Distance measures to compare real and ideal quantum processes

Alexei Gilchrist,[1,*] Nathan K. Langford,[1,†] and Michael A. Nielsen[2,‡]
[1]*Centre for Quantum Computer Technology and Department of Physics, The University of Queensland, Brisbane, Queensland 4072, Australia*
[2]*School of Physical Sciences and School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Queensland 4072, Australia*
(Received 26 August 2004; published 13 June 2005)

With growing success in experimental implementations it is critical to identify a "gold standard" for quantum information processing, a single measure of distance that can be used to compare and contrast different experiments. We enumerate a set of criteria that such a distance measure must satisfy to be both experimentally and theoretically meaningful. We then assess a wide range of possible measures against these criteria, before making a recommendation as to the best measures to use in characterizing quantum information processing.

DOI: 10.1103/PhysRevA.71.062310 PACS number(s): 03.67.Lx

I. INTRODUCTION

Many real-world imperfections arise when experimentally performing a quantum information processing task. These may arise either in the creation or measurement of a quantum state, or in the manipulation of the state via some quantum process. It is important to quantitatively measure and characterize these imperfections in a way that is theoretically meaningful and experimentally practical.

How can this be done? Quantum states can be completely determined using quantum state tomography [1,2] and compared using a variety of well-known measures [3]. Quantum processes can be measured using an analogous procedure called quantum process tomography [3–5]. However, the problem of developing quantitative measures to compare real and idealized quantum processes has not been comprehensively addressed.

Ideally there would be a single good measure, a "gold standard" [6,7], enabling sensible comparison of different experimental implementations of quantum information processing, and agreed upon by experimentalists and theorists alike. We will refer to candidates for such a gold standard as "distance measures" for quantum processes, or as "error measures" when we want to stress the comparison of real and idealized processes.

Such an error measure would be extremely useful both when comparing experiments with the theoretical ideal, and in comparing different experiments that attempt to perform the same task. Existing experiments in quantum information processing have typically been assessed on a rather *ad hoc* basis. For example, some implementations of quantum logic gates have relied on demonstrating that those gates act in the correct way on computational basis states (i.e., verifying the truth table of the gate) and a few superposition states. Such demonstrations are important, but it is clear that a figure of merit that is standardized, theoretically well motivated and experimentally practical would be a considerable step forward. Parenthetically, we note that such a measure would also be of great use in concretely connecting real experiments to results such as the fault-tolerance threshold for quantum computation [8].

The purpose of this paper is to comprehensively address the problem of developing such error measures. There is a sizeable previous literature on this subject, but we believe that there has been a consistent gap between work motivated primarily by theoretical considerations, and work constrained by experimental realities. Our paper aims to address both theoretical and experimental desiderata.

The key to our work is to introduce a list of six simple, physically motivated criteria that should be satisfied by any good measure of distance between quantum processes. These criteria enable us to eliminate many approaches to the definition of an error measure that *a priori* appear highly plausible.

The criteria are as follows. Suppose Δ is a candidate measure of the distance between two quantum processes. Such processes are described by maps between input and output quantum states, e.g., $\rho_{\text{out}}=\mathcal{E}(\rho_{\text{in}})$, where the map $\mathcal{E}$ is known as a *quantum operation* [3,9]. Physically, $\Delta(\mathcal{E},\mathcal{F})$ may be thought of in two ways, as a measure of error in quantum information processing when one wants to do the ideal process $\mathcal{F}$ but does $\mathcal{E}$ instead, or of distinguishability between the two processes $\mathcal{E}$ and $\mathcal{F}$. We believe that any such measure must satisfy the following six properties, motivated by both physical and mathematical concerns.

(1) *Metric:* Δ should be a metric. This requires three properties, (i) $\Delta(\mathcal{E},\mathcal{F})\geqslant 0$ with $\Delta(\mathcal{E},\mathcal{F})=0$ if and only if $\mathcal{E}=\mathcal{F}$; (ii) symmetry, $\Delta(\mathcal{E},\mathcal{F})=\Delta(\mathcal{F},\mathcal{E})$; and (iii) the triangle inequality, $\Delta(\mathcal{E},\mathcal{G})\leqslant\Delta(\mathcal{E},\mathcal{F})+\Delta(\mathcal{F},\mathcal{G})$.

(2) *Easy to calculate:* it should be possible to evaluate Δ in a direct manner.

(3) *Easy to measure:* there should be a clear and achievable experimental procedure for determining the value of Δ.

(4) *Physical interpretation:* Δ should have a well-motivated physical interpretation.

(5) *Stability [10]:* $\Delta(\mathcal{I}\otimes\mathcal{E},\mathcal{I}\otimes\mathcal{F})=\Delta(\mathcal{E},\mathcal{F})$, where $\mathcal{I}$ rep-

*Electronic address: alexei@physics.uq.edu.au
†Electronic address: langford@physics.uq.edu.au
‡Electronic address: nielsen@physics.uq.edu.au; URL: www.qinfo.org/people/nielsen

resents the identity operation on an additional quantum system. Physically, this means that unrelated ancillary quantum systems do not affect the value of Δ.

(6) *Chaining:* $\Delta(\mathcal{E}_2\circ\mathcal{E}_1,\mathcal{F}_2\circ\mathcal{F}_1)\leqslant\Delta(\mathcal{E}_1,\mathcal{F}_1)+\Delta(\mathcal{E}_2,\mathcal{F}_2)$. Thus, for a process composed of many smaller steps, the total error will be less than the sum of the errors in the individual steps.

The chaining and stability criteria are key properties for estimating the error in a complex quantum information processing task. Because quantum information processing tasks are typically broken down into a sequence of simpler component operations, a conservative bound on the total error can be found by simply analyzing the individual components. This is critical for applications such as quantum computation, where full process tomography on an *n*-qubit computation requires exponentially many measurements, and is thus infeasible. Chaining and stability enable one to instead benchmark the constituent processes involved in the computation, which can then be used to infer that the entire computation is robust.

Many other properties follow from these six criteria. For example, from the metric and chaining criteria we see that $\Delta(\mathcal{R}\circ\mathcal{E},\mathcal{R}\circ\mathcal{F})\leqslant\Delta(\mathcal{E},\mathcal{F})$, where $\mathcal{R}$ is any quantum operation. This corresponds to the requirement that postprocessing by $\mathcal{R}$ cannot *increase* the distinguishability of two processes $\mathcal{E}$ and $\mathcal{F}$. Another elementary consequence of the metric and chaining criteria is *unitary invariance*, i.e., $\Delta(\mathcal{U}\circ\mathcal{E}\circ\mathcal{V},\mathcal{U}\circ\mathcal{F}\circ\mathcal{V})=\Delta(\mathcal{E},\mathcal{F})$, where $\mathcal{U}$ and $\mathcal{V}$ are unitary operations.

For both theoreticians and experimentalists, there are strong motivations to find a gold standard satisfying these criteria—the need for a physically sensible way of evaluating the performance of a quantum process, and the need to compare the success of a theoretical model to the operation of a real, experimental system. For the experimentalist, however, there is also another important consideration. That is the need for *diagnostic measures* which can be used to build insight into the source of imperfections in experimental implementations. Diagnostic measures may not necessarily be good candidates for our sought-after gold standard—they may fail to satisfy one or more of our criteria—but they still may be extremely useful in the experimental context. Thus, some of the measures we discard as unsuitable for use as a gold standard may still be useful as diagnostic measures. Furthermore, it is not difficult to construct other examples of useful diagnostic measures which are different to any considered in this paper. The detailed investigation of such diagnostic measures is, however, beyond the scope of the present paper.

Prior work: The principal contribution of our paper is to comprehensively evaluate many plausible error measures for quantum information processing, within the broad framework of the criteria we have identified. So far as we are aware, none of the prior work has surveyed and compared error measures against such a broad array of theoretical and experimental concerns.

Error measures for quantum teleportation have received particular attention in the prior literature, perhaps spurred by controversy over which experiments should be regarded as definitively demonstrating the teleportation effect [11]. Examples of this line of development include [12–17], and references therein. With the exception of Ref. [17] this work differs from ours in that it is focused primarily on the problem of teleportation. Reference [17] has a more general focus, but is not primarily concerned with the development of error measures, but rather with the question of when quantum information processing can be modeled classically.

More mathematical investigations of error measures have also been mounted, especially in the context of quantum communication and fault-tolerant quantum computation. Examples of this work include [10,18–26], and references therein. This work (often embedded in some larger investigation) typically focuses on one or a few measures of specific interest for the problem at hand. These papers thus differ from our work in that they do not attempt a comprehensive survey of possible error measures against some set of abstract criteria; nor, typically, do they address experimental criteria such as ease of measurement. Nonetheless, while this prior work is different in character from ours, it has greatly informed our point of view, and we will have occasion to cite it on specific points throughout this paper. Of particular relevance is Ref. [10], which introduced one of the key measures we use, the stabilized process distance, or *S* distance (referred to as the diamond norm in Ref. [10]), and emphasized some of the important properties satisfied by that measure.

Structure of the paper: Secs. II and III summarize background material on quantum operations and distance measures for quantum states.

Section IV is the core of the paper, comprehensively surveying possible approaches to the definition of error measures. Our strategy is to cast a wide net, considering many different possible approaches to the definition of a distance measure, and then to use our list of criteria to eliminate as many approaches as possible. This means a certain amount of tedium as we propose and then reject certain *a priori* plausible candidate error measures. The benefit of going through this process of elimination is considerable, however. First, it gives us confidence that the few measures we identify as particularly promising should be preferred over all other measures. Indeed, we quickly eliminate all but four of the measures we define as follows: the *Jamiolkowski process fidelity* (*J* fidelity), the *Jamiolkowski process distance* (*J* distance), the *stabilized process fidelity* (*S* fidelity), and the *stabilized process distance* (*S* distance). Second, in several instances we show that error measures proposed previously in the literature (in one case, by one of the authors of this paper) should be rejected as inadequate.

Section V applies the four promising measures identified in Sec. IV to the concrete problem of quantum computation, showing that each measure has a useful operational interpretation in terms of the success or failure of a quantum computation.

Section VI concludes the paper with a summary of our results, and the identification of the *S* distance and the *S* fidelity as the two measures whose properties make them the most attractive candidates for use as a gold standard in quantum information processing. We do not make a final recommendation as to which of these two measures should be used, since they have extremely similar strengths and weaknesses.

However, we do discuss and make definite recommendations regarding the reporting of quantum information processing experiments. Furthermore, we sketch future research directions which may ameliorate some of the weaknesses of one or both measures, and which may therefore make it possible to definitively choose a single measure as a gold standard.

II. DESCRIBING QUANTUM PROCESSES

Quantum operations describe the most general physical processes that may occur in a quantum system [3], including unitary evolution, measurement, noise, and decoherence. Any quantum operation may be given the *operator-sum representation* relating input ρ_{in} and output ρ_{out} states,

$$\rho_{\text{out}} = \mathcal{E}(\rho_{\text{in}}) = \sum_j E_j \rho_{\text{in}} E_j^\dagger, \tag{1}$$

where the operators E_j are known as *operation elements*, and obey the condition that $\Sigma_j E_j^\dagger E_j \leqslant I$ [27]. Note that the operation elements $\{E_j\}$ completely describe the effect of the process. We will mostly be concerned with the case of trace-preserving operations, for which $\Sigma_j E_j^\dagger E_j = I$. Physically, this corresponds to the requirement that $\mathcal{E}$ represents a physical process without post-selection [28]. Many of our results extend easily to the case of non-trace-preserving operations, but to ease the exposition we assume processes are trace-preserving unless otherwise noted.

The operator-sum representation has the drawback that it is not unique, in the sense that there is a freedom in the choice of operation elements [3]. This is inconvenient if we are trying to compare two processes. To alleviate this, let us fix a basis $\{A_j\}$ for the space of operators, choosing for convenience a basis orthonormal under the Hilbert-Schmidt inner product, i.e., $\text{tr}(A_j^\dagger A_k) = \delta_{jk}$ [29]. We can use this basis to expand the operation elements, $E_j = \Sigma_m a_{jm} A_m$, and rewrite Eq. (1),

$$\mathcal{E}(\rho) = \sum_{mn} (\chi_{\mathcal{E}})_{mn} A_m \rho A_n^\dagger, \tag{2}$$

where $(\chi_{\mathcal{E}})_{mn} \equiv \Sigma_j a_{jm} a_{jn}^*$ are the elements of the *process matrix*, $\chi_{\mathcal{E}}$. Equation (2) tells us that the process matrix completely describes the action of the quantum process. The big advantage of the process matrix representation is that, unlike the operator-sum representation, once the basis $\{A_j\}$ is chosen the process matrix can be shown to be unique to the process [30]; i.e., it depends only on $\mathcal{E}$, not on the particular choice of operation elements $\{E_j\}$. We will not give an explicit proof of this fact here, but note that this result follows easily from the discussion below.

The process matrix gives a convenient way of representing the operation $\mathcal{E}$. A closely related but more abstract representation is provided by the *Jamiolkowski isomorphism* [31], which relates a quantum operation $\mathcal{E}$ to a quantum state, $\rho_{\mathcal{E}}$,

$$\rho_{\mathcal{E}} \equiv [\mathcal{I} \otimes \mathcal{E}](|\Phi\rangle\langle\Phi|), \tag{3}$$

where $|\Phi\rangle = \Sigma_j |j\rangle|j\rangle / \sqrt{d}$ is a maximally entangled state of the (d-dimensional) system with another copy of itself, and $\{|j\rangle\}$ is some orthonormal basis set. The map $\mathcal{E} \to \rho_{\mathcal{E}}$ is invertible, that is, knowledge of $\rho_{\mathcal{E}}$ is equivalent to knowledge of $\mathcal{E}$ [32]. This isomorphism thus allows us to treat quantum operations using the same tools as are ordinarily used to treat quantum states. For later use we note the useful property $\rho_{\mathcal{E}\otimes\mathcal{F}} = \rho_{\mathcal{E}} \otimes \rho_{\mathcal{F}}$.

The state $\rho_{\mathcal{E}}$ and the process matrix $\chi_{\mathcal{E}}$ are closely related. A direct calculation shows that if one chooses the operator basis sets $\{A_j\} = \{|m\rangle\langle n|\}$, then $\chi_{\mathcal{E}} = d\rho_{\mathcal{E}}$, as matrices. Thus we shall refer to both $\chi_{\mathcal{E}}$ and $\rho_{\mathcal{E}}$ as the process matrix, and treat them interchangeably. This is very convenient, as $\rho_{\mathcal{E}}$ is easy to work with mathematically, using the expression Eq. (3), while the elements of $\chi_{\mathcal{E}}$ have an obvious physical significance, expressed by Eq. (2).

We conclude this section with a comment on our notational conventions. We often use notation like ψ to denote either a pure state $|\psi\rangle$ or the corresponding density matrix $|\psi\rangle\langle\psi|$, with the meaning to be determined from context. Thus, for example, we may write $\psi = \alpha|0\rangle + \beta|1\rangle$ to indicate a pure state of a single qubit, while also writing $\mathcal{E}(\psi)$ to indicate a quantum operation $\mathcal{E}$ acting on the density matrix corresponding to that pure state.

III. DISTANCE MEASURES FOR QUANTUM STATES

A natural starting place for an attempt to define a measure of distance for quantum processes is to consider measures of distance for quantum states. The quantum information science community has identified the *trace distance* and the *fidelity* as particularly important approaches to the definition of a distance measure for states [33], and these two measures will serve as the basis for our later definitions of distance measures for quantum operations. In keeping with the aims of the paper, we do not make a choice between the trace distance and the fidelity at the outset. Instead, our preference is to develop distance measures for quantum operations based on *both* the trace distance and the fidelity, and then assess them using the criteria discussed in the introduction. We now briefly review the basic properties of the trace distance and the fidelity.

The trace distance: The *trace distance* between density matrices ρ and σ is defined by $D(\rho,\sigma) \equiv \frac{1}{2}\text{tr}|\rho - \sigma|$, where $|X| \equiv \sqrt{X^\dagger X}$. From this definition it follows that the trace distance is a genuine metric on quantum states, with $0 \leqslant D \leqslant 1$. The trace distance also has many other attractive properties that make it a particularly good measure of distance between quantum states. We now briefly describe three of these.

First, the trace distance has a compelling physical interpretation as a measure of state distinguishability. Suppose Alice prepares a quantum system in the state ρ with probability $\frac{1}{2}$, and in the state σ with probability $\frac{1}{2}$. She gives the system to Bob, who performs a POVM measurement [3] to distinguish the two states. It can be shown that Bob's probability of correctly identifying which state Alice prepared is $1/2 + D(\rho,\sigma)/2$. That is, $D(\rho,\sigma)$ can be interpreted, up to the factor $1/2$, as the optimal *bias* in favor of Bob correctly determining which of the two states was prepared. This physical interpretation follows from the identity

$D(\rho,\sigma)=\max_{E\leqslant I}\mathrm{tr}[E(\rho-\sigma)]$ [34], where the maximum is over all positive operators E satisfying $E\leqslant I$.

Second, the trace distance possesses the *contractivity* property [35], that is, $D(\mathcal{E}(\rho),\mathcal{E}(\sigma))\leqslant D(\rho,\sigma)$ whenever $\mathcal{E}$ is a trace-preserving quantum operation. This statement expresses the physical fact that a quantum process acting on two quantum states cannot increase their distinguishability. Contractivity follows from the physical interpretation of $D(\rho,\sigma)$ described above.

Third, the trace distance is *doubly convex*, i.e., if p_j are probabilities then $D(\Sigma_j p_j\rho_j,\Sigma_j p_j\sigma_j)\leqslant\Sigma_j p_j D(\rho_j,\sigma_j)$. This inequality can be physically interpreted as the statement that the distinguishability between the states $\Sigma_j p_j\rho_j$ and $\Sigma_j p_j\sigma_j$, where j is not known, can never be greater than the average distinguishability when j is known, but has been chosen at random according to the distribution p_j.

Fidelity: The *fidelity* between density matrices ρ and σ is defined by

$$F(\rho,\sigma)\equiv\mathrm{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2. \tag{4}$$

When $\rho=\psi$ is a pure state, this reduces to $F(\psi,\sigma)=\langle\psi|\sigma|\psi\rangle$, the overlap between ψ and σ.

The fidelity also has many attractive properties. It can be shown that $0\leqslant F(\rho,\sigma)\leqslant 1$, with equality in the second inequality if and only if $\rho=\sigma$. The fidelity is thus not a metric as such, but serves rather as a generalized measure of the overlap between two quantum states. The fidelity is also symmetric in its inputs, $F(\rho,\sigma)=F(\sigma,\rho)$, a fact that is not obvious from the definition we have given, but which follows from other equivalent definitions.

There is an ambiguity in the literature in the definition of fidelity that is worth commenting on here. Both the quantity defined above and its square root have been referred to as the fidelity, and both have many appealing properties [36].

Nevertheless, we strongly advocate using the definition of Eq. (4), despite the other definition being used in references such as Ref. [3]. As we will see in Sec. V, adopting the definition of Eq. (4) gives rise to a measure of distance between quantum processes with a physically compelling interpretation in terms of the *probability of success* of a quantum computation. Adopting the other definition of fidelity would make about as much sense as reporting the square root of the probability that the quantum computation succeeded.

Although not a metric, the fidelity can easily be turned into a metric. Two common ways of doing this are the *Bures metric*, defined by $B(\rho,\sigma)\equiv\sqrt{2-2\sqrt{F(\rho,\sigma)}}$, and the *angle*, defined by $A(\rho,\sigma)\equiv\arccos\sqrt{F(\rho,\sigma)}$. The origin of these metrics can be seen intuitively by considering the case when ρ and σ are both pure states. The Bures metric is just the Euclidean distance between the two pure states, with respect to the usual norm on state space [37], while the angle is, as the name suggests, just the angle between the two states, with respect to the usual inner product on state space.

In addition to the angle and the Bures metric we will find it convenient to introduce a third metric based on the fidelity. This metric does not seem to have been previously recognized in the literature, but arises naturally later in this paper in the context of quantum computation. It is defined by $C(\rho,\sigma)\equiv\sqrt{1-F(\rho,\sigma)}$. The only difficult step in proving this is a metric is the proof of the triangle inequality [38,39].

In later sections our discussion will sometimes focus on the fidelity, and sometimes on metrics derived from the fidelity. We will say that a metric $\Delta^F(\rho,\sigma)$ on state space is a *fidelity-based* metric if it is a monotonically decreasing function of the fidelity $F(\rho,\sigma)$. Obviously the angle, the Bures metric and $C(\cdot,\cdot)$ are all fidelity-based metrics. It is often the case that the specific details of the metric used are not important, and whenever possible we state results using the fidelity as a single unifying concept. However, sometimes it will prove advantageous to use the fidelity-based metrics directly. In particular, they have the advantage of satisfying the triangle inequality, which turns out to be useful proving the chaining criterion [property (6)].

Like the trace distance, the fidelity and its derived metrics have many other nice properties. It can be shown [40] that $F(\mathcal{E}(\rho),\mathcal{E}(\sigma))\geqslant F(\rho,\sigma)$ for any trace-preserving quantum operation $\mathcal{E}$. We call this the *monotonicity* property of the fidelity. It follows that any fidelity-based metric satisfies a contractivity property analogous to that satisfied by the trace distance.

The fidelity also satisfies a property analogous to the double convexity of the trace distance. Precisely, the square root of the fidelity is *doubly concave*, that is, $F(\Sigma_j p_j\rho_j,\Sigma_j p_j\sigma_j)^{1/2}\geqslant\Sigma_j p_j F(\rho_j,\sigma_j)^{1/2}$. This double concavity can be used to prove double convexity of certain fidelity-based metrics. In particular, supposing Δ^F is a fidelity-based metric which is convex in the square root of the fidelity [the angle, the Bures metric and $C(\cdot,\cdot)$ are all easily verified to have this property], then it is easy to verify that Δ^F is doubly convex.

One drawback of the fidelity is that it is difficult to find a compelling physical interpretation. When ρ and σ are mixed states, no completely satisfactory interpretation of the fidelity is known (but cf. Refs. [41,42]). When $\rho=\psi$ is a pure state, we have $F(\psi,\sigma)=\langle\psi|\sigma|\psi\rangle$, the overlap between ψ and σ. Physically, we might imagine σ is an attempt to prepare the pure state ψ. In this case the fidelity coincides with the probability that a perfect measurement testing whether the state is ψ will succeed. It is this property of the fidelity that is used in Sec. V to connect our fidelity-based error measures for quantum processes to the probability of success of a quantum computation.

General comments: The fidelity is, at present, perhaps somewhat more widely used in the quantum information science community than is the trace distance. However, we shall see below that the trace distance and the fidelity have complementary advantages as a basis for developing measures of distance for quantum operations, and so it is useful to investigate both. In any case, the two measures are, as one might expect, quite closely related. In particular, it is possible to show that they are related by the inequalities [43],

$$1-\sqrt{F(\rho,\sigma)}\leqslant D(\rho,\sigma)\leqslant\sqrt{1-F(\rho,\sigma)}. \tag{5}$$

It is not difficult to construct examples of saturation for both inequalities. Note that the second inequality is always satu-

rated for pure states, i.e., $D(\psi,\phi)=\sqrt{1-F(\psi,\phi)}$ for pure states ψ and ϕ.

IV. ERROR MEASURES FOR QUANTUM PROCESSES

Our goal in this paper is to recommend a single error measure enabling researchers to compare the performance of quantum information processing experiments against the theoretical ideal. As the basis for such a recommendation, in this section we comprehensively survey possible definitions of such error measures, and do a preliminary assessment of each measure against the criteria introduced earlier in this paper.

We take three basic approaches to defining an error measure for processes. In Sec. IV A we investigate approaches based on the process matrix, $\rho_{\mathcal{E}}$. In Sec. IV B we investigate approaches based on the *average* behavior of a process. Finally, in Sec. IV C we investigate approaches based on the *worst-case* behavior of a process. In each case we investigate measures based on both the trace distance and the fidelity. We will describe connections between the various measures, and identify four measures of particular merit. The properties of these four measures will be discussed in more detail in the next section.

Nomenclature: in the following treatment we shall use the unadorned symbol Δ to mean a metric between states. Our approach is to use state-based metrics to form metrics between processes, and these will also be represented by Δ but with a subscript denoting the method used, e.g., Δ_{ave} is a process metric based on the average over input states. Where we need to specialize to a specific state metric we will use a superscript with the symbol representing that metric (*A*, *B*, *C*, and *D* from Sec. III), or use that symbol directly with a subscript for the method, e.g., $\Delta^{D}_{\text{ave}}\equiv D_{\text{ave}}$ is the *process* metric based on the average trace distance. The chief departure from these conventions will be due to the fidelity, which is not a metric. We will use the notation Δ^{F} to mean any *metric* derived from the fidelity (e.g., *A*, *B*, and *C*) and the symbol *F* with a subscript to mean a process measure based on fidelity, for example, F_{ave} is the average fidelity.

A. Error measures based on the process matrix

Suppose $\Delta(\rho,\sigma)$ is any metric on the space of quantum states. A natural approach to defining a measure Δ_{pro} of the distance between two quantum processes is

$$\Delta_{\text{pro}}(\mathcal{E},\mathcal{F})\equiv\Delta(\rho_{\mathcal{E}},\rho_{\mathcal{F}}). \tag{6}$$

Defining Δ_{pro} in this way automatically gives Δ_{pro} the metric property. Provided $\Delta(\cdot,\cdot)$ is easy to calculate, Δ_{pro} is also easy to calculate. Furthermore, since $\mathcal{E}$ can be experimentally determined using quantum process tomography, it follows that Δ_{pro} can be experimentally measured, at least in principle.

What about the other properties? The properties of stability and chaining can be obtained by making some natural extra assumptions about the state metric Δ, which we now describe. Suppose first that the metric Δ is *stable* in the sense that $\Delta(\rho\otimes\tau,\sigma\otimes\tau)=\Delta(\rho,\sigma)$. This is easily seen to be the case for the trace distance and for any fidelity-based metric, for example. The stability property for Δ_{pro} follows immediately:

$$\Delta_{\text{pro}}(I\otimes\mathcal{E},I\otimes\mathcal{F})=\Delta(\rho_I\otimes\rho_{\mathcal{E}},\rho_I\otimes\rho_{\mathcal{F}})=\Delta(\rho_{\mathcal{E}},\rho_{\mathcal{F}})=\Delta_{\text{pro}}(\mathcal{E},\mathcal{F}).$$

The chaining property can be proved, with some caveats to be described below, by assuming that $\Delta(\cdot,\cdot)$ is *contractive*, i.e., $\Delta(\mathcal{E}(\rho),\mathcal{E}(\sigma))\leqslant\Delta(\rho,\sigma)$, for trace-preserving operations $\mathcal{E}$. We have already seen that this is a natural physical assumption satisfied by the trace distance and any fidelity-based metric.

Suppose then that Δ is contractive with respect to trace-preserving operations. We claim that Δ_{pro} satisfies the chaining property,

$$\Delta_{\text{pro}}(\mathcal{E}_2\circ\mathcal{E}_1,\mathcal{F}_2\circ\mathcal{F}_1)\leqslant\Delta_{\text{pro}}(\mathcal{E}_2,\mathcal{F}_2)+\Delta_{\text{pro}}(\mathcal{E}_1,\mathcal{F}_1),$$

provided $\mathcal{F}_1$ is *doubly stochastic*, i.e., $\mathcal{F}_1$ is trace preserving and satisfies $\mathcal{F}_1(I)=I$; this assumption is used at a certain point in our proof of chaining. This may seem like a significant assumption, since physical processes such as relaxation to a finite temperature are not doubly stochastic. However, in quantum information science we are typically interested in the case when $\mathcal{F}_1$ and $\mathcal{F}_2$ are ideal unitary processes, and we are using Δ_{pro} to compare the composition of these two ideal processes to the experimentally realized process $\mathcal{E}_2\circ\mathcal{E}_1$. Since unitary processes are automatically doubly stochastic, it follows that chaining holds in this case, which is the case of usual interest.

The proof of chaining begins by applying the triangle inequality to obtain

$$\Delta_{\text{pro}}(\mathcal{E}_2\circ\mathcal{E}_1,\mathcal{F}_2\circ\mathcal{F}_1)=\Delta(\rho_{\mathcal{E}_2\circ\mathcal{E}_1},\rho_{\mathcal{F}_2\circ\mathcal{F}}) \tag{7}$$

$$\leqslant\Delta(\rho_{\mathcal{E}_2\circ\mathcal{E}_1},\rho_{\mathcal{E}_2\circ\mathcal{F}_1})+\Delta(\rho_{\mathcal{E}_2\circ\mathcal{F}_1},\rho_{\mathcal{F}_2\circ\mathcal{F}_1}). \tag{8}$$

Then note the easily verified identity $\rho_{\mathcal{E}\circ\mathcal{F}}=(\mathcal{F}^{T}\otimes\mathcal{E})(\Phi)$, where Φ is the maximally entangled state defined earlier, we define $\mathcal{F}^{T}(\rho)\equiv\sum_j F_j^{T}\rho F_j^{*}$, and F_j are the operation elements for $\mathcal{F}$ [cf. Eq. (1)]. Applying this identity to both density matrices in the second term on the right-hand side of Eq. (8) gives

$$\Delta_{\text{pro}}(\mathcal{E}_2\circ\mathcal{E}_1,\mathcal{F}_2\circ\mathcal{F}_1)\leqslant\Delta(\rho_{\mathcal{E}_2\circ\mathcal{E}_1},\rho_{\mathcal{E}_2\circ\mathcal{F}_1})+\Delta((\mathcal{F}_1^{T}\otimes\mathcal{E}_2)(\Phi),(\mathcal{F}_1^{T}\otimes\mathcal{F}_2)(\Phi)). \tag{9}$$

The double stochasticity of $\mathcal{F}_1$ implies that $\mathcal{F}_1^{T}$ is a trace-preserving quantum operation. We can therefore apply contractivity to both the first and the second terms on the right-hand side of Eq. (9), giving the desired result.

Only one property of Δ_{pro} remains in question, and that is whether or not it has a good physical interpretation. We will see in Sec. V that D_{pro} and F_{pro} can both be related in a natural way to the average probability with which a quantum computation fails or succeeds, providing a good physical interpretation for these quantities.

Although Δ_{pro} may be calculated easily in principle for both the trace distance and fidelity-based approaches, the fidelity-based measures have some substantial advantages.

The reason is that, so far as we are aware, experimentally determining D_{pro} requires doing full process tomography, which for a d-dimensional quantum system requires the estimation of d^4-d^2 observable averages. By contrast, when U is a unitary operation it turns out that the fidelity $F_{\text{pro}}(\mathcal{E},U)$ (and related error measures) can be determined based upon the estimation of at most $2d^2$ observable averages, and in particular, d^2 observable averages for qubits. This makes $F_{\text{pro}}(\mathcal{E},U)$ and related error measures substantially easier to determine experimentally than D_{pro}. The key to proving this is the observation [44]

$$F_{\text{pro}}(\mathcal{E},U)=\frac{1}{d^3}\sum_j \text{tr}[UU_j^\dagger U^\dagger \mathcal{E}(U_j)], \tag{10}$$

where the $\{U_j\}$ are a basis of unitary operators orthogonal under the Hilbert-Schmidt inner product, satisfying $\text{tr}(U_j^\dagger U_k)=d\delta_{jk}$. Up to scaling we saw an example of such a set in Sec. II, the n-qubit tensor products formed from the Pauli matrices and the identity matrix. Eq. (10) does not provide a direct way of estimating F_{pro}. But suppose we expand the U_j in terms of a set of *input states*, ρ_k: $U_j=\Sigma_k a_{jk}\rho_k$. These input states must span the entire operator space, and thus there must be d^2 of them; we will see an explicit example below for two qubits. We also expand $UU_jU^\dagger$ in terms of a set of *observables*, σ_l, $UU_jU^\dagger=\Sigma_l b_{jl}\sigma_l$. These observables must also span the entire operator space. Substitution into Eq. (10) gives

$$F_{\text{pro}}(\mathcal{E},U)=\frac{1}{d^3}\sum_{kl} M_{kl}\,\text{tr}[\sigma_l\mathcal{E}(\rho_k)], \tag{11}$$

where $M_{kl}\equiv\Sigma_j b_{jl}a_{jk}$. This equation gives a method to evaluate F_{pro}, choose a spanning set of d^2 input states ρ_k which can be prepared experimentally, and a set of observables σ_l whose averages we can reliably measure; determine the matrix $M=(M_{kl})$, whose elements depend only on known quantities (ρ_k, σ_l, and the idealized operation U), not on the unknown $\mathcal{E}$. The nonzero matrix elements in M will determine which observable averages need to be estimated for calculating $F_{\text{pro}}(\mathcal{E},U)$. In general, d^4 observable averages will need to be estimated. However, suppose we choose some fixed set of ρ_k, and then define $\sigma_l\equiv\Sigma_k a_{kl}UU_kU^\dagger$ [45]. In this case it is easily verified that Eq. (11) simplifies to

$$F_{\text{pro}}(\mathcal{E},U)=\frac{1}{d^3}\sum_k \text{tr}[\sigma_k\mathcal{E}(\rho_k)], \tag{12}$$

which only requires between d^2 and $2d^2$ measurements. The drawback is that in this method we are not free to choose the σ_l; they are determined by U and the ρ_k.

In practical situations, certain input states and measurements are easier to use than others. We envisage an experimentalist choosing the set of input states and measurements according to convenience and using the prescription above to calculate which combinations are necessary. This in general will be less than what is required to perform full process tomography. This direct method has the additional advantage of making it easier to estimate the experimental error in F_{pro}.

For example, consider an n-qubit process, U. Suppose we select the U_j to range over the n-fold tensor products of Pauli matrices (including the identity matrix). Suppose furthermore that for each qubit we select the input states from the set $\{I,I+X,I+Y,I+Z\}$ (where X, Y, Z are the usual Pauli operators), so that we choose ρ_k from the set of all possible tensor products of the single qubit input states. Now, choosing $\sigma_l\equiv\Sigma_k a_{kl}UU_kU^\dagger$, we see that the a_{kl} will always be real, and since the U_k are Hermitian then the σ_l are also Hermitian. Thus Eq. (12) tells us that we need to estimate only d^2 observable averages to evaluate F_{pro} for *any* U, much fewer than the d^4-d^2 observable averages necessary to do full process tomography on n qubits.

It is an interesting problem deserving further exploration to find the minimum number of measurements required to estimate F_{pro} when there are constraints on what input states and observables are available. For instance, it would be useful to know the optimal number for the case where we are restricted to separable inputs and product observables, i.e., inputs and observables that can be given direct local implementations.

B. Error measures based on the average case

Another natural approach for defining error measures for quantum operations is to compare output states and average over all input state, where the output states can be compared using the distance measures for states described in Sec. III. We define

$$\Delta_{\text{ave}}(\mathcal{E},\mathcal{F})\equiv\int d\psi\,\Delta(\mathcal{E}(\psi),\mathcal{F}(\psi)), \tag{13}$$

where the integral is over the uniform (Haar) measure on state space.

While this approach seems intuitively sensible, it turns out that the resulting measures satisfy few of our criteria. The only two properties these measures appear to satisfy in general, for an arbitrary state metric Δ, are the metric and chaining criteria, both of which follow immediately from the metric property of Δ.

The average-based metrics are less successful in meeting the other criteria. Even when Δ is easy to calculate, it is not obvious that the integral in Eq. (13) will have a simple form that enables easy calculation of Δ_{ave}. This, in turn, means that Δ_{ave} may not be so easy to determine experimentally. So far as we are aware, no simple expressions are known for Δ_{ave} for any of the metrics we have discussed.

It is not surprising that the physical interpretations of these metrics rely heavily on the possible interpretations of the corresponding state metrics as discussed in Sec. III. The earlier discussion of the trace distance, for example, follows on to give a meaning for D_{ave}. Suppose we are asked to distinguish between $\mathcal{E}(\psi)$ and $\mathcal{F}(\psi)$ for some ψ which is known, but has been chosen uniformly at random. On average, the optimal probability of successfully distinguishing the two processes will be $1/2+D_{\text{ave}}(\mathcal{E},\mathcal{F})/2$. Thus, $D_{\text{ave}}(\mathcal{E},\mathcal{F})$ may be interpreted as a measure of the average bias in favor of correctly distinguishing which process was applied to a state ψ. With regard to the fidelity-based metrics,

however, there does not appear to be any clear physical interpretation for Δ_{ave} because of the lack of any clear meaning for the fidelity-based metrics.

Finally, completing the checklist of criteria, our numerical analysis shows that Δ_{ave} is not stable for any of the four candidate state metrics we have investigated. Later in the paper we describe in detail a method for "stabilizing" measures which are not stable; we now briefly note the results that are obtained when this procedure is applied in the present context. The idea is to introduce an ancillary system A, and consider the quantity

$$\Delta_{\text{stab-ave}}(\mathcal{E},\mathcal{F}) \equiv \lim \Delta_{\text{ave}}(\mathcal{I}\otimes\mathcal{E},\mathcal{I}\otimes\mathcal{F}),$$

where the limit is that of large ancilla dimension. Using the well-known result that a randomly chosen state of a composite system AQ ($\dim A \gg \dim Q$) has very close to maximal entanglement [46,47], it follows that $\Delta_{\text{stab-ave}}(\mathcal{E},\mathcal{F}) = \Delta_{\text{pro}}(\mathcal{E},\mathcal{F})$, i.e., the stabilized average distance reduces to the process distance considered earlier.

There is an alternative approach, available because the fidelity-based metrics are nonlinear functions of the fidelity, which is to create a measure based on the average fidelity,

$$F_{\text{ave}}(\mathcal{E},\mathcal{F}) \equiv \int d\psi F(\mathcal{E}(\psi),\mathcal{F}(\psi)). \tag{14}$$

When $\mathcal{F}$ is a unitary operation, U, the average fidelity has a physical interpretation that is at least plausible, as the average overlap between $U|\psi\rangle$ and $\mathcal{E}(\psi)$. It was shown in Ref. [48] (see also Ref. [19]) that F_{ave} and F_{pro} are related by the equation

$$F_{\text{ave}}(\mathcal{E},U) = \frac{F_{\text{pro}}(\mathcal{E},U)d+1}{d+1}, \tag{15}$$

where d is the dimension of the quantum system, and we are restricting ourselves to the case where U is a unitary operation. This relationship makes $F_{\text{ave}}(\mathcal{E},U)$ easy to calculate [19,20] and also easy to measure experimentally, using the techniques described in the preceding section for $F_{\text{pro}}(\mathcal{E},U)$.

Although F_{ave} has several advantages (ease of calculation, ease of measurement, and a physical interpretation), the outlook for the other criteria is not so good. Not only is F_{ave} not a metric, it is not stable either, a fact that follows from Eq. (15) and the knowledge that F_{pro} is stable. The same argument shows that measures analogous to A, B, and C based on F_{ave} will also not be stable. We do not know of any stable metrics that may be derived as a function of F_{ave}, and Eq. (15) renders any such metrics equivalent in content to functions based on F_{pro} so the only reason to use them would be if they had better characteristics.

To summarize the results of this section, they show that none of the average-case error measures we have defined are particularly attractive. However, these negative results are vital because these approaches are all fairly natural solutions one might take to defining a plausible error measure. It was therefore important to consider them carefully before choosing to reject them.

C. Error measures based on the worst case

Our final approach to defining error measures is based on the worst-case distance between $\mathcal{E}(\psi)$ and $\mathcal{F}(\psi)$. We define

$$\Delta_{\max}(\mathcal{E},\mathcal{F}) \equiv \max_{\psi} \Delta(\mathcal{E}(\psi),\mathcal{F}(\psi)), \tag{16}$$

where the maximum is over all possible pure state inputs, ψ, and Δ is a metric on quantum states.

When $\Delta=\Delta^F$ is a fidelity-based metric, we see $\Delta^F_{\max}$ is a function of the *minimal fidelity*, defined by

$$F_{\min}(\mathcal{E},\mathcal{F}) \equiv \min_{\psi} F(\mathcal{E}(\psi),\mathcal{F}(\psi)). \tag{17}$$

In the definition of $\Delta_{\max}$, we maximize over all *pure* state inputs. Is this maximum the same if *all* physical inputs, including mixed states, are considered? In fact, it is fairly simple to show that this is true, and therefore that it does not matter if we optimize over pure or mixed states [49]. Suppose Δ is a *doubly convex* metric, as are all the metrics discussed in this paper (cf. Sec. III). If the maximum is achieved at some mixed state, ρ, then we have $\Delta_{\max} = \Delta(\mathcal{E}(\rho),\mathcal{F}(\rho))$. Expanding $\rho=\Sigma_j p_j\psi_j$ as a mixture of pure states, and applying double convexity we see that the maximum must also be attained at some pure state ψ_j. A similar argument holds for $F_{\min}$, based on the double concavity of the fidelity.

To assess the suitability of these measures, it is useful to first note that $D_{\max}$ has already been shown in general not to be stable [10], and similar arguments can be made to extend this to the fidelity-based measures. In Ref. [10], Aharonov *et al.* resolve this difficulty by constructing a variant of $D_{\max}$ which is stable, but which otherwise has extremely similar properties to $D_{\max}$. We now describe how this procedure can be extended to define a stable version of $\Delta_{\max}$ for an arbitrary state metric Δ, and defer for the moment discussion of the other criteria.

Suppose the original system Q on which $\mathcal{E}$ and $\mathcal{F}$ act has state space dimension d. It will be convenient to use subscripts to indicate the system on which operations act (e.g., $\mathcal{E}=\mathcal{E}_Q, \mathcal{F}=\mathcal{F}_Q$). We introduce a fictitious d-dimensional ancillary system A, acted on by the identity operation $\mathcal{I}_A$, and define the stabilized quantity [50]

$$\Delta_{\text{stab}}(\mathcal{E}_Q,\mathcal{F}_Q) \equiv \Delta_{\max}(\mathcal{I}_A\otimes\mathcal{E}_Q,\mathcal{I}_A\otimes\mathcal{F}_Q). \tag{18}$$

The proof that Δ_{stab} is stable under addition of systems is simple and has been included in Appendix A 1. In the same way, we can also define a stable form of the minimum fidelity, $F_{\text{stab}}(\mathcal{E}_Q,\mathcal{F}_Q) \equiv F_{\min}(\mathcal{I}_A\otimes\mathcal{E}_Q,\mathcal{I}_A\otimes\mathcal{F}_Q)$, with the proof of stability following similar lines. Note that the stabilized fidelity-based metrics Δ^F_{stab} are functions of F_{stab} in the obvious way (e.g. we define as usual A_{stab}, B_{stab}, and C_{stab}).

Which of the other criteria for an error measure does Δ_{stab} satisfy? It is straightforward to show that Δ_{stab} satisfies the metric and chaining criteria. Furthermore, the stabilized trace-distance D_{stab} has an appealing physical interpretation, it is the worst-case bias in the probability of being able to distinguish $(\mathcal{I}\otimes\mathcal{E})(\psi)$ from $(\mathcal{I}\otimes\mathcal{F})(\psi)$, where we allow an ancilla of arbitrary size. We defer discussion of the physical interpretation of the fidelity-based measures until the next

section, where we will see that both they and D_{stab} can be given an elegant interpretation in the context of quantum computation.

What of the remaining criteria—ease of calculation and ease of measurement? Unfortunately, no powerful general formulas for calculating Δ_{stab} are known. Reference [10] gives a general formula for the distance D_{stab} between two unitary operations, but the more interesting case of the distance between an idealized unitary operation and a noisy quantum process has not been solved, even for single-qubit operations.

The good news is that D_{stab} and F_{stab} (and thus A_{stab}, B_{stab}, and C_{stab}) are easy to calculate numerically, because they can all be reduced to *convex optimization* problems [51]. For this special class of problem, where the task is to minimize a convex function defined on a convex set, extremely efficient numerical techniques are available. Among many other nice properties, it is possible to show that a local minimum of a convex optimization problem is always a global minimum, and thus techniques such as gradient descent typically converge extremely rapidly, with no danger of finding false minima. In Appendix A 2, we prove explicitly that finding F_{stab} belongs to this class of problems, and the proof for D_{stab} follows similar lines.

We have seen that numerical calculation of D_{stab} and F_{stab} can easily be carried out, and this enables a two-step procedure for experimental measurement of either quantity, process tomography, followed by a numerical optimization. Of course, finding general formulas along the lines of $F_{\text{pro}}(\mathcal{E},U)$ or D_{pro} is still a highly desirable goal. Aside from the intrinsic benefit, finding general formulas would simplify the experimental measurement and determination of error bars for D_{stab} and F_{stab}, and perhaps obviate the need for a full process tomography, as Eq. (10) did for $F_{\text{pro}}(\mathcal{E},U)$.

V. APPLICATION TO QUANTUM COMPUTING

Can we find a good physical interpretation for any of the error measures that we have identified? In this section we will focus on interpretations that arise within the context of quantum computation and we will find that of the error measures we have discussed, four have particularly outstanding properties, D_{pro}, F_{pro}, D_{stab}, and F_{stab}. (Note that in the case of the fidelity, it will actually be more convenient to state our results in terms of the equivalent measures C_{pro} and C_{stab}.)

Assessed according to the criteria described in the introduction, these four measures have already been found to be superior to all the other measures we have studied. The additional fact that each arises naturally in the context of quantum computation strongly indicates that these four measures are the most deserving of consideration as measures of error in quantum information processing. We will return in the conclusion, Sec. VI, to the question of which of these four measures is the best possible measure of error.

There are a variety of different ways of describing quantum computations, and it turns out that each of the four error measures arises naturally in different contexts. We will discuss separately two broad divisions of quantum computation, *function computation* and *sampling computation* looking at both worst-case and average-case performance for each division.

Most algorithms on classical computers are framed as function computations. We will see that our error measures can be given particularly compelling interpretations relating to the probability of error in a function computation. However, in the context of simulating quantum systems it is often more natural to consider sampling computations, where the goal is to reproduce the statistics obtained from a measurement of the system in some specified configuration. Again, we will see that our error measures can be given good interpretations in this context, albeit somewhat more complex interpretations than for function computation.

The reason for treating the two types of computation separately is at least partially a practical one, since both types of computation arise naturally in the context of quantum computation. However, a more fundamental reason is that it does not appear to be known how to reduce sampling computation to function computation. Rather remarkably, even when there is an efficient way of *computing* a probability distribution, there does not appear to be any general way to convert that into an efficient way of *sampling* from that distribution.

A. Function computation

In function computation, the goal of the quantum computation is to *compute a function, f*, exactly or with high probability of success. More precisely, the goal is to take as input an instance, x, of the problem, and to produce a final state ρ_x of the computer that is either equal to $|f(x)\rangle$, or sufficiently close that when a measurement in the computational basis is performed, the outcome is $f(x)$ with high probability. Grover's algorithm is usually cast in this way, where we want to determine the identity of the state marked by the oracle.

Function computation in the worst case: Suppose we attempt to perform a quantum computation represented by an ideal operation $\mathcal{F}$ that acts on an input $|x\rangle$, where x represents the instance of the problem to be solved, e.g., a number to be factored [52]. This process succeeds in computing $f(x)$ with an error probability of at most p_e^{id}, where "id" indicates that this is the *ideal* worst-case error probability. Of course, in reality some nonideal operation $\mathcal{E}$ is performed. A good measure of error in the real computation is the *actual* probability p_e that the measured output of the computation is not equal to $f(x)$. In Appendix B 1, we show that

$$p_e \leqslant p_e^{\text{id}} + D_{\text{stab}}(\mathcal{E},\mathcal{F}), \tag{19}$$

$$p_e \leqslant [\sqrt{p_e^{\text{id}}} + C_{\text{stab}}(\mathcal{E},\mathcal{F})]^2. \tag{20}$$

Which of these inequalities is better depends upon the exact circumstances. For example, when $p_e^{\text{id}}=0$, we see that it depends upon whether $D_{\text{stab}}(\mathcal{E},\mathcal{F})$ is larger or smaller than $C_{\text{stab}}(\mathcal{E},\mathcal{F})^2$. With Eq. (5) in mind, it is not difficult to convince oneself that either of these possibilities may occur.

Function computation in the average case: Once again our goal is to compute a function $f(x)$ using an approximation $\mathcal{E}$ to some ideal operation $\mathcal{F}$. However, we now look at

the average-case error probability $\overline{p_e}$ that the measured output of $\mathcal{E}(|x\rangle\langle x|)$ is not equal to $f(x)$, where the average is taken with respect to a uniform distribution over instances x. Correspondingly, we introduce $\overline{p_e^{\mathrm{id}}}$, the average-case error probability for the idealized operation $\mathcal{F}$. We show that (Appendix B 2):

$$\overline{p_e} \leqslant \overline{p_e^{\mathrm{id}}} + D_{\mathrm{pro}}(\mathcal{E},\mathcal{F}). \tag{21}$$

Unfortunately, we have been unable to develop a full natural analogue of Eq. (20) based on the fidelity. However, we have proved a partial analogue for when the ideal computation succeeds with probability one ($\overline{p_e^{\mathrm{id}}}=0$). In this case,

$$\overline{p_e} \leqslant C_{\mathrm{pro}}(\mathcal{E},\mathcal{F})^2 = 1 - F(\mathcal{E},\mathcal{F}). \tag{22}$$

The proof uses very similar techniques to those used to establish Eqs. (21) and (20), and is therefore omitted.

B. Sampling computation

In sampling quantum computation, the goal is to *sample* from some ideal distribution $\{p_x(y)\} \equiv p_x$ on measurement outcomes y, with x representing input data for the problem. For instance, x might represent the coupling strengths and temperature of some spin glass model, with the goal being to sample from the thermal distribution of configurations y for that spin glass. This type of computation is particularly useful for simulating the dynamics of another quantum system.

Unlike Grover's algorithm, Shor's algorithm is usually described as a sampling computation. The goal is not to directly produce a factor or list of factors, but rather to produce a distribution over measurement outcomes. By sampling from this distribution and doing classical post-processing it is possible to extract factors of some number x. Of course, as noted in Ref. [53], it is possible to modify Shor's algorithm to be a function computation, taking an instance x and producing a list of all the factors of x.

The desired result in sampling computation is that the measurement outcomes y are distributed according to the *ideal* probabilities $p_x(y)$, for a given problem instance x. Suppose, however, that they are instead distributed according to some nonideal set of *real* probabilities $q_x(y)$. How should we compare these two distributions? There are two widely used classical measures enabling comparison of probability distributions p and q. The first is the *Kolmogorov* or l_1 distance, defined by $D(p,q) \equiv \Sigma_y |p(y)-q(y)|/2$. The second is the *Bhattacharya overlap*, defined by $F(p,q) \equiv \Sigma_y \sqrt{p(y)q(y)}$. Since these measures are in fact commutative analogues of the trace distance and fidelity, respectively, we represent them with the same symbols as their quantum analogues (D and F). As with the trace distance, the Kolmogorov distance can be given an appealing interpretation as the bias in probability when trying to distinguish the distributions p and q. No similarly simple interpretation for the Bhattacharya overlap seems to be known, although it is related to the Kolmogorov distance through inequalities analogous to Eq. (5).

The Kolmogorov distance and Bhattacharya overlap, together with the quantum error measures we have introduced, can be used to relate ideal and real probability distributions obtained as the result of a quantum computation.

Sampling computation in the worst case: Suppose we attempt to perform a quantum computation represented by an ideal operation $\mathcal{F}$ that acts on an input $|x\rangle$, where x represents the instance of the problem to be solved. The goal is to produce a final state $\mathcal{F}(|x\rangle\langle x|)$ which, when measured in the computational basis, gives rise to an ideal distribution p_x. Instead, we perform the operation $\mathcal{E}$, giving rise to a distribution q_x on measurement outcomes. In Appendix B 3 we prove that

$$\max_x D(q_x,p_x) \leqslant D_{\mathrm{stab}}(\mathcal{E},\mathcal{F}), \tag{23}$$

$$\max_x [1 - F(q_x,p_x)] \leqslant C_{\mathrm{stab}}(\mathcal{E},\mathcal{F})^2. \tag{24}$$

Just as for function computation, which of these is the better inequality depends upon the details of the situation under study.

Sampling computation in the average case: Given the same situation as for the worst case, we now assume that problem instances are chosen uniformly at random. We will therefore use the Kolmogorov distance and Bhattacharya overlap between the *joint* distributions $\{p(x,y)\} \equiv p$ and $\{q(x,y)\} \equiv q$ to measure how well $\mathcal{E}$ has approximated $\mathcal{F}$. Arguments analogous to that used in the worst case establish

$$D(q,p) \leqslant D_{\mathrm{pro}}(\mathcal{E},\mathcal{F}), \tag{25}$$

$$1 - F(q,p) \leqslant C_{\mathrm{pro}}(\mathcal{E},\mathcal{F})^2. \tag{26}$$

VI. SUMMARY, RECOMMENDATIONS, AND CONCLUSION

We have formulated a list of criteria that must be satisfied by a good measure of error in quantum information processing. These criteria provide a broad framework that can be used to assess candidate error measures, incorporating both theoretical and experimental desiderata.

We have used this framework to comprehensively survey possible approaches to the definition of an error measure, rejecting many *a priori* plausible error measures as they fail to satisfy many of our criteria. Although many of these rejected error measures are of some interest as diagnostic measures, none are suitable for use as a *primary* measure of the error in a quantum information processing task.

Four error measures were identified which have particular merit, each of which satisfies most or all of the criteria we identified. These measures are the *J distance* (Jamiolkowski process distance), the *J fidelity* (Jamiolkowski process fidelity), the *S distance* (stabilized process distance) and the *S fidelity* (stabilized process fidelity), denoted D_{pro}, F_{pro}, D_{stab}, and F_{stab}, respectively.

All four measures either are metrics (in the case of the process distances) or give rise to a variety of associated metrics (for the process fidelities). Moreover, all of the metrics can be shown to satisfy stability and chaining properties which greatly simplify the analysis of multistage quantum information processing tasks, as described in the introduc-

tion. The main differences arise in the criteria of easy calculation, measurement and sensible physical interpretation. We now briefly summarize these remaining properties for the four measures. Throughout this section, we assume that the goal in each case is to compare a quantum operation $\mathcal{E}$ to an ideal unitary operation U; the results vary somewhat when $\mathcal{E}$ is being compared to an arbitrary process $\mathcal{F}$.

(i) *J distance:* There is a straightforward formula enabling D_{pro} to be calculated directly from the process matrix, thus also allowing it to be experimentally determined using quantum process tomography. The J distance can be given an operational interpretation as a bound on the average probability of error $\overline{p_e}$ experienced during quantum computation of a function, or as a bound on the distance between the real and ideal joint distributions of the computer in a sampling computation:

$$\overline{p_e} \leqslant \overline{p_e^{\text{id}}} + D_{\text{pro}}(\mathcal{E},U), \tag{27}$$

$$D(q,p) \leqslant D_{\text{pro}}(\mathcal{E},U). \tag{28}$$

In the first expression $\overline{p_e^{\text{id}}}$ is the average probability of error in the *ideal* computation, represented by U. In the second expression, $D(q,p)$ is the Kolmogorov distance between the real joint probability distribution $\{p(x,y)\} \equiv p$ on problem instances x and measurement outcomes y and the ideal joint distribution $\{q(x,y)\} \equiv q$, for a uniform distribution on problem instances.

(ii) *J fidelity:* Once again, the J fidelity can be calculated directly from the process matrix. However, there is also a simpler formula for F_{pro}, Eq. (11), allowing easy calculation and measurement, without the need for full process tomography. This is much more straightforward than the calculation for the J distance, and is likely to simplify the determination of experimental errors. As for the J distance, the J fidelity can be given an operational interpretation related to average error probabilities,

$$\overline{p_e} \leqslant 1 - F_{\text{pro}}(\mathcal{E},U). \tag{29}$$

$$F(q,p) \geqslant F_{\text{pro}}(\mathcal{E},U). \tag{30}$$

In the first expression we are now restricted to ideal computations U which succeed perfectly, i.e., $\overline{p_e^{\text{id}}}=0$. In the second expression, $F(q,p)$ is the Bhattacharya overlap between the real and ideal joint probability distributions, p and q, again for a uniform distribution on problem instances.

(iii) *S distance:* There is no known elementary formula for D_{stab}, but we have proved that calculating the S distance is equivalent to a convex optimization problem, which can be efficiently solved numerically, given knowledge of the process. This, in turn, enables D_{stab} to be measured experimentally, by performing full quantum process tomography. The S distance can be simply interpreted as a bound on the worst-case error probability p_e for a function computation, and as a bound on the maximum distance between the real and ideal output distributions of a sampling computation,

$$p_e \leqslant p_e^{\text{id}} + D_{\text{stab}}(\mathcal{E},U), \tag{31}$$

$$\max_x D(q_x,p_x) \leqslant D_{\text{stab}}(\mathcal{E},U). \tag{32}$$

In the first expression p_e^{id} is the worst-case error probability in the ideal computation, U. In the second expression $D(q_x,p_x)$ is the Kolmogorov distance between the real and ideal output probability distributions $\{q_x(y)\} \equiv q_x$ and p_x, and we take the worst case over all problem instances x.

(iv) *S fidelity:* Once again, no elementary formula for the S fidelity is known, but we have proved that the determination of F_{stab} can be formulated as a convex optimization problem, and thus F_{stab} can be efficiently determined numerically. As a result, F_{stab} can again be determined experimentally, using process tomography. As with the S distance, F_{pro} has an operational interpretation related to worst-case error probabilities,

$$p_e \leqslant [\sqrt{p_e^{\text{id}}} + C_{\text{stab}}(\mathcal{E},U)]^2. \tag{33}$$

$$\min_x F(q_x,p_x) \geqslant F_{\text{stab}}(\mathcal{E},U). \tag{34}$$

The notation here is the same as above, with the definition $C_{\text{stab}}(\mathcal{E},U) \equiv \sqrt{1-F_{\text{stab}}(\mathcal{E},U)}$.

Which of these four error measures is the best? Our recommendation is necessarily tentative, for we do not yet have a complete understanding of the properties of these measures. In particular, the discovery of simpler formulas for calculating the measures or simpler procedures for measuring them experimentally remain possibilities which could make it necessary to reconsider their relative merits.

The fact that they all four measures obey the stability and chaining criteria means that in all cases it is only necessary to characterize the component processes in order to bound the total error in a complex quantum information processing task. This makes conceivable the idea of using these measures for assessing processes in large-scale systems.

One important difference between the measures is that the S distance and S fidelity bound worst-case error probabilities, as compared to the average-case error probabilities for which the J distance and J fidelity provide bounds. This would seem to be a significant advantage for the S distance and S fidelity, since worst-case errors are usually of more interest than the average case. On the other hand, given the linear nature of quantum mechanics, it seems likely that in low dimensions relatively tight ways may be found to use the average errors to bound the worst-case errors.

The measure which is simplest to calculate is the J fidelity, which has a simple formula, and is relatively easy to determine experimentally compared with the other measures. Unfortunately, this measure has the weakest operational interpretation of the four. As well as being only related to the average-case probability of error, our expression Eq. (29) does not hold true for function computations where the ideal case suffers an intrinsic error. For this reason we believe that the J fidelity is of particular interest for early, proof-of-principle experimental demonstrations, but that other measures with more desirable properties will eventually supersede it.

The J distance has different strengths and weaknesses than the J fidelity. On the one hand, it does allow the analysis of function computations with intrinsic errors in the ideal case. However, it requires a full process tomography to be determined experimentally, it is not as easy to calculate, and is still only related to average errors.

The S distance and S fidelity have the most attractive operational interpretations, since they relate to worst-case error probabilities. Unfortunately, they are also more difficult to determine experimentally than the J fidelity, requiring full process tomography, and no elementary formula for either is known. However, they are easy to calculate numerically, and although full process tomography is a time-consuming task, it is becoming a standard technique in quantum information experiments.

On the basis of their compelling operational interpretations, and other attractive theoretical and experimental properties, we believe that the S distance and S fidelity are the two best error measures, and should be used as the basis for comparison of real quantum information processing experiments to the theoretical ideal.

Is it possible to make a definite recommendation as regards which of these two measures to use? At the moment, we know of no convincing argument to choose one over the other. For instance, it is straightforward to find examples of different processes where either the S distance or the S fidelity give the better bound in Eqs. (31) and (33). Further work on the relative merits of these measures is required before a definitive choice can be made.

As a consequence, at the present time we believe that *both* measures should be reported in experiments. Note that determining two measures rather than one imposes little additional burden on experimentalists, since determining either measure requires (at present) process tomography to be performed, and once process tomography has been performed it is straightforward to numerically calculate both measures.

Much work remains to be done. Tasks of obvious importance include (a) obtaining closed-form formulas and simple experimental measurement procedures for the S distance and S fidelity, (b) finding procedures which can be used to calculate experimental error bars for the S distance and S fidelity, (c) expressing the threshold condition for fault-tolerant quantum computation and communication using the error measures we have identified, and (d) extending our work so that it applies to quantum operations which are not trace preserving, such as those which arise naturally in certain optical proposals for quantum computation [54,55], where measurements and post-selection are critical elements.

Broadening the scope, it would also be useful to develop additional diagnostic measures, which could be used experimentally to understand and improve specific aspects of a process's operation, while not being suitable as general-purpose measures of how well a process has been performed. An example of such a measure is the *process purity*, $\mathrm{tr}(\rho_\varepsilon^2)$, which can be regarded as a measure of the extent to which a quantum operation $\mathcal{E}$ maintains the purity of the quantum state. Although this measure is easily seen to be deficient in terms of the criteria developed in the introduction, and thus is not suitable as a general-purpose measure, it may be useful as a diagnostic measure that provides information about one specific aspect of $\mathcal{E}$'s performance.

ACKNOWLEDGMENTS

One of the authors (A.G.) acknowledges support from the New Zealand Foundation for Research, Science and Technology under Grant No. UQSL0001. One of the authors (M.A.N.) thanks Carl Caves, who has repeatedly emphasized the significance of obtaining suitable criteria for quantum information processing.

APPENDIX A: WORST-CASE PROOFS

1. Proof of worst-case stabilization

Let $\mathcal{E}_Q$ and $\mathcal{F}_Q$ be trace-preserving quantum operations acting on a d-dimensional system Q. We will show, following Ref. [10], that $\Delta_{\mathrm{stab}}(\mathcal{E}_Q,\mathcal{F}_Q)$ is stable under the addition of an arbitrary d'-dimensional system Q', i.e, $\Delta_{\mathrm{stab}}(\mathcal{E}_Q,\mathcal{F}_Q) = \Delta_{\mathrm{stab}}(\mathcal{I}_{Q'}\otimes\mathcal{E}_Q,\mathcal{I}_{Q'}\otimes\mathcal{F}_Q)$.

To see this, recall the definition of $\Delta_{\mathrm{stab}}(\mathcal{E}_Q,\mathcal{F}_Q)$. We introduce a fictitious d-dimensional ancillary system A, acted upon by the identity operation $\mathcal{I}_A$. Then by definition $\Delta_{\mathrm{stab}}(\mathcal{E}_Q,\mathcal{F}_Q)\equiv\Delta_{\max}(\mathcal{I}_A\otimes\mathcal{E}_Q,\mathcal{I}_A\otimes\mathcal{F}_Q)$.

By definition of Δ_{stab} we see that $\Delta_{\mathrm{stab}}(\mathcal{I}_{Q'}\otimes\mathcal{E}_Q,\mathcal{I}_{Q'}\otimes\mathcal{F}_Q)$ is equal to $\Delta_{\max}(\mathcal{I}_B\otimes\mathcal{I}_{Q'}\otimes\mathcal{E}_Q,\mathcal{I}_B\otimes\mathcal{I}_{Q'}\otimes\mathcal{F}_Q)$, where $\mathcal{I}_B$ acts as the identity on a $d\times d'$-dimensional ancilla B. Thus, to prove stability it suffices to show that the quantity $\Delta_{\max}(\mathcal{I}_S\otimes\mathcal{E}_Q,\mathcal{I}_S\otimes\mathcal{F}_Q)$ is independent of the dimension of the system S that $\mathcal{I}_S$ acts on, provided S is at least d dimensional.

To see this independence, let ψ be a state achieving the maximum in $\Delta_{\max}(\mathcal{I}_S\otimes\mathcal{E}_Q,\mathcal{I}_S\otimes\mathcal{F}_Q)$, with a Schmidt decomposition $\psi=\Sigma_j\psi_j|e_j\rangle|f_j\rangle$, where $|e_j\rangle$ are orthonormal states of S, and $|f_j\rangle$ is an orthonormal basis set for Q. Since Q is d dimensional, the state ψ has at most d Schmidt coefficients, and so we can restrict our attention to that d-dimensional subspace of S spanned by the states $|e_j\rangle$ with nonzero Schmidt coefficients. We see that the maximum can be obtained working only in this subspace, concluding the proof.

2. Proof of convex optimization property for F_{stab}

Our goal is to show that the problem of computing F_{stab} can be reduced to the minimization of a convex function defined on a convex set. To show this we introduce a function, denoted $F(\rho_Q,\mathcal{E}_Q,\mathcal{F}_Q)$, where subscripts indicate the system on which the variable is defined. The value of $F(\rho_Q,\mathcal{E}_Q,\mathcal{F}_Q)$ is defined to be the state fidelity $F\big((\mathcal{I}_A\otimes\mathcal{E}_Q)(\psi),(\mathcal{I}_A\otimes\mathcal{F}_Q)(\psi)\big)$, where A is an ancilla of at least the same dimension as Q, and ψ is any purification of ρ_Q to AQ. It is easily verified that this definition is independent of which purification ψ of ρ_Q is used.

From this definition, it can be seen that the problem of computing $F_{\mathrm{stab}}(\mathcal{E}_Q,\mathcal{F}_Q)$ is equivalent to minimizing $F(\rho_Q,\mathcal{E}_Q,\mathcal{F}_Q)$ over all density matrices ρ_Q of system Q. Therefore, to prove that finding F_{stab} is a convex optimization problem, we simply need to show that $F(\rho_Q,\mathcal{E}_Q,\mathcal{F}_Q)$ is a convex function of ρ_Q, which takes values in a convex set.

To do this, let p_j be probabilities, and let ρ_Q^j be corresponding states of the system Q, with purifications ψ_j to a system AQ. It is helpful to introduce another ancillary system

A' with an orthonormal basis $|j\rangle$ in one-to-one correspondence with the index on the states ρ_Q^j, and we define a state $|\psi\rangle \equiv \Sigma_j \sqrt{p_j}|j\rangle|\psi_j\rangle$ of the joint system $A'AQ$. By observing that $|\psi\rangle$ is a purification of $\Sigma_j p_j \rho_Q^j$, we see that

$$F\left(\sum_j p_j \rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q\right) = F((\mathcal{I}_{A'A} \otimes \mathcal{E}_Q)(\psi), (\mathcal{I}_{A'A} \otimes \mathcal{F}_Q)(\psi)). \tag{A1}$$

We then apply the monotonicity of the fidelity (cf. Sec. III) under decoherence in the $|j\rangle$ basis, giving

$$F\left(\sum_j p_j \rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q\right) \leqslant F\left(\sum_j p_j |j\rangle\langle j| \otimes (\mathcal{I}_A \otimes \mathcal{E}_Q)(\psi_j), \sum_j p_j |j\rangle\langle j| \otimes (\mathcal{I}_A \otimes \mathcal{F}_Q)(\psi_j)\right). \tag{A2}$$

Finally, applying some elementary algebra to simplify the right-hand side, we obtain

$$F\left(\sum_j p_j \rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q\right) \leqslant \sum_j p_j F(\rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q), \tag{A3}$$

which implies that $F(\rho_Q, \mathcal{E}_Q, \mathcal{F}_Q)$ is convex in ρ_Q, as desired.

A similar construction shows that the computation of D_{stab} is equivalent to the maximization of a concave function over a convex set, and thus is also a convex optimization problem, with concomitant numerical benefits. The construction is sufficiently similar that we omit the details.

APPENDIX B: APPLICATION TO QUANTUM COMPUTING

1. Function computation in the worst case

Suppose $\mathcal{E}$ and $\mathcal{F}$ are real and ideal quantum operations, respectively, that act on an input $|x\rangle$, where x represents a problem instance. $\mathcal{E}$ succeeds in computing the desired function $f(x)$ with an error probability of at most p_e, whereas $\mathcal{F}$ succeeds with an (ideal) error probability of at most p_e^{id}.

We wish to show

$$p_e \leqslant p_e^{\text{id}} + D_{\text{stab}}(\mathcal{E}, \mathcal{F}), \tag{B1}$$

$$p_e \leqslant [\sqrt{p_e^{\text{id}}} + C_{\text{stab}}(\mathcal{E}, \mathcal{F})]^2. \tag{B2}$$

To prove the first inequality (B1), we introduce a quantum operation $\mathcal{M}$ representing the process of measurement, $\mathcal{M}(\rho) = \Sigma_y |y\rangle\langle y|\rho|y\rangle\langle y|$, where the sum is over all possible measurement outcomes y. Now observe that

$$p_e = D((\mathcal{M} \circ \mathcal{E})(|x\rangle\langle x|), |f(x)\rangle\langle f(x)|) \tag{B3}$$

$$\leqslant D((\mathcal{M} \circ \mathcal{E})(|x\rangle\langle x|), (\mathcal{M} \circ \mathcal{F})(|x\rangle\langle x|)) + D((\mathcal{M} \circ \mathcal{F})(|x\rangle\langle x|), |f(x)\rangle\langle f(x)|) \tag{B4}$$

$$\leqslant D(\mathcal{E}(|x\rangle\langle x|), \mathcal{F}(|x\rangle\langle x|)) + p_e^{\text{id}}, \tag{B5}$$

where we used simple algebra in the first line, the triangle inequality in the second line, and contractivity of trace distance and some simple algebra in the third line. The desired result, Eq. (B1), now follows from the definition of D_{stab}.

To prove the second inequality, Eq. (B2), note that

$$p_e = 1 - F(\mathcal{E}(|x\rangle\langle x|), |f(x)\rangle\langle f(x)|) \tag{B6}$$

$$= C(\mathcal{E}(|x\rangle\langle x|), |f(x)\rangle\langle f(x)|)^2 \tag{B7}$$

$$\leqslant [C(\mathcal{E}(|x\rangle\langle x|), \mathcal{F}(|x\rangle\langle x|)) + C(\mathcal{F}(|x\rangle\langle x|), |f(x)\rangle\langle f(x)|)]^2, \tag{B8}$$

where the first line follows from the definition of p_e and the state fidelity, the second line follows from the definition of the metric $C(\cdot, \cdot)$, and the third line follows from the triangle inequality for $C(\cdot, \cdot)$. The proof of Eq. (B2) is completed by noting that $C(\mathcal{E}(|x\rangle\langle x|), \mathcal{F}(|x\rangle\langle x|)) \leqslant C_{\text{stab}}(\mathcal{E}, \mathcal{F})$ and $C(\mathcal{F}(|x\rangle\langle x|), |f(x)\rangle\langle f(x)|) \leqslant \sqrt{p_e^{\text{id}}}$.

2. Function computation in the average case

As in the worst case, $\mathcal{E}$ and $\mathcal{F}$ are real and ideal quantum operations that act on an input $|x\rangle$ to compute a desired function $f(x)$. $\mathcal{E}$ succeeds with an average error probability $\overline{p_e}$, whereas $\mathcal{F}$ succeeds with an average error probability $\overline{p_e^{\text{id}}}$.

The first steps in the proof of Eq. (21) are directly analogous to the proof of Eq. (19), resulting in the inequality

$$\overline{p_e} \leqslant \overline{p_e^{\text{id}}} + \frac{1}{d} \sum_x D(\mathcal{E}(|x\rangle\langle x|), \mathcal{F}(|x\rangle\langle x|)), \tag{B9}$$

where d is the total number of possible inputs x. Recall that

$$D_{\text{pro}}(\mathcal{E}, \mathcal{F}) = D((\mathcal{I} \otimes \mathcal{E})(\Phi), (\mathcal{I} \otimes \mathcal{F})(\Phi)), \tag{B10}$$

where $\mathcal{I}$ acts on an ancilla which is a copy of the system $\mathcal{E}$ and $\mathcal{F}$ act on, and $|\Phi\rangle = \Sigma_x |x\rangle|x\rangle / \sqrt{d}$ is a maximally entangled state of the two systems. Now let $\mathcal{M}$ be a quantum operation representing measurement on the ancilla system, defined similarly to the definition of $\mathcal{M}$ just above. By contractivity of the trace distance,

$$D_{\text{pro}}(\mathcal{E}, \mathcal{F}) \geqslant D((\mathcal{M} \otimes \mathcal{E})(\Phi), (\mathcal{M} \otimes \mathcal{F})(\Phi)). \tag{B11}$$

Elementary algebra gives

$$D((\mathcal{M} \otimes \mathcal{E})(\Phi), (\mathcal{M} \otimes \mathcal{F})(\Phi)) = \frac{1}{d} \sum_x D(\mathcal{E}(|x\rangle\langle x|), \mathcal{F}(|x\rangle\langle x|)). \tag{B12}$$

Combining these results, we obtain Eq. (21).

As already remarked we have not found a natural average-case analogue of Eq. (20). However, if $\overline{p_e^{\text{id}}} = 0$, i.e., our computation succeeds with probability one, then it is possible to prove an average-case analogue. The result is

$$\overline{p_e} \leqslant C_{\text{pro}}(\mathcal{E}, \mathcal{F})^2 = 1 - F(\mathcal{E}, \mathcal{F}). \tag{B13}$$

The proof uses very similar techniques to those used to establish Eqs. (21) and (20), and is therefore omitted.

3. Sampling computation in the worst case

The quantum operation $\mathcal{E}$ is an imperfect attempt to reproduce the statistics of the ideal operation $\mathcal{F}$ which acts on an

input $|x\rangle$. Measured in the computational basis, $\mathcal{F}$ gives rise to a distribution $\{p_x(y)\} \equiv p_x$, whereas $\mathcal{E}$ gives a distribution $\{q_x(y)\} \equiv q_x$.

The inequalities Eqs. (23) and (24) that we want to prove may be stated as follows:

$$\max_x D(q_x, p_x) \leqslant D_{\text{stab}}(\mathcal{E}, \mathcal{F}), \tag{B14}$$

$$\min_x F(q_x, p_x) \geqslant F_{\text{stab}}(\mathcal{E}, \mathcal{F}). \tag{B15}$$

To prove the first inequality (B14), let $\mathcal{M}$ again be a quantum operation representing measurement in the computational basis. Note that for all x,

$$D(q_x, p_x) = D((\mathcal{M} \circ \mathcal{E})(|x\rangle\langle x|), (\mathcal{M} \circ \mathcal{F})(|x\rangle\langle x|)) \tag{B16}$$

$$\leqslant D(\mathcal{E}(|x\rangle\langle x|), \mathcal{F}(|x\rangle\langle x|)) \tag{B17}$$

$$\leqslant D_{\text{stab}}(\mathcal{E}, \mathcal{F}), \tag{B18}$$

where we used simple algebra in the first line, contractivity in the second line, and the definition of D_{stab} in the third line. An analogous argument can be used to establish the second inequality (B15).

[1] K. R. W. Jones, Ann. Phys. (N.Y.) **207**, 140 (1991).
[2] U. Leonhardt, Phys. Rev. A **53**, 2998 (1996).
[3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[4] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).
[5] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).
[6] Our use of this term was suggested by a talk of Caves [7]; see also Ref. [12], which discusses possible gold standards for quantum teleportation.
[7] C. M. Caves, Quantum information processing: Is it always quantum, Talk at the Kavli Institute for Theoretical Physics, UC Santa Barbara, 2001, viewgraphs, video, and audio available online at http://online.kitp.ucsb.edu/online/qinfo01/caves/
[8] See Chap. 10 of Ref. [3], and references therein.
[9] We prefer the physically motivated term quantum operation over two other commonly used and equivalent terms, *completely positive map* and *superoperator*.
[10] D. Aharonov, A. Kitaev, and N. Nisan, STOC 1998 e-print quant-ph/9806029.
[11] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
[12] C. M. Caves and K. Wodkiewicz, Phys. Rev. Lett. **93**, 040506 (2004).
[13] S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and P. van Loock, Phys. Rev. A **64**, 022321 (2001).
[14] T. Rudolph and B. C. Sanders, Phys. Rev. Lett. **87**, 077903 (2001).
[15] F. Grosshans and P. Grangier, Phys. Rev. A **64**, 010301(R) (2001).
[16] T. C. Ralph, R. E. S. Polkinghorne, and P. K. Lam, J. Opt. B: Quantum Semiclassical Opt. **1**, 483 (1999).
[17] R. Schack and C. M. Caves, Phys. Rev. A **60**, 4354 (1999).
[18] B. M. Terhal and G. Burkard, e-print quant-ph/0402104.
[19] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).
[20] M. D. Bowdrey, D. K. L. Oi, A. J. Short, K. Banaszek, and J. A. Jones, Phys. Lett. A **294**, 258 (2002).
[21] M. Raginsky, Phys. Lett. A **290**, 11 (2001).
[22] A. M. Childs, I. L. Chuang, and D. W. Leung, Phys. Rev. A **64**, 012314 (2001).
[23] A. M. Childs, J. Preskill, and J. Renes, J. Mod. Opt. **47**, 155 (2000).
[24] E. Bernstein and U. Vazirani, SIAM J. Comput. **26**, 1411 (1997).
[25] B. W. Schumacher, Phys. Rev. A **54**, 2614 (1996).
[26] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[27] The operator inequality $A \leqslant B$ means that $B-A$ is a positive matrix. Thus $A \leqslant I$ means that A is a Hermitian matrix whose eigenvalues are at most 1.
[28] Post-selection is usually associated with some specific measurement outcome being obtained.
[29] An n-qubit example of an orthogonal operator basis is the set of all possible tensor products formed from the Pauli matrices and the identity. These operators may be normalized by multiplying by an appropriate constant, giving an orthonormal operator basis.
[30] The condition $\Sigma_j E_j^\dagger E_j \leqslant I$ for the operation elements translates to $\Sigma \chi_{nm} A_m^\dagger A_n \leqslant I$ for the process matrix.
[31] A. Jamiolkowski, Rep. Math. Phys. **3**, 275 (1972).
[32] Note, however, that not all density matrices may be realized as states $\rho_{\mathcal{E}}$. It turns out that the class $\mathcal{E}$ of quantum states that may arise in this way from a trace-preserving operation is simply all those $\rho_{\mathcal{E}}$ whose reduced density matrix on the copy of the original system is the completely mixed state, I/d [48].
[33] A review of the properties of the trace distance and fidelity may be found in Chap. 9 of Ref. [3].
[34] See Eq. (9.22) in Ref. [3].
[35] M. B. Ruskai, Rev. Math. Phys. **6**, 1147 (1994).
[36] It is noteworthy that using the square root of the definition in Eq. (4) results in "better" values for the fidelity, and this may account for some of the attractiveness of that quantity. Needless to say, this is not a compelling reason for using it.
[37] Strictly, this statement is only true in projective state space, in which states that are the same up to overall phase are regarded as identical.
[38] To prove the triangle inequality for $C(\rho, \sigma)$, we note that by Uhlmann's theorem (see Ref. [39] or Theorem 9.4 on p. 410 of Ref. [3]) given any three states ρ, σ, and τ there exist corresponding purifications $|r\rangle$, $|s\rangle$, and $|t\rangle$ such that $F(\rho, \sigma) = |\langle r|s\rangle|^2$, $F(\rho, \tau) = |\langle r|t\rangle|^2$, and $F(\sigma, \tau) = |\langle s|t\rangle|^2$. Let θ_{rs} be an angle in the range 0 to $\pi/2$ so that $\cos(\theta_{rs}) = |\langle r|s\rangle|$.

Define θ_{rt} and θ_{st} similarly. With these definitions it is easy to check that $C(\rho,\sigma)=\sin(\theta_{rs})$, with similar relations holding for the other pairs of density matrices, and so the triangle inequality reduces to proving $\sin(\theta_{rt})\leqslant\sin(\theta_{rs})+\sin(\theta_{st})$. But elementary geometry implies that $\theta_{rt}\leqslant\theta_{rs}+\theta_{st}$, and simple algebra can be used to deduce the triangle inequality from that.

[39] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[40] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996).

[41] J. L. Dodd and M. A. Nielsen, Phys. Rev. A **66**, 044301 (2002).

[42] C. A. Fuchs, Ph.D thesis, The University of New Mexico, Albuquerque, NM, 1996.

[43] C. A. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[44] This observation seems to be folklore. We do not know where it originated, but a proof is provided in Ref. [19].

[45] Note that the σ_l defined in this way are not, in general, Hermitian, and thus may not be observables. However, we can easily split them into the form $\sigma_l=A_l+iB_l$, where A_l and B_l are Hermitian, and thus measuring $\mathrm{tr}(\sigma_l\rho)$ can be achieved simply by measuring two separate observables.

[46] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).

[47] S. Sen, Phys. Rev. Lett. **77**, 1 (1996).

[48] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[49] Obviously, taking the maximum over pure states will be numerically easier as it will involves only $2d-2$ real parameters for a d-dimensional state space, as opposed to d^2-1 for a normalized mixed state.

[50] When Δ is the trace distance, Aharonov *et al.* [10] refer to this as the diamond norm.

[51] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).

[52] One can consider an alternate model in which the starting state of the computer is some standard state, $|s\rangle$, and the operation to be performed $\mathcal{F}_x$ depends on the problem instance. Exactly analogous results to those we will describe here can be proved in this alternate context.

[53] N. Gershenfeld and I. L. Chuang, Science **275**, 350 (1997).

[54] E. Knill, R. Laflamme, and G. J. Milburn, Nature (London) **409**, 46 (2001).

[55] M. A. Nielsen, e-print quant-ph/0402005.