S. L. Braunstein, Nature Phys. **4**, 726 (2008).
[21] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).
[22] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Europhys. Lett. **83**, 30004 (2008).
[23] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).
[24] M.-D. Choi, Linear Algebra Appl. **10**, 285 (1975).
[25] G. Gutoski and J. Watrous, in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computation* (ACM, New York, 2007), pp. 565–574.
[26] A. Bisio, G. Chiribella, G. M. D'Ariano, and P. Perinotti, e-print arXiv:1006.1780v1.
[27] M. Ozawa, J. Math. Phys. **25**, 79 (1984).
[28] M. Raginsky, Phys. Lett. A **290**, 11 (2001).
[29] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **72**, 042338 (2005).