Robustness of quantum gates in the presence of noise

Aram W. Harrow[1,2,*] and Michael A. Nielsen[2,†]
[1]*MIT Physics, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[2]*School of Physical Sciences, University of Queensland, Queensland 4072, Australia*
(Received 24 January 2003; published 11 July 2003)

We define several quantitative measures of the *robustness* of a quantum gate against noise. Exact analytic expressions for the robustness against depolarizing noise are obtained for all bipartite unitary quantum gates, and it is found that the controlled-NOT gate is the most robust two-qubit quantum gate, in the sense that it is the quantum gate which can tolerate the *most* depolarizing noise and still generate entanglement. Our results enable us to place several analytic upper bounds on the value of the threshold for quantum computation, with the best bound in the most pessimistic error model being $p_{\text{th}} \leqslant 0.5$.

DOI: 10.1103/PhysRevA.68.012308 PACS number(s): 03.67.Lx, 03.65.Ud

I. INTRODUCTION

An ideal quantum computer [1] is usually described as a sequence of unitary quantum gates applied to the qubits making up the computer. A typical universal set of quantum gates is the controlled-NOT (CNOT) gate, and single-qubit unitary operations [2]. A crucial element in a universal gate set is that it be capable of generating *entanglement* between the qubits making up the computer.

In the real world, quantum gates suffer from noise [3], which can inhibit the creation of entanglement. This problem led to the development of fault-tolerant methods for quantum computation (see Ref. [1]) based on quantum error-correcting codes [4,5]. One of the outstanding achievements of work on fault tolerance is the *threshold theorem* for quantum computation [6–10]. The threshold theorem states that, under reasonable physical assumptions about noise in the computer, it is possible to correct for the effects of that noise, provided the strength of the noise is below some constant threshold, p_{th}. (Roughly speaking, p_{th} can be thought of as the maximal probability of error during a single quantum gate that can be corrected using the methods of fault tolerance.) The exact value of the threshold depends on what assumptions are made about the noise in the quantum computer, and estimates of the value of the threshold, therefore, vary quite a bit. Typical current estimates place it in the range 10^{-4}–10^{-6}.

Motivated by the practical problem of noise, and the theory of fault-tolerant quantum computation, in this paper we consider the problem of quantifying how robust a quantum gate is to the effects of noise. More precisely, for a given gate U we attempt to quantify how much noise the gate can tolerate while preserving the ability to generate entanglement. Since, in a sense we make precise below, entanglement generation is necessary for quantum computation to be possible, even if the methods of fault-tolerant computation are used, this program allows us to determine *upper bounds* on the value of the threshold.

Our work is different from most of the other works on estimating thresholds, which usually aims to determine *lower bounds*. The interest in lower bounds stems from their more immediate practical interest: if we know that $p_{\text{th}} > 10^{-6}$, for example, then that gives experimentalists a target to shoot for in pursuit of a working quantum computer. Nonetheless, as emphasized in Ref. [11], from a fundamental point of view it would be extremely interesting to have exact values for the threshold, and this requires techniques for obtaining upper bounds.

Our work is based upon the results of Vidal and Tarrach [12], who investigated the *robustness* of entangled quantum states, that is, how much noise can be added to a quantum state before it becomes unentangled, i.e., separable. Our work also naturally extends and complements the work of Aharonov and Ben-Or [11], who, to our knowledge, have done the only prior work obtaining upper bounds on the value of the threshold.

Another interesting context in which our measures of gate robustness may be placed is the program of defining ''dynamic strength measures'' for quantum dynamical operations [13]. Dynamic strength measures quantify the intrinsic power or strength of a quantum dynamical operation as a physical resource, much as an entanglement measure quantifies the entanglement in a quantum state. Reference [13] developed a framework for the analysis of dynamic strength measures, and we will see that gate robustness can be regarded as a measure of dynamic strength, and analyzed within this framework.

The structure of the paper is as follows. Section II reviews background material on the Schmidt decomposition for operators. This decomposition is central to our later work on the robustness of quantum gates. Section III reviews the notion of *separable* quantum gates, which may be defined as the class of gates that cannot generate entanglement in a quantum computer. Furthermore, this section proves that a quantum circuit containing only separable gates can be efficiently simulated on a classical computer. Section IV reviews Vidal and Tarrach's work on the robustness of quantum states. This section also introduces an alternative measure of the robustness of quantum states useful in our later work on

*Electronic address: aram@mit.edu;
URL: http://web.mit.edu/aram/
†Electronic address: nielsen@physics.uq.edu.au;
URL: http://www.qinfo.org/people/nielsen/

 ©2003 The American Physical Society

gate robustness, and proves some elementary properties of the new measure. Section V gives our definitions and results on the robustness of quantum gates, and relates the results to the theory of fault-tolerant quantum computation. Section VI concludes.

II. THE OPERATOR-SCHMIDT DECOMPOSITION

The *operator-Schmidt* decomposition is an operator analog of the well-known Schmidt decomposition for pure quantum states [1]. The present treatment of the operator-Schmidt decomposition is based on the discussion in Refs. [13,14], with the addition of a result on the continuity of the Schmidt coefficients of a unitary operator.

We begin by introducing the Hilbert-Schmidt inner product on $d\times d$ operators, $(Q,P)\equiv \mathrm{tr}(Q^\dagger P)$, for any operators Q and P. We define an orthonormal operator basis to be a set $\{Q_j\}$ which satisfies the condition $(Q_j,Q_k)=\mathrm{tr}(Q_j^\dagger Q_k)=\delta_{jk}$. For example, an orthonormal basis for the space of single-qubit operators is the set $\{I/\sqrt{2},X/\sqrt{2},Y/\sqrt{2},Z/\sqrt{2}\}$, where X, Y, and Z are the Pauli σ operators, and I is the identity.

The operator-Schmidt decomposition states that any operator Q acting on systems A and B may be written as [14]

$$Q=\sum_l q_l A_l\otimes B_l, \tag{1}$$

where $q_l\geqslant 0$, and A_l and B_l are orthonormal operator bases for A and B, respectively. A constructive proof of the operator-Schmidt decomposition may be found in Ref. [14].

To better understand the coefficients q_l in the operator-Schmidt decomposition, imagine that associated with each system, A and B, there are *reference systems* R_A and R_B, with the same state space dimensionalities d_A and d_B as A and B. Let

$$|\alpha\rangle\equiv\frac{\sum_j |j_{R_A}j_A\rangle}{\sqrt{d_A}} \quad\text{and}\quad |\beta\rangle\equiv\frac{\sum_j |j_B j_{R_B}\rangle}{\sqrt{d_B}} \tag{2}$$

denote normalized, maximally entangled states of R_AA and BR_B, respectively. Now let $\mathcal{E}$ be a general quantum operation.[1] We define $\rho(\mathcal{E})$ to be the density operator resulting when $\mathcal{E}$ acts on $|\alpha\rangle|\beta\rangle$. We write this out explicitly, with subscripts to make it clear which operations are acting on which systems:

$$\rho(\mathcal{E})\equiv(\mathcal{I}_{R_A}\otimes\mathcal{E}_{AB}\otimes\mathcal{I}_{R_B})\circ(|\alpha\rangle\langle\alpha|\otimes|\beta\rangle\langle\beta|), \tag{3}$$

where $\mathcal{I}_S$ denotes the identity quantum operation on a system S. In the special case when $\mathcal{E}$ represents a unitary operation U on AB, we define $\psi(U)$ to be the quantum state obtained when U acts on $|\alpha\rangle|\beta\rangle$, and let $\rho(U)$ be the corresponding density operator. Note that we will interchange notations such as $\psi(U)$ and $|\psi(U)\rangle$, depending on which is more convenient in a particular context.

The Schmidt coefficients of $\psi(U)$ are closely connected to the operator-Schmidt coefficients of U, which we denote u_j. Letting $U=\Sigma_j u_j A_j\otimes B_j$ be an operator-Schmidt decomposition, we see that

$$\psi(U)=(I_{R_A}\otimes U\otimes I_{R_B})|\alpha\rangle|\beta\rangle \tag{4}$$

$$=\sum_j u_j(I_{R_A}\otimes A_j)|\alpha\rangle(B_j\otimes I_{R_B})|\beta\rangle. \tag{5}$$

Direct calculation shows that $\sqrt{d_A}(I_{R_A}\otimes A_j)|\alpha\rangle$ and $\sqrt{d_B}(B_j\otimes I_{R_B})|\beta\rangle$ form orthonormal bases for R_AA and BR_B, respectively. Thus, the quantum state $\psi(U)$ has Schmidt coefficients $u_j/\sqrt{d_Ad_B}$ equal, up to the factor $1/\sqrt{d_Ad_B}$, to the Schmidt coefficients of U.

The following proposition shows that the Schmidt coefficients of U are continuous functions of U. In the statement of the proposition, $\|M\|=\max_{\||\psi\rangle\|=1}\|M|\psi\rangle\|$ denotes the usual operator norm.

Proposition 1. Let U and V be operators on AB, with respective Schmidt coefficients u_j and v_j, ordered into decreasing order, $u_1\geqslant u_2\geqslant\cdots$, and $v_1\geqslant v_2\geqslant\cdots$. Then

$$2\left(1-\frac{\sum_j u_jv_j}{d_Ad_B}\right)\leqslant\|U-V\|^2. \tag{6}$$

To understand why Eq. (6) can be interpreted as a statement about continuity requires a little thought. Note that $\mathrm{tr}(U^\dagger U)=\mathrm{tr}(V^\dagger V)=d_Ad_B$, and thus $\Sigma_j u_j^2=\Sigma_j v_j^2=d_Ad_B$. It follows that we can think of u_j^2/d_Ad_B and v_j^2/d_Ad_B as probability distributions. With this interpretation, the quantity $\Sigma_j u_jv_j/d_Ad_B$ is just the fidelity of these two probability distributions, and it follows from Eq. (6) that if $U\approx V$ then $u_j\approx v_j$ for all j.

Proof. The key is to observe that the norm $\|\cdot\|$ is *stable* when extended trivially to an ancilla system, i.e., $\|M\|=\|M\otimes I\|$. Using this observation, we have

$$\|U-V\|=\|I_{R_A}\otimes(U-V)\otimes I_{R_B}\| \tag{7}$$

$$\geqslant\|[I_{R_A}\otimes(U-V)\otimes I_{R_B}]|\alpha\rangle|\beta\rangle\| \tag{8}$$

$$=\|\psi(U)-\psi(V)\|. \tag{9}$$

Squaring both sides of the inequality, and interchanging the roles of the two sides, we obtain

$$\|\psi(U)\|^2+\|\psi(V)\|^2-2\mathrm{Re}[\langle\psi(U)|\psi(V)\rangle]\leqslant\|U-V\|^2. \tag{10}$$

Since $\|\psi(U)\|^2=\|\psi(V)\|^2=1$, this implies

[1]Quantum operations are sometimes known as *completely positive maps*. We use the more physically oriented terminology, since it is physical applications we have in mind. Note that we use "quantum gate" and "quantum operation" interchangeably, depending on whether the context is quantum computation or more general. A review of the theory of quantum operations may be found in Ref. [1].

$$2[1-|\langle\psi(U)|\psi(V)\rangle|]\leqslant\|U-V\|^2. \quad (11)$$

Since $\psi(U)$ and $\psi(V)$ have Schmidt coefficients $u_j/\sqrt{d_A d_B}$ and $v_j/\sqrt{d_A d_B}$, respectively, it follows from the results of Refs. [15,16] that $|\langle\psi(U)|\psi(V)\rangle|\leqslant\Sigma_j u_j v_j/d_A d_B$. Combining this inequality with Eq. (11) gives the desired result. ■

III. SEPARABLE AND SEPARABILITY-PRESERVING QUANTUM GATES

We now formally introduce the notion of separable quantum gates, and study their basic properties, in Sec. III A. Section III B states and proves a theorem showing that quantum circuits built entirely out of separable quantum gates can be efficiently simulated on a classical computer. Finally, Sec. III C notes that the classical simulation theorem of the preceding section can be extended to a somewhat larger class of gates, the ''separability-preserving'' gates, and considers some of the implications of this fact.

A. Definition and basic properties

Suppose $\mathcal{E}$ is a quantum operation acting on a composite quantum system with two components labeled A and B. $\mathcal{E}$ is said to be *separable* if it can be given an operator-sum representation of the form

$$\mathcal{E}(\rho)=\sum_j (A_j\otimes B_j)\rho(A_j^\dagger\otimes B_j^\dagger). \quad (12)$$

Separable quantum operations were independently introduced in Refs. [17,18], where it was speculated that trace-preserving separable quantum operations might correspond to the class of quantum operations that can be implemented on a bipartite system using local operations and classical communication. This speculation was false [19]. However, a related conjecture is true, namely, that trace-preserving separable quantum operations correspond to the class of trace-preserving quantum operations which cannot be used to generate quantum entanglement. This follows from an elegant characterization theorem of Cirac *et al.* [20] linking separability of a quantum operation $\mathcal{E}$ to separability of the quantum state $\rho(\mathcal{E})$ introduced in Eq. (3).

Theorem 1 (operation-separability theorem [20]). A trace-preserving quantum operation $\mathcal{E}$ is separable if and only if $\rho(\mathcal{E})$ is a separable quantum state, that is, $\rho(\mathcal{E})$ can be written in the form

$$\rho(\mathcal{E})=\sum_j p_j\rho_j^{R_AA}\otimes\rho_j^{BR_B}, \quad (13)$$

where the p_j are probabilities, $\rho_j^{R_AA}$ are quantum states of system R_AA, and $\rho_j^{BR_B}$ are quantum states of system BR_B.

When we say in the statement of the theorem that $\rho(\mathcal{E})$ is separable, there is initially some ambiguity, due to the multiple ways the system R_AABR_B can be decomposed into subsystems. To avoid this ambiguity, it is convenient to introduce notational conventions as follows. Let σ be a state of a composite system CD. We say σ is *separable with respect to the C:D cut* if σ can be written as $\sigma=\Sigma_j p_j\rho_j^C\otimes\rho_j^D$ for probabilities p_j, and quantum states ρ_j^C,ρ_j^D of systems C and D, respectively. The advantage of this notation comes when more systems are introduced. For example, in the operation-separability theorem, the assertion is that $\mathcal{E}$ is separable if and only if $\rho(\mathcal{E})$ is separable with respect to the R_AA:BR_B cut.

We have stated the operation-separability theorem for trace-preserving quantum operations, but a similar result also holds for non-trace-preserving quantum operations $\mathcal{E}$. The only change is that the p_j are no longer probabilities, but can be any set of non-negative real numbers. We have also restricted our attention to bipartite quantum operations, that is, $\mathcal{E}$ which act on quantum systems with just two components, A and B; it is not difficult to show that an analogous statement also holds for k-party quantum operations $\mathcal{E}$.

A nice corollary of the operation-separability theorem is that a quantum operation is separable if and only if it is incapable of producing entangled states. Furthermore, by connecting gate separability to state separability, the operation-separability theorem allows us to apply results from the theory of state separability to prove that certain gates are separable, and thus incapable of producing entanglement.

The operation-separability theorem tells us that a trace-preserving quantum operation $\mathcal{E}$ is separable precisely when $\rho(\mathcal{E})$ is separable. However, it does not follow that all separable states of R_AA:BR_B can be written as $\rho(\mathcal{E})$ for some trace-preserving quantum operation. To understand this, observe that when $\mathcal{E}$ is trace preserving, $\mathrm{tr}_{AB}[\rho(\mathcal{E})]$ must be the completely mixed state of R_AR_B. In general, however, it is easy to find separable states σ of R_AA:BR_B such that $\mathrm{tr}_{AB}(\sigma)$ is not completely mixed.

An elegant result of Horodecki *et al.* [21] can be used to characterize precisely which separable states can be written in the form $\rho(\mathcal{E})$ for trace-preserving, separable $\mathcal{E}$. Their result, which we have restated in the context of multipartite systems, is as follows.

Theorem 2. The set of density matrices, σ, of R_AABR_B such that $\sigma=\rho(\mathcal{E})$ for some trace-preserving quantum operation $\mathcal{E}$ is precisely the set such that $\mathrm{tr}_{AB}(\sigma)$ is the completely mixed state of R_AR_B.

Combining this theorem with the operation-separability theorem we obtain the following result.

Theorem 3. The set of density matrices, σ, of R_AABR_B such that $\sigma=\rho(\mathcal{E})$ for some trace-preserving and separable quantum operation $\mathcal{E}$ is precisely the set such that (a) σ is separable with respect to the R_AA:BR_B cut; and (b) $\mathrm{tr}_{AB}(\sigma)$ is the completely mixed state of R_AR_B.

B. Separable gates and quantum computation

Having discussed the basic properties of separable quantum operations, we turn to their utility for quantum computation. Imagine that a quantum circuit is built entirely out of separable quantum gates and single-qubit gates. It is intuitively plausible that such a quantum circuit can be efficiently simulated on a classical computer, and we now prove

this result. The major technical difficulty is the accuracy required in the simulation, and the associated computational overhead.

Our model of quantum computation is as follows. Let $\mathcal{G}$ be a fixed set of one- and two-qubit quantum gates. By "quantum gate" we mean a trace-preserving quantum operation. We assume that all the two-qubit gates in $\mathcal{G}$ are separable. We let $\{C_n\}$ be a uniform family of quantum circuits [1,22] containing $p(n)$ gates, and acting on $q(n)$ qubits, where $p(n)$ and $q(n)$ are polynomials in some parameter n. The initial state of the computer is assumed to be a computational basis state, $|x\rangle$. The computation is concluded by performing a measurement in the computational basis, yielding a probability distribution $p_x(y)$ over measurement outcomes y. The measurement may be either on all the qubits, or on some prespecified subset. For instance, if one is solving a decision problem, it is only necessary to measure the first qubit of the computer, to get a single 0 or 1 as output.

What does it mean to simulate this computation efficiently on a classical computer? Suppose we have a classical computer that, on input of x, produces output y with probability distribution $\tilde{p}_x(y)$. A good measure of how well this simulates the quantum computation is the L_1 *distance*. For probability distributions $r(y)$ and $s(y)$, the L_1 distance is defined by $D(r(y),s(y)) \equiv \Sigma_y |r(y)-s(y)|/2$. Thus, we require that $D(p_x(y),\tilde{p}_x(y)) = \Sigma_y |p_x(y)-\tilde{p}_x(y)|/2$ satisfies

$$D(p_x(y),\tilde{p}_x(y)) \leq \epsilon \tag{14}$$

for some parameter $\epsilon > 0$. We will show that the computational resources to achieve this accuracy on a classical computer scale as $O(\mathrm{poly}(p(n)/\epsilon))$, where $\mathrm{poly}(\cdot)$ is some polynomial of fixed degree not depending on the circuit family $\{C_n\}$. Thus, high accuracies in the simulation can be achieved with modest computational cost.

As an example of the practical implications of this result, suppose $\{C_n\}$ is a uniform family of quantum circuits solving a decision problem, outputting the correct answer to an instance x of the decision problem with probability at least 3/4. Our result implies that there is a classical simulation using $O(\mathrm{poly}(p(n)))$ gates, and outputting the correct solution to the decision problem with probability 2/3. (The probability of obtaining the correct answer may easily be boosted up beyond 3/4 by a constant number of repetitions.)

To analyze the method described below for classical simulation, we need the notion of the *trace distance*, a quantum generalization of the L_1 distance. The trace distance $D(\rho,\sigma)$ between density matrices ρ and σ is defined by [1] $D(\rho,\sigma) \equiv \mathrm{tr}|\rho-\sigma|/2$. Note that we use the same notation $D(\cdot,\cdot)$ for the trace distance and the L_1 distance, with the meaning to be determined from context. The properties of the trace distance are discussed in detail in Ref. [1], and we need only a few properties here.

(1) The trace distance satisfies the *triangle inequality*, $D(\rho,\tau) \leq D(\rho,\sigma) + D(\sigma,\tau)$.

(2) The trace distance is *doubly convex*, meaning that if p_j are probabilities, and ρ_j and σ_j are corresponding density matrices, then

$$D\left(\sum_j p_j \rho_j, \sum_j p_j \sigma_j\right) \leq \sum_j p_j D(\rho_j,\sigma_j). \tag{15}$$

(3) The trace distance is *contractive*. That is, if $\mathcal{E}$ is a trace-preserving quantum operation, then $D(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \leq D(\rho,\sigma)$.

(4) The trace distance has the *stability property*, $D(\rho_1 \otimes \sigma, \rho_2 \otimes \sigma) = D(\rho_1,\rho_2)$.

(5) Suppose E_y are positive-operator-valued measure elements describing the statistics from an arbitrary quantum measurement. Let $r(y) \equiv \mathrm{tr}(\rho E_y)$ and $s(y) \equiv \mathrm{tr}(\sigma E_y)$ be the corresponding probability distributions for ρ and σ. Then the L_1 distance and the trace distance are related by the inequality

$$D(r(y),s(y)) \leq D(\rho,\sigma). \tag{16}$$

We now describe how the classical simulation is performed, followed by an analysis to determine the accuracy of the simulation.

Variables used in the classical simulation. For each $j = 1,\ldots,q(n)$, we let $\vec{s}_j$ be a three-dimensional real vector. Each vector $\vec{s}_j$ is *valid*, meaning that it has the following three properties: (a) Each component of $\vec{s}_j$ is in the range $[-1,1]$; (b) each component is specified to l bits of precision, where l is a number that will be fixed by the later analysis, in order to ensure the overall accuracy is at least ϵ; and (c) $\|\vec{s}_j\| \leq 1$.

We use the notation $\vec{s} \equiv (\vec{s}_1,\ldots,\vec{s}_{q(n)})$ to denote the $3q(n)$-dimensional real vector containing all the $\vec{s}_j$'s as subvectors. We say that $\vec{s}$ is *valid* if each $\vec{s}_j$ is valid. It will also be convenient to introduce the notation

$$\rho(\vec{s}) \equiv \frac{I+\vec{s}_1\cdot\vec{\sigma}}{2} \otimes \cdots \otimes \frac{I+\vec{s}_{q(n)}\cdot\sigma}{2}. \tag{17}$$

Note that $\rho(\vec{s})$ is a legitimate density operator of $q(n)$ qubits, whenever $\vec{s}$ is valid. The idea of the classical simulation is that the variables $\vec{s}$ will be used to represent the state $\rho(\vec{s})$. Note that $\rho(\vec{s})$ is *not* a variable used in the classical simulation; it is simply a mathematical notation convenient in the analysis of the simulation.

Initial state of the classical variables. Suppose the initial state of the quantum computer is $|x\rangle$, where x has binary expansion $x_1 \cdots x_{q(n)}$. If $x_j = 0$ we set $\vec{s}_j = (0,0,1)$ initially, while if $x_j = 1$ we set $\vec{s}_j = (0,0,-1)$ initially.

Simulating a single-qubit gate. A single-qubit gate can be regarded as a two-qubit separable gate in which one of the qubits is acted on trivially. Thus, we need only consider the case of two-qubit separable gates.

Simulating a two-qubit separable gate. Suppose $\mathcal{E}$ is a two-qubit separable gate, and it acts on qubits A and B. We simulate this gate by using $\vec{s}$ as input to the following stochastic gate simulation procedure, which produces a valid $3q(n)$-dimensional vector $\vec{s}\,'$ as output. We then set $\vec{s} = \vec{s}\,'$, and repeat over, going through each gate, $\mathcal{E}_1,\ldots,\mathcal{E}_{p(n)}$, in

the computation, until a final output value of $\vec{s}$ is produced, at which point we proceed to the simulation of the final measurement, described below.

Gate simulation procedure. The procedure is as follows.

(1) Input to the procedure: A valid vector, $\vec{s}$.

(2) Body of the procedure: Find valid three-vectors $\vec{s}_A^j$ and $\vec{s}_B^j$, and a probability distribution p_j containing at most 16 elements, and with each p_j specified to l bits of precision, such that

$$D\left(\mathcal{E}\left(\frac{I+\vec{s}_A\cdot\vec{\sigma}}{2}\otimes\frac{I+\vec{s}_B\cdot\vec{\sigma}}{2}\right)\sum_j p_j\frac{I+\vec{s}_A^j\cdot\vec{\sigma}}{2}\otimes\frac{I+\vec{s}_B^j\cdot\vec{\sigma}}{2}\right)\leqslant c2^{-l}, \tag{18}$$

for some constant c that does not depend on $\mathcal{E}$, A, or B. To see that this is possible, we make use of the fact that

$$\mathcal{E}\left(\frac{I+\vec{s}_A\cdot\vec{\sigma}}{2}\otimes\frac{I+\vec{s}_B\cdot\vec{\sigma}}{2}\right) \tag{19}$$

is a separable, two-qubit state, and therefore, by Carathéodory's theorem [23], can be written in the form

$$\sum_j q_j\frac{I+\vec{t}_A^j\cdot\vec{\sigma}}{2}\otimes\frac{I+\vec{t}_B^j\cdot\vec{\sigma}}{2},$$

where the q_j are probabilities, $\vec{t}_A^j,\vec{t}_B^j$ are real three-vectors satisfying $\|\vec{t}_A^j\|,\|\vec{t}_B^j\|\leqslant 1$, and there are at most 16 terms in the sum. Choosing the p_j to be probabilities which are l-bit approximations to the q_j, and the $\vec{s}_A^j,\vec{s}_B^j$ to be valid vectors which approximate $\vec{t}_A^j,\vec{t}_B^j$ also to l bits, we obtain the result.

Note that while Carathéodory's theorem ensures that such probabilities and vectors exist, finding them may be nontrivial. The obvious technique, a brute force search over probability distributions and valid vectors, requires poly(2^l) operations, where poly$(\cdot)$ is some fixed polynomial function. Although we believe that likely better techniques—perhaps even polynomial in l—are possible, for the purposes of the present simulation poly(2^l) turns out to be sufficient.

(3) Output of the procedure: For $k\neq A,B$ we define $\vec{s}_k^j\equiv\vec{s}_k$. Set $\vec{s}^j=(\vec{s}_1^j,\ldots,\vec{s}_{q(n)}^j)$. Note that $\vec{s}^j$ is valid, by construction. With probability p_j, output $\vec{s}\,'=\vec{s}^j$.

Simulating the final measurement in the computational basis. Let S be the subset of qubits that is measured at the output of the quantum computation. For each $k\in S$, let s_k^3 be the third component of $\vec{s}_k$. The measurement result for that qubit is 0 with probability $(1+s_k^3)/2$, and 1 with probability $(1-s_k^3)/2$. Note that, by definition, $\tilde{p}_x(y)$ is the distribution over possible outcomes, y, produced by following this procedure.

Analysis. The key to the analysis of the classical simulation is a simple equivalence between the classical simulation and certain measurements on quantum states. Suppose we define $\tilde{p}^m(\vec{s})$ to be the probability distribution on valid vectors after m steps of the simulation procedure, that is, after $\mathcal{E}_1,\ldots,\mathcal{E}_m$ have been simulated. For $m=0,\ldots,p(n)$ define

$$\tilde{\sigma}^m\equiv\sum_{\vec{s}}\tilde{p}^m(\vec{s})\rho(\vec{s}). \tag{20}$$

It is not difficult to see that the distribution obtained by measuring $\tilde{\sigma}^{p(n)}$ in the computational basis of the subset S is exactly the same as the output distribution $\tilde{p}_x(y)$ produced by the classical simulation.

For $m=0,\ldots,p(n)$ define σ^m to be the state of the actual quantum computer after m gates have been applied. Thus $\sigma^0=|x\rangle\langle x|$, $\sigma^1=\mathcal{E}_1(\sigma^0)$, and so on. The idea of the proof that the classical simulation works well is to bound the distance between σ^m and $\tilde{\sigma}^m$. We do this using the following lemma.

Lemma 1. Suppose a valid vector $\vec{s}$ is used as input to the gate simulation procedure with probability $p(\vec{s})$, and let $p(\vec{s}\,')$ be the corresponding output distribution on valid vectors. Define

$$\sigma\equiv\sum_{\vec{s}}p(\vec{s})\rho(\vec{s}),\quad \sigma'\equiv\sum_{\vec{s}\,'}p(\vec{s}\,')\rho(\vec{s}\,'). \tag{21}$$

If the gate simulation procedure simulates the gate $\mathcal{E}$, then we have

$$D(\mathcal{E}(\sigma),\sigma')\leqslant c2^{-l}, \tag{22}$$

where c is the constant introduced earlier in the discussion of the gate simulation procedure.

Proof. Let $p(\vec{s}\,'|\vec{s})$ be the probability that $\vec{s}\,'$ is output by the gate simulation procedure, given that $\vec{s}$ is input. Then we have $p(\vec{s}\,')=\Sigma_{\vec{s}}p(\vec{s}\,'|\vec{s})p(\vec{s})$, so

$$\sigma'=\sum_{\vec{s}}p(\vec{s})\sum_{\vec{s}\,'}p(\vec{s}\,'|\vec{s})\rho(\vec{s}\,'). \tag{23}$$

Applying the double convexity of the trace distance gives

$$D(\mathcal{E}(\sigma),\sigma')\leqslant\sum_{\vec{s}}p(\vec{s})D\left(\mathcal{E}(\rho(\vec{s})),\sum_{\vec{s}\,'}p(\vec{s}\,'|\vec{s})\rho(\vec{s}\,')\right). \tag{24}$$

By inspection of the construction used in the gate simulation procedure, notably Eq. (18), and the stability property for trace distance, we have

$$D\left(\mathcal{E}(\rho(\vec{s})),\sum_{\vec{s}\,'}p(\vec{s}\,'|\vec{s})\rho(\vec{s}\,')\right)\leqslant c2^{-l}. \tag{25}$$

Combining this observation with Eq. (24) gives $D(\mathcal{E}(\sigma),\sigma')\leqslant c2^{-l}$, which was the desired result. ■

Proposition 2. For $m=0,\ldots,p(n)$, $D(\sigma^m,\tilde{\sigma}^m)\leqslant cm2^{-l}$.

Proof. We induct on m. For $m=0$, the result follows from the fact that $\sigma^0=\tilde{\sigma}^0$. Assuming that the result is true for m, we now prove it for $m+1$. By the triangle inequality

$$D(\sigma^{m+1},\tilde{\sigma}^{m+1}) \leq D(\sigma^{m+1},\mathcal{E}_{m+1}(\tilde{\sigma}^m)) + D(\mathcal{E}_{m+1}(\tilde{\sigma}^m),\tilde{\sigma}^{m+1}). \quad (26)$$

By definition $\sigma^{m+1}=\mathcal{E}_{m+1}(\sigma^m)$, so this equation may be rewritten as

$$D(\sigma^{m+1},\tilde{\sigma}^{m+1}) \leq D(\mathcal{E}_{m+1}(\sigma^m),\mathcal{E}_{m+1}(\tilde{\sigma}^m)) + D(\mathcal{E}_{m+1}(\tilde{\sigma}^m),\tilde{\sigma}^{m+1}). \quad (27)$$

Applying the contractivity of the trace distance to the first term, and Lemma 1 to the second term, we obtain

$$D(\sigma^{m+1},\tilde{\sigma}^{m+1}) \leq D(\sigma^m,\tilde{\sigma}^m) + c2^{-l}. \quad (28)$$

Applying the inductive hypothesis to the first term gives

$$D(\sigma^{m+1},\tilde{\sigma}^{m+1}) \leq cm2^{-l} + c2^{-l} = c(m+1)2^{-l}, \quad (29)$$

which completes the induction. ■

We conclude from the proposition that $D(\sigma^{p(n)},\tilde{\sigma}^{p(n)}) \leq cp(n)2^{-l}$. It follows from Eq. (16) that the simulated distribution $\tilde{p}_x(y)$ and the actual distribution $p_x(y)$ are related by the inequality $D(p_x(y),\tilde{p}_x(y)) \leq cp(n)2^{-l}$. Choosing l to be the least integer greater than $\log_2[cp(n)/\epsilon]$, we therefore have

$$D(p_x(y),\tilde{p}_x(y)) \leq \epsilon. \quad (30)$$

The total number of times the gate simulation procedure is performed is $p(n)$, and the number of operations performed in one iteration of the gate simulation procedure scales as $\mathrm{poly}(2^l)$, so the total number of operations in the classical simulation is $O(\mathrm{poly}(p(n)/\epsilon))$, where we change the notation by letting $\mathrm{poly}(\cdot)$ be a (new) polynomial function. We have proved the following theorem.

Theorem 4. Let $\mathcal{G}$ be a fixed set of one- and two-qubit gates. Suppose all two-qubit gates in $\mathcal{G}$ are separable. Let $\{C_n\}$ be a uniform family of quantum circuits of size $p(n)$, acting on $q(n)$ qubits, where both $p(n)$ and $q(n)$ are polynomials. The initial state of the computer is a computational basis state, $|x\rangle$. The computation is concluded by performing a measurement in the computational basis on some prespecified subset S of the qubits, yielding a probability distribution $p_x(y)$ over possible measurement outcomes y. Then for any $\epsilon>0$, it is possible to sample from a distribution $\tilde{p}_x(y)$ satisfying $D(p_x(y),\tilde{p}_x(y))<\epsilon$ using a classical algorithm taking $O(\mathrm{poly}(p(n)/\epsilon))$ steps, where $\mathrm{poly}(\cdot)$ is a fixed polynomial.

Results related to Theorem 4 have been obtained in the past, but, so far as we have determined, no proof of this result has previously been published. In particular, Aharonov and Ben-Or [11] studied the role of entanglement in quantum computation, proving that many-party entanglement must be present in order for a quantum computation to be difficult to simulate classically. This conclusion was subsequently clarified and extended by Jozsa and Linden [24]. However, the conclusions of both Refs. [11,24] are not applicable in the present context, since they apply in the context of pure state entanglement of a quantum computer, rather than the mixed-state case considered in this paper.

The issue of mixed-state quantum information processing was considered by Braunstein *et al*. [25], who raised, without answering, the question of what role mixed-state entanglement can play in quantum computation.This line of thought has been carried further by many authors, without completely answering the question. See Refs. [26,27] for recent work and further references.

C. Separability-preserving gates

It is straightforward to extend the proof of Theorem 4 in a variety of ways, without changing the conclusion that an efficient classical simulation of the quantum circuit is possible. In particular, we can change the gates in $\mathcal{G}$ so they act on any *bounded number* of *qudit* systems, rather than *two-qubit* systems.

Furthermore, the proof relies on properties of gates in $\mathcal{G}$ that are weaker than separability. In particular, the gates in $\mathcal{G}$ need only be *separability preserving*, that is, $\mathcal{E}(\rho)$ is separable for any separable state ρ. We denote the class of separability-preserving gates by SP. To see that this is a weaker property, note that SWAP is separability preserving since it maps product states to product states, but SWAP is not separable, since it can generate entanglement with the aid of local ancilla systems. More generally, note that $\mathcal{E}_{AB}$ is separable with respect to $A:B$ if and only if $\mathcal{E}_{AB}\otimes\mathcal{I}_{A'B'}$ is separability preserving with respect to $AA':BB'$.

Since the proof of Theorem 4 only relied on the state in Eq. (19) being separable, it still holds when the available gates are all separability preserving. However, no simple and easy-to-use characterization of the separability-preserving gates is known, which is why we prefer, for most of the remainder of this paper, to work with the separable gates. We do make occasional later use of separability-preserving gates, so it is convenient to note here a few properties. Note that all separable gates are in SP, and for gates operating on multiple qudits, any permutation of the qudits (for example, SWAP) is in SP. Furthermore, SP is convex and is closed under composition, so

$$SP \supseteq \mathrm{Hull}\{\mathcal{E}\circ\mathcal{P}: \mathcal{E} \text{ separable and } \mathcal{P} \text{ a permutation}\}. \quad (31)$$

However, it is unclear whether this convex hull describes all of SP. For example, the operation which measures a pair of qubits in the Bell basis and stores the answer in the computational basis [i.e., $(|00\rangle+|11\rangle)/\sqrt{2}$ becomes $|00\rangle$, $(|00\rangle-|11\rangle)/\sqrt{2}$ becomes $|01\rangle$, etc., . . .] is certainly in SP though it does not seem that it can be expressed as a convex combination of $\mathcal{E}_k\circ\mathcal{P}_k$ for separable $\mathcal{E}_k$ and permutations $\mathcal{P}_k$.[2]

[2]We thank Keiji Matsumoto for pointing this out to us.

IV. ROBUSTNESS OF QUANTUM STATES

To understand how robust quantum gates are to noise, it is useful to first review prior work on the robustness of entangled quantum states. This section describes Vidal's and Tarrach's [12] definitions and results on the robustness of quantum states, introduces a measure of robustness, and relates that measure to Vidal and Tarrach's measure. The measure and its properties will be of special interest in applications to gate robustness.

Let ρ be a quantum state of a bipartite system AB, and let σ be a state of AB. Vidal and Tarrach [12] define the *robustness of ρ relative to σ*, $R(\rho\|\sigma)$, to be the smallest nonnegative number t such that the state

$$\frac{1}{1+t}\rho+\frac{t}{1+t}\sigma \tag{32}$$

is separable. Equivalently, we can define $R(\rho\|\sigma)$ to be the smallest non-negative t such that $\rho+t\sigma$ is separable; this latter definition in terms of unnormalized quantum states is frequently useful. Note that Ref. [12] specifies that σ be separable; however, we will find it convenient to extend the definition to nonseparable σ also, specifying that $R(\rho\|\sigma)\equiv+\infty$ if no value of t exists such that the state in Eq. (32) is separable. At first sight one is tempted to ask why we choose this definition for the robustness, and not the related quantity

$$\min\{p:p\geqslant 0,(1-p)\rho+p\sigma \text{ is separable}\}. \tag{33}$$

This latter definition has a more obvious physical interpretation as the minimal probability with which σ can be mixed with ρ to obtain a separable state. It follows from the definitions that the quantity of Eq. (33) is equal to $R(\rho\|\sigma)/[1+R(\rho\|\sigma)]$. The reason we do not work with the quantity of Eq. (33), despite its apparently more compelling physical interpretation, is that the robustness defined in Eq. (32) has useful and easy-to-prove convexity properties not satisfied by Eq. (33), namely, $R(\rho\|\sigma)$ is convex in both the first and the second entry.

A special case of $R(\rho\|\sigma)$ of particular interest is the *random robustness*, defined to be the robustness of ρ relative to the maximally mixed state I/d_Ad_B. We denote the random robustness of a state ρ by $R_r(\rho)\equiv R(\rho\|I/d_Ad_B)$. Vidal and Tarrach [12] found a useful formula for the random robustness of a pure state ψ of AB in terms of a Schmidt decomposition $\psi=\Sigma_j\psi_j|j\rangle|j\rangle$ with ordered Schmidt coefficients $\psi_1\geqslant\psi_2\geqslant\cdots\geqslant 0$:

$$R_r(\psi)=\psi_1\psi_2d_Ad_B. \tag{34}$$

So far we have discussed the robustness of a state ρ *relative* to another fixed state σ. We now define *the robustness* of ρ, $R(\rho)$, to be the *minimum* relative robustness $R(\rho\|\sigma)$ over all separable σ. Thus, the robustness of ρ is a measure of how much local noise can be mixed with ρ before it becomes separable.

We have defined three notions of robustness for quantum states, $R(\rho\|\sigma)$,$R_r(\rho)$, and $R(\rho)$. All three definitions have assumed that ρ is a state of a *bipartite* quantum system AB. However, robustness is easily extended to more than two parties, and it is convenient to have a notation to express the extended notion. Suppose, for example, that ρ and σ are states of a tripartite system ABC. Then $R^{A:B:C}(\rho\|\sigma)$ is defined to be the minimal value of t such that $\rho+t\sigma$ is separable with respect to $A:B:C$. Different ways of grouping the components of many-party quantum systems are handled in the obvious way. So, for example, we can define a notion of robustness, $R^{A:BC}(\rho\|\sigma)$, when systems B and C are grouped together. Explicitly, $R^{A:BC}(\rho\|\sigma)$ is the minimal value of t such that $\rho+t\sigma$ is separable with respect to $A:BC$.

These examples may be extended in a natural way to the random robustness and robustness, as well as to the case where more systems are present, and to more complicated groupings of subsystems. Most of our work concerns two-party robustness, and so we usually do not explicitly include superscripts in expressions such as $R^{A:B}(\rho)$.

The robustness has many useful properties, which are explored in detail in Ref. [12]. We mention just a few of the more striking properties here. The robustness is invariant under local unitary operations. Moreover, it is an entanglement monotone that cannot be increased under local operations and classical communication. It is also a convex function of ρ. As for the random robustness, Vidal and Tarrach [12] have obtained an elegant formula for the robustness in the special case of a pure state, ψ, of a bipartite system AB,

$$R(\psi)=\left(\sum_j\psi_j\right)^2-1, \tag{35}$$

where ψ_j are the Schmidt coefficients for ψ. In the course of their proof, Vidal and Tarrach explicitly construct a state σ_ψ, such that $|\psi\rangle\langle\psi|+R(\psi)\sigma_\psi$ is separable. σ_ψ may be expressed in terms of the Schmidt decomposition $|\psi\rangle=\Sigma_j\psi_j|j\rangle|j\rangle$ by

$$\sigma_\psi=\frac{1}{R(\psi)}\sum_{k\neq l}\psi_k\psi_l|k\rangle\langle k|\otimes|l\rangle\langle l|. \tag{36}$$

In the definition of robustness we mixed ρ with a *separable* quantum state σ, trying to determine what minimal level of mixing will produce separability. Another natural definition of robustness would allow σ to range over *arbitrary* density matrices, not just separable density matrices. That is, we can define $R_g(\rho)\equiv\min_\sigma R(\rho\|\sigma)$, where the g subscript indicates that we are minimizing *globally* over all possible density matrices σ.

How are $R_g(\rho)$ and $R(\rho)$ related? It is clear from the definitions that $R_g(\rho)\leqslant R(\rho)$. We will prove that the reverse inequality is also true when $\rho=\psi$ is a pure state:

$$R_g(\psi)=R(\psi)=\left(\sum_j\psi_j\right)^2-1. \tag{37}$$

We do not know whether $R_g(\rho)=R(\rho)$ in general. To complete the proof of Eq. (37), we show that if there exists a density operator σ such that $\psi+t\sigma$ is separable, then $t\geqslant(\Sigma_j\psi_j)^2-1$. (Our proof both extends and simplifies a similar proof in Ref. [12] for the robustness $R(\rho)$.)

The proof is based on the positive partial transpose criterion of Peres [28]. Let us denote the partial transpose on systems A and B by T_A and T_B, respectively. Then the positive partial transpose criterion implies that if $\psi+t\sigma$ is separable, then $\psi^{T_B}+t\sigma^{T_B}$ is a positive operator.

Next, we introduce an operator M defined by $M\equiv I-U_{\mathrm{SWAP}}$, where $U_{\mathrm{SWAP}}\equiv \Sigma_{jk}|j\rangle\langle k| \otimes |k\rangle\langle j|$ is the linear operator interchanging states of system A and system B. Note that M is positive, since $U_{\mathrm{SWAP}}^2=I$ implies that U_{SWAP} has eigenvalues ± 1, and thus M is a diagonalizable operator with eigenvalues 0 and 2.

We now combine the results of the previous two paragraphs. Since the trace of a product of two positive operators is non-negative, it follows that $0\leqslant \mathrm{tr}(M\psi^{T_B})+t\ \mathrm{tr}(M\sigma^{T_B})$. Using a little algebra and the observation that for any two operators K and L $\mathrm{tr}(KL^{T_B})=\mathrm{tr}(K^{T_A}L)$, this inequality may be rewritten as

$$-\mathrm{tr}(M^{T_A}\psi)\leqslant t\ \mathrm{tr}(M^{T_A}\sigma). \tag{38}$$

Direct calculation shows that $M^{T_A}=I-|\alpha\rangle\langle\alpha|$, where $|\alpha\rangle \equiv\Sigma_j|j\rangle|j\rangle$ is the (unnormalized) maximally entangled state. It follows that $\mathrm{tr}(M^{T_A}\psi)=1-(\Sigma_j\psi_j)^2$ and $\mathrm{tr}(M^{T_A}\sigma) \leqslant\mathrm{tr}(\sigma)=1$. Substituting these results into Eq. (38) gives $(\Sigma_j\psi_j)^2-1\leqslant t$, which was the desired bound.

V. ROBUSTNESS OF QUANTUM GATES

We now extend state robustness to *quantum gates*. Suppose $\mathcal{E}$ and $\mathcal{F}$ are trace-preserving quantum operations on a composite system AB. Then we define the *robustness of $\mathcal{E}$ relative to $\mathcal{F}$*, $R(\mathcal{E}\|\mathcal{F})$, to be the minimum value of t such that

$$\frac{1}{1+t}\mathcal{E}+\frac{t}{1+t}\mathcal{F} \tag{39}$$

is separable. Equivalently, $R(\mathcal{E}\|\mathcal{F})$ can be defined to be the minimal value of t such that $\mathcal{E}+t\mathcal{F}$ is separable. Applying the operation-separability theorem, we immediately find the useful formula

$$R(\mathcal{E}\|\mathcal{F})=R^{R_AA:BR_B}(\rho(\mathcal{E})\|\rho(\mathcal{F})). \tag{40}$$

Just as for quantum states, the notion of gate robustness extends in a natural way to systems of more than two parties, and we use notations analogous to those introduced earlier, such as $R^{A:B:C}(\mathcal{E}\|\mathcal{F})$ and $R^{A:BC}(\mathcal{E}\|\mathcal{F})$, to describe this scenario. Note that these notations will also be extended in a natural way to the random robustness and robustness of a quantum gate, as defined below. As for quantum states, when identifying superscripts is omitted, we assume that the quantum gate in question acts on a bipartite system AB.

Motivated by several different classes of noise commonly occurring in physical systems, we now use the notion of relative gate robustness to define and study several different measures of robustness for quantum gates. First is the random robustness, which we define and study in Sec. V A. Also in this section, we use results on the random robustness to place bounds on the threshold for quantum computation. Two other measures of robustness are the *separable robustness* and the *global robustness*, which we define in Sec. V B, and use to prove bounds on the threshold for quantum computation.Our results on these measures of robustness are less complete, and so our discussion is more limited.

A. Random robustness of quantum gates

1. Definition and basic properties

The random robustness of $\mathcal{E}$, $R_r(\mathcal{E})$, is defined to be equal to the robustness of $\mathcal{E}$ relative to the completely depolarizing channel, $\mathcal{D}(\rho)=I/d_Ad_B$ for all states ρ of system AB:

$$R_r(\mathcal{E})\equiv R(\mathcal{E}\|\mathcal{D}). \tag{41}$$

The random robustness is especially interesting because it measures the robustness of $\mathcal{E}$ against complete randomization of systems A and B. Another way of stating this is to imagine that we are applying the operation $\mathcal{E}$ with probability $1-p$, and randomizing the systems A and B with probability p. Then the threshold probability at which this gate crosses the separable-inseparable threshold is

$$p_{\mathrm{th}}=\frac{R_r(\mathcal{E})}{1+R_r(\mathcal{E})}. \tag{42}$$

From Eq. (40), we see that the random robustness for an operation is related to the random robustness of a state by

$$R_r(\mathcal{E})=R_r^{R_AA:BR_B}(\rho(\mathcal{E})). \tag{43}$$

Specializing to the case where $\mathcal{E}$ is a unitary quantum operation U, we see that $R_r(U)=R_r^{R_AA:BR_B}(\rho(U))$. However, $\rho(U)$ is a pure state. We showed earlier that $\rho(U)$ has Schmidt coefficients $u_j/\sqrt{d_Ad_B}$, where u_j are the Schmidt coefficients of U. This observation, together with Eqs. (43) and (34) implies the formula

$$R_r(U)=d_Ad_Bu_1u_2, \tag{44}$$

where we order the Schmidt coefficients of U so that $u_1 \geqslant u_2\geqslant\cdots\geqslant 0$. [Note that in deriving this equation, we have replaced d_A by d_A^2, and d_B by d_B^2 in Eq. (34), since we are working with robustness for the $R_AA:BR_B$ system.]

It is, perhaps, not immediately clear what the physical relevance of the random robustness is.After all, in real physical systems, the effects of noise on a quantum gate will not usually be to simply mix in some depolarization, together with the gate. Despite this, there is still a very good physical reason to be interested in the random robustness. The reason is that, as we show in more explicit detail below, the random robustness can be used to analyze the particular noise models which have been used in estimating bounds on the threshold for quantum computation. In turn, it has been argued [6–10] that by analyzing and correcting for the effects of noise in those *particular* models, it is possible to make general statements about a wide class of physically reasonable noise models. Thus, although the physical scenario considered in

the definition of the random robustness appears rather specialized, it will enable insight into much more general physical situations.

As an example, we may ask how robust the CNOT is against the effects of depolarizing noise? The CNOT has the Schmidt decomposition [13] $\sqrt{2}|0\rangle\langle 0|\otimes I/\sqrt{2}+\sqrt{2}|1\rangle\langle 1|\otimes X/\sqrt{2}$, so Eq. (44) implies that $R_r(\text{CNOT})=8$. Interestingly, we can also show that the CNOT is the most robust two-qubit gate. Indeed, a more general bound on the random robustness may be proved for an arbitrary quantum operation, $\mathcal{E}$, of a composite system AB. The argument is as follows. Suppose ψ is a quantum state of a system CD, where C has dimension d_C and D has dimension d_D. We have $R_r(\psi)=d_C d_D \psi_1 \psi_2 \leq d_C d_D/2$. By convexity of the random robustness, we have $R_r(\rho)\leq d_C d_D/2$ for any state ρ of CD. Applying this result to the state $\rho(\mathcal{E})$ of $R_A A{:}BR_B$ gives $R_r(\mathcal{E})=R_r(\rho(\mathcal{E}))\leq d_A^2 d_B^2/2$. These results may be summarized as a proposition.

Proposition 3. Let $\mathcal{E}$ be any trace-preserving quantum operation acting on systems A and B. Then $R_r(\mathcal{E})\leq d_A^2 d_B^2/2$. If $d_A=d_B=2$ then $R_r(\mathcal{E})\leq R_r(\text{CNOT})=8$.

The random robustness has many physically interesting properties. Below we list six easily proved properties, before discussing in more depth two less easily proved properties. Our discussion of these properties is, in part, motivated by the framework of ''dynamic strength'' measures introduced in Ref. [13], although the properties we discuss are interesting independent of that motivation. In Ref. [13] it was argued that these properties, especially the property of *chaining*, discussed below, are essential if a measure can be said to quantify the strength of a quantum dynamical operation as a physical resource. By showing that these properties are satisfied, we thus show that the random robustness is a good measure of dynamic strength.

(1) Non-negativity and locality: $R_r(\mathcal{E})\geqslant 0$ with equality if and only if $\mathcal{E}$ is a separable quantum operation.

(2) Local unitary invariance: If $\mathcal{U}_A,\mathcal{U}_B,\mathcal{V}_A,\mathcal{V}_B$ are all local unitary quantum operations, with the system being acted on indicated by the subscript, then $R_r((\mathcal{U}_A\otimes\mathcal{U}_B)\circ\mathcal{E}\circ(\mathcal{V}_A\otimes\mathcal{V}_B))=R_r(\mathcal{E})$.

(3) Exchange symmetry: $R_r(\mathcal{E})=R_r(\text{SWAP}\circ\mathcal{E}\circ\text{SWAP})$, that is, the random robustness is not affected if we interchange the role of the systems.

(4) Time-reversal invariance: For a unitary U, $R_r(U)=R_r(U^\dagger)$.

(5) Convexity: The random robustness $R_r(\mathcal{E})$ is *convex* in $\mathcal{E}$.

(6) Reduction: Suppose a trace-preserving quantum operation $\mathcal{E}$ acting on AB is obtained from a trace-preserving quantum operation $\mathcal{F}$ acting on ABC as follows: $\mathcal{E}(\rho_{AB})=\text{tr}_C[\mathcal{F}(\rho_{AB}\otimes\sigma_C)]$, for some fixed state σ_C of system C. Then the random robustness satisfies the reduction property, namely, $R_r^{A:B}(\mathcal{E})\leq R_r^{A:BC}(\mathcal{F})$.

The random robustness satisfies two other physically interesting properties that are more difficult to prove. First of all, the random robustness is *continuous* in $\mathcal{E}$. Physically, this is self-evident: making a small change in $\mathcal{E}$ should not too drastically affect its robustness against the effects of noise. We now prove a quantitative form of this statement for unitary gates.

Proposition 4 (continuity of random robustness). Let U and V be unitary gates acting on a system A of dimension d_A, and a system B of dimension d_B. Then

$$|R_r(U)-R_r(V)|\leq d_M d_A^3 d_B^3 \|U-V\|^2, \tag{45}$$

where $d_M\equiv\min(d_A,d_B)$.

Proof. Let u_j and v_j be the ordered Schmidt coefficients of U and V, respectively. From Eq. (44),

$$\begin{aligned}|R_r(U)-R_r(V)| &= d_A d_B |u_1 u_2 - v_1 v_2| \\ &= d_A d_B |(u_1-v_1)u_2 + v_1(u_2-v_2)| \\ &\leq d_A d_B |u_1-v_1||u_2| + |v_1||u_2-v_2| \\ &\leq d_A^2 d_B^2 (|u_1-v_1|+|u_2-v_2|) \\ &\leq d_A^2 d_B^2 \sum_j |u_j-v_j|. \end{aligned} \tag{46}$$

The second part of the proof is to observe that by the Cauchy-Schwartz inequality,

$$\sum_j |u_j-v_j| \leq d_M \sum_j (u_j^2+v_j^2-2u_j v_j) \tag{47}$$

$$= 2 d_M d_A d_B \left(1-\frac{\sum_j u_j v_j}{d_A d_B}\right). \tag{48}$$

Applying Proposition 1, we obtain $\Sigma_j|u_j-v_j| \leq d_M d_A d_B \|U-V\|^2$. Combining with Eq. (46) gives the result. ■

Another physically interesting question is to ask how the random robustness of a gate $\mathcal{E}_1\circ\mathcal{E}_2$ composed of quantum gates $\mathcal{E}_1$ and $\mathcal{E}_2$ relates to the random robustness of the individual gates. The following proposition bounds the random robustness of the combined operation.

Proposition 5 (chaining for random robustness). Let $\mathcal{E}_1$ be a doubly stochastic quantum operation, that is, a quantum operation which is both trace preserving and unital [i.e., $\mathcal{E}_1(I)=I$], and let $\mathcal{E}_2$ be an arbitrary trace-preserving quantum operation. Then

$$R_r(\mathcal{E}_1\circ\mathcal{E}_2)\leq R_r(\mathcal{E}_1)+R_r(\mathcal{E}_2)+R_r(\mathcal{E}_1)R_r(\mathcal{E}_2). \tag{49}$$

Note that unitary operations are trace preserving and unital, so the proposition is true when $\mathcal{E}_1$ and $\mathcal{E}_2$ are unitary. There is an equivalent way of phrasing Proposition 5 that is physically more intuitive. Define

$$C_r(\mathcal{E})\equiv\ln[1+R_r(\mathcal{E})]. \tag{50}$$

Then $C_r(\mathcal{E})$ is monotonically related to the random robustness of $\mathcal{E}$, and thus can be thought of as carrying the same

qualitative information about the robustness of the gate. Simple algebra shows that the conclusion of Proposition 5 may be recast in the form

$$C_r(\mathcal{E}_1\circ\mathcal{E}_2)\leqslant C_r(\mathcal{E}_1)+C_r(\mathcal{E}_2). \tag{51}$$

The simplicity and clarity of this form may, perhaps, make it more useful in some circumstances.

Proof. By definition of the random robustness, the quantum operations

$$\mathcal{E}_1+R_r(\mathcal{E}_1)\mathcal{D} \tag{52}$$

and

$$\mathcal{E}_2+R_r(\mathcal{E}_2)\mathcal{D} \tag{53}$$

are separable quantum operations. Furthermore, since the composition of two separable quantum operations is separable, and $\mathcal{E}_1\circ\mathcal{D}=\mathcal{D}\circ\mathcal{E}_2=\mathcal{D}\circ\mathcal{D}=\mathcal{D}$ (using the unitality of $\mathcal{E}_1$), we can compose the operations of Eqs. (52) and (53) to see that

$$\mathcal{E}_1\circ\mathcal{E}_2+[R_r(\mathcal{E}_1)+R_r(\mathcal{E}_2)+R_r(\mathcal{E}_1)R_r(\mathcal{E}_2)]\mathcal{D} \tag{54}$$

is separable, and thus

$$R_r(\mathcal{E}_1\circ\mathcal{E}_2)\leqslant R_r(\mathcal{E}_1)+R_r(\mathcal{E}_2)+R_r(\mathcal{E}_1)R_r(\mathcal{E}_2), \tag{55}$$

as required. ■

2. Random robustness and the threshold for quantum computation

Suppose we are trying to do fault-tolerant quantum computation using single-qubit gates and some entangling two-qubit unitary gate U. U might be the CNOT gate; it can also be any other entangling two-qubit gate, at least in principle [29,30], and still be capable of universal quantum computation when assisted by single-qubit gates. Suppose, furthermore, that the U gates are afflicted with noise of a special type, namely, immediately after a gate acts, each qubit is independently depolarized with probability p. Let $\mathcal{U}(\rho)\equiv U\rho U^\dagger$ denote the quantum operation corresponding to U. Then the quantum operation describing this noise process is

$$\begin{aligned}\mathcal{E}(\rho)=&(1-p)^2\mathcal{U}(\rho)+p(1-p)(\mathcal{D}\otimes\mathcal{I})\circ\mathcal{U}(\rho)\\&+p(1-p)(\mathcal{I}\otimes\mathcal{D})\circ\mathcal{U}(\rho)+p^2(\mathcal{D}\otimes\mathcal{D})\circ\mathcal{U}(\rho).\end{aligned} \tag{56}$$

Note that $(\mathcal{D}\otimes\mathcal{D})\circ\mathcal{U}(\rho)=(\mathcal{D}\otimes\mathcal{D})(\rho)$, so this expression can be simplified to

$$\begin{aligned}\mathcal{E}(\rho)=&(1-p)^2\mathcal{U}(\rho)+p(1-p)(\mathcal{D}\otimes\mathcal{I})\circ\mathcal{U}(\rho)\\&+p(1-p)(\mathcal{I}\otimes\mathcal{D})\circ\mathcal{U}(\rho)+p^2(\mathcal{D}\otimes\mathcal{D})(\rho).\end{aligned} \tag{57}$$

This expression cannot immediately be analyzed using our expressions for the random robustness of a gate, due to the two terms in which a single qubit is depolarized. Fortunately, we can simplify the analysis by showing that these terms are always separability preserving, that is, $(\mathcal{D}\otimes\mathcal{I})\circ\mathcal{U}$ and $(\mathcal{I}\otimes\mathcal{D})\circ\mathcal{U}$ are both in SP. This holds because for any ρ, $(\mathcal{D}\otimes\mathcal{I})\circ\mathcal{U}(\rho)=(I/d_A)\otimes\mathrm{tr}_A U\rho U^\dagger$, which is manifestly separable, and a similar result holds for $(\mathcal{I}\otimes\mathcal{D})\circ\mathcal{U}$. Note that such gates may not be separable: for example, $(\mathcal{D}\otimes\mathcal{I})\circ\mathrm{SWAP}+(\mathcal{I}\otimes\mathcal{D})\circ\mathrm{SWAP}$ is separability preserving, but not separable.

From this observation, and Eq. (57), it follows that $\mathcal{E}$ is in SP if $(1-p)^2\mathcal{U}+p^2(\mathcal{D}\otimes\mathcal{D})$ is separable. Comparing with the earlier results on random robustness, we see that this becomes true when $p^2/(1-p)^2=R_r(U)=8$. We see that $\mathcal{E}$ will be separability preserving when

$$p\geqslant\frac{R_r(U)-\sqrt{R_r(U)}}{R_r(U)-1}=\frac{8-\sqrt{8}}{7}\approx 0.74, \tag{58}$$

and thus, when this condition is satisfied, the quantum computation may be efficiently simulated on a classical computer. If we assume, as is usually done, that quantum computers may not be efficiently simulated on a classical computer, then it follows that the threshold for quantum computation is guaranteed to be less than 0.74.

In their work on obtaining upper bounds for the threshold, Aharonov and Ben-Or [11] considered a similar model of quantum computation, in which each qubit is independently dephased after each quantum gate. The main difference between their model and ours is that we have used depolarizing, rather than dephasing noise.Which of these more accurately describes the noise occurring in a real physical system depends, of course, upon the physical system in question. Aharonov and Ben-Or obtained an upper bound of $p_{\mathrm{th}}<0.97$; of course, this cannot be directly compared to our upper bound, since the noise models are different.

B. Robustness against more general noise

1. Definitions and general results

Depolarization is only one of many kinds of noise that may afflict a quantum gate. Other classes of noise motivate other measures of gate robustness. We now introduce two more measures of robustness, based on two natural classes of noise. The first measure is the separable robustness, which measures the gate's resilience against separable noise. The separable robustness $R_s(\mathcal{E})$ is defined to be the minimum relative robustness $R(\mathcal{E}\|\mathcal{F})$ over all separable, trace-preserving quantum operations $\mathcal{F}$. The second measure is the global robustness, which measures the resilience of the gate against arbitrary noise. The global robustness $R_g(\mathcal{E})$ is defined to be the minimum relative robustness $R(\mathcal{E}\|\mathcal{F})$ over all trace-preserving quantum operations $\mathcal{F}$.

A priori, it is apparent that $R_g(\mathcal{E})\leqslant R_s(\mathcal{E})$, but it is not clear whether or not the two quantities are equal. Furthermore, the gate robustnesses may be related to state robustness by the following inequalities:

$$R(\rho(\mathcal{E}))\leqslant R_s(\mathcal{E}), \tag{59}$$

$$R(\rho(U))\leqslant R_g(U). \tag{60}$$

To see the first of these inequalities, note that $\mathcal{E}+R_s(\mathcal{E})\mathcal{F}$ is separable, for some separable quantum operation $\mathcal{F}$. It follows that $\rho(\mathcal{E}+R_s(\mathcal{E})\mathcal{F})=\rho(\mathcal{E})+R_s(\mathcal{E})\rho(\mathcal{F})$ is a separable quantum state. Since $\rho(\mathcal{F})$ is separable, Eq. (59) follows from the definition of $R(\rho(\mathcal{E}))$. The proof of Eq. (60) is similar, but also makes use of the fact, noted in Eq. (37), that $R_g(\psi)=R(\psi)$ for any pure state ψ.

Do inequalities (59) and (60) hold with equality? We do not know the answer to this question, but suspect that the answer is, in general, ''no,'' in both cases. Our reasoning for this suspicion is as follows. Recall from Sec. III, in particular, Theorem 3, that not all separable states can be written as $\rho(\mathcal{F})$ for some separable quantum operation $\mathcal{F}$. Recall also the construction, Eq. (36), used in finding the separable σ_ψ which minimizes $R(\psi\|\sigma_\psi)$. Using this construction, it is not difficult to find examples of unitary U for which the separable state $\sigma_{\rho(U)}$ does not correspond to any trace-preserving, separable quantum operation, as characterized in Theorem 3.

Fortunately, there is a large and interesting class of gates for which inequalities (59) and (60) hold with equality. This class includes the CNOT and SWAP gates.

Theorem 5. Let U be a bipartite unitary gate acting on systems A and B with dimensions d_A and d_B. Assume that U has the Schmidt decomposition $U=\Sigma_j u_j A_j\otimes B_j$, where the A_j satisfy $A_jA_j^\dagger=I/d_A$ and the B_j satisfy $B_jB_j^\dagger=I/d_B$. That is, the A_j and B_j are all proportional to unitary operators. Then

$$R_g(U)=R_s(U)=R(\rho(U))=\frac{\left(\sum_j u_j\right)^2}{d_Ad_B}-1. \tag{61}$$

Furthermore, the quantum operation $\mathcal{F}$ defined by

$$\mathcal{F}(\rho)\equiv\frac{\sum_{k\neq l} u_ku_l(A_k\otimes B_l)\rho(A_k^\dagger\otimes B_l^\dagger)}{\sum_{k\neq l} u_ku_l} \tag{62}$$

is an instance of the type of noise against which U is least robust. That is, $\mathcal{F}$ is trace preserving, and $\mathcal{U}+R(\rho(U))\mathcal{F}$ is separable. Note that $\mathcal{F}$ is manifestly separable.

The application of the theorem of most interest for us is the CNOT. It is not necessarily obvious that the CNOT has a Schmidt decomposition with the properties required by the theorem; after all, we earlier wrote the Schmidt decomposition for the CNOT as $\sqrt{2}|0\rangle\langle 0|\otimes I/\sqrt{2}+\sqrt{2}|1\rangle\langle 1|\otimes X/\sqrt{2}$, and this is not of the required form. However, while the Schmidt coefficients are unique, the operators appearing in the Schmidt decomposition may not be unique, when two or more of the coefficients are degenerate. It turns out that there is an alternative form of the Schmidt decomposition for the CNOT which is of the right form. This follows, for example, from Proposition 4 of Ref. [13], and can also be verified directly, with a little algebra. The explicit form is not particularly illuminating, so we omit it here.

Equation (61) now tells us that $R_g(\text{CNOT})=R_s(\text{CNOT})=1$. Comparing with the random robustness, $R_r(\text{CNOT})=8$, we see that the CNOT is substantially less robust against general noise than depolarizing noise; the worst-case noise is easily calculated from Eq. (62).

Proof. We already know that $R(\rho(U))\leqslant R_g(U)\leqslant R_s(U)$, so it suffices to prove that $R_s(U)\leqslant R(\rho(U))$. To prove this, we use the construction of Vidal and Tarrach, Eq. (36), to see that $\rho(U)+R(\rho(U))\sigma_{\rho(U)}$ is separable, where

$$\sigma(\rho(U))=\frac{1}{R(\rho(U))}\sum_{k\neq l} u_ku_k|k\rangle\langle k|\otimes|l\rangle\langle l|, \tag{63}$$

$$|k\rangle\equiv(I_{R_A}\otimes A_k)|\alpha\rangle,\quad |l\rangle\equiv(B_l\otimes I_{R_B})|\beta\rangle. \tag{64}$$

Using the fact that the A_k and B_l are proportional to unitary operations, a calculation shows that $\text{tr}_{AB}(\sigma_{\rho(U)})$ is a completely mixed, separable state. By Theorem 3 we conclude that there exists a trace-preserving, separable quantum operation $\mathcal{F}$ such that $\rho(\mathcal{F})=\sigma_{\rho(U)}$. [Another way of seeing this is to directly verify that $\mathcal{F}$ as defined by Eq. (62) satisfies $\rho(\mathcal{F})=\sigma_{\rho(U)}$.] Thus

$$\rho(U)+R(\rho(U))\rho(\mathcal{F})=\rho[\mathcal{U}+R(\rho(U))\mathcal{F}] \tag{65}$$

is separable, whence $\mathcal{U}+R(\rho(U))\mathcal{F}$ is separable. It follows from the definition that $R_s(U)\leqslant R(\rho(U))$, which completes the proof. ■

It is not difficult to verify that $R_s(\mathcal{E})$ and $R_g(\mathcal{E})$ satisfy properties similar to those satisfied by the random robustness, and thus can be regarded as measures of dynamic strength. The major difference is continuity: the lack of an explicit formula for the separable and global robustness has prevented us from obtaining quantitative continuity statements like those we obtained for the random robustness, although it is still not difficult to argue that both quantities are continuous.

2. General robustness and the threshold for quantum computation

As with the random robustness, we can use R_s and R_g to obtain bounds on the threshold for quantum computation. The method for obtaining a bound is similar. Suppose we have a quantum computer capable of arbitrary single-qubit gates and a single two-qubit gate, U. Then there exists $\mathcal{E}$ such that $U+R_g(U)\mathcal{E}$ is separable. Suppose that whenever we apply U, there is probability p that instead $\mathcal{E}$ occurs. If $p\geqslant R_g(U)/[1+R_g(U)]$ then this set of operations can be efficiently simulated classically, and we conclude that $p_{\text{th}}\leqslant R_g(U)/[1+R_g(U)]$. Similar remarks apply for $R_s(U)$, only the noise in that case is restricted to be separable.

Note that both these noise models are more adversarial, or pessimistic, than the noise model in Sec. V A 2, and the threshold bounds are thus tighter. In particular, these models allow correlated two-qubit noise, while the earlier model assumes independent noise on the two qubits. Which model is more realistic obviously depends upon which system a gate is implemented in. However, we do expect correlated errors similar to those in the present models to play a role in many real-world two-qubit gates, due to interactions occurring during the gate.

The bounds obtained using R_s and R_g are, in general, tighter than those obtained by studying R_r, as in Sec. V A 2. However, without specific formulas for $R_s(U)$ and $R_g(U)$ it is difficult to derive bounds on the threshold without resorting to numerical calculation. Fortunately, if the only entangling gate available is of the form described by Theorem 5, then we can calculate the optimal noise process, and the corresponding robustness $R_s(U)=R_g(U)=R(\rho(U))$. For example, for the CNOT, this gives the bound $p_{\text{th}}\leqslant 1/2$ on the threshold, since $R_g(\text{CNOT})=R_s(\text{CNOT})=1$.

An alternative approach to proving bounds on the threshold is provided by the following general bound on the robustness. The bound says, roughly, that if all two-qubit unitary gates are available, then without loss of generality the worst noise is depolarizing noise.

Theorem 6. For any trace-preserving quantum operation $\mathcal{E}$, $\max_U R(U\|\mathcal{E})\geqslant \max_U R_r(U)=d_A^2 d_B^2/2$.

As a corollary, if all one- and two-qubit gates are available, but we do not make any assumptions about the noise, the worst possible noise will be depolarizing noise, $\mathcal{D}\otimes\mathcal{D}$, and the corresponding bound on the threshold is $p_{\text{th}}\leqslant 8/9$.

Proof. Completely depolarizing noise can be represented as applying a random unitary operation V_k with probability p_k, where each $V_k=V_k^A\otimes V_k^B$ is a product of local gates and $\Sigma p_k V_k \rho V_k^\dagger \propto I$ for any density operator ρ. Thus $\mathcal{D}=\Sigma p_k \mathcal{V}_k$ where $\mathcal{V}_k(\rho)=V_k\rho V_k^\dagger$.

Since R is convex in the second argument and $\mathcal{D}\circ\mathcal{E}=\mathcal{D}$ for any operation $\mathcal{E}$, it follows that for any unitary U,

$$\begin{aligned} R(U\|\mathcal{D}) &= R(U\|\mathcal{D}\circ\mathcal{E}) \\ &= R\left(U\Big\|\sum_k p_k \mathcal{V}_k\mathcal{E}\right) \\ &\leqslant \sum_k p_k R(U\|\mathcal{V}_k\circ\mathcal{E}) = \sum_k p_k R(V_k^\dagger U\|\mathcal{E}), \end{aligned} \quad (66)$$

where the last equality follows from the fact that V_k is a product of local gates.

Let $R_0=\max_U R(U\|\mathcal{E})=\max_U R(V_k^\dagger U\|\mathcal{E})$. Then Eq. (66) implies that $R(U\|\mathcal{D})\leqslant \Sigma_k p_k R_0 = R_0$ for any U, so $\max_U R(U\|\mathcal{D})\leqslant \max_U R(U\|\mathcal{E})$ for any trace-preserving operation $\mathcal{E}$.

We conclude with a result tying our techniques more closely to the physical situation. Suppose we are attempting to perform quantum computation in the laboratory using a noisy gate $\mathcal{E}$ meant to approximate an ideal, unitary quantum gate U. U is known exactly, for it is a theoretical construct, and $\mathcal{E}$ has been experimentally determined using quantum process tomography [31,32]. For what values of p is it possible to find a trace-preserving quantum operation $\mathcal{G}$, such that $\mathcal{E}=p\mathcal{U}+(1-p)\mathcal{G}$? The answer to a generalization of this question is provided by the following theorem.

Theorem 7. Let $\mathcal{E}$ and $\mathcal{F}$ be trace-preserving quantum operations, and let $0\leqslant p\leqslant 1$. Then there exists a trace-preserving quantum operation $\mathcal{G}$ such that $\mathcal{E}=p\mathcal{F}+(1-p)\mathcal{G}$ if and only if the support of $\rho(\mathcal{F})$ is contained within the support of $\rho(\mathcal{E})$, and

$$p\leqslant \frac{1}{\lambda_1(\rho(\mathcal{E})^{-1}\rho(\mathcal{F}))}, \quad (67)$$

where $\lambda_1(\cdot)$ denotes the largest eigenvalue, and the inverse is a generalized inverse if $\rho(\mathcal{E})$ is not invertible.

The theorem is a straightforward consequence of the following theorem, and the Jamiolkowski [33] isomorphism between states and operations.

Theorem 8. Let ρ and σ be density matrices, and let $0\leqslant p\leqslant 1$. Then there exists a density matrix τ such that $\rho=p\sigma+(1-p)\tau$ if and only if the support of σ is contained within the support of ρ, and $p\leqslant 1/\lambda_1(\rho^{-1}\sigma)$, where $\lambda_1(\cdot)$ denotes the largest eigenvalue, and the inverse is a generalized inverse if ρ is not invertible.

Proof. Suppose $\rho=p\sigma+(1-p)\tau$. Since σ and τ are positive, it is clear that the support of both σ and τ must be contained within the support of ρ. It will be convenient to work in the vector space corresponding to the support of ρ, so ρ is invertible. Since τ is positive, we have $\rho\geqslant p\sigma$, as an operator inequality. Premultiplying and postmultiplying by $\rho^{-1/2}$ gives $I\geqslant p\rho^{-1/2}\sigma\rho^{-1/2}$. Comparing the largest eigenvalues of these two operators gives the desired inequality. The converse is proved by running the argument backward. ■

VI. CONCLUSION

We have defined several measures of the robustness of quantum gates against the effects of noise, and used these measures to prove that certain noisy quantum gate sets can be efficiently simulated on a classical computer, even if the methods of fault-tolerant computation are used. Our results imply an upper bound on the threshold for quantum computation, $p_{\text{th}}\leqslant 0.5$. A key component in proving these results was a proof that any quantum computation involving only separable quantum gates can be efficiently simulated on a classical computer. Furthermore, we have studied gate robustness as a measure of the strength of a quantum operation, considered as a physical resource, and shown that robustness satisfies many properties such a strength measure is expected to have.

Note added. Recently, we learnt that Eq. (37) was proved independently by Steiner [34], and, in the case of qubits, by Verstraete and Verschelde [35].

ACKNOWLEDGMENTS

Thanks to the quantum information theory group at the University of Queensland for valuable discussions. M.A.N. thanks Michael Hall for enlightening discussions which led to the proof of Theorem 8. A.W.H. thanks the Center for Quantum Computer Technology at the University of Queensland for its hospitality and acknowledges support from the NSA and ARDA under Army Research Office Contract No. DAAD19-01-1-06.

[1] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[2] A. Barenco, C.H. Bennett, R. Cleve, D.P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J.A. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).
[3] R. Landauer, in *Proceedings of the Drexel-4 Symposium on Quantum Nonintegrability—Quantum Classical Correspondence*, edited by D.H. Feng and B.-L. Hu (International Press, Boston, 1995).
[4] P.W. Shor, Phys. Rev. A **52**, 2493 (1995).
[5] A.M. Steane, Proc. R. Soc. London, Ser. A **452**, 2551 (1996).
[6] D. Aharonov and M. Ben-Or, e-print quant-ph/9906129.
[7] D. Gottesman, e-print quant-ph/9705052.
[8] A.Y. Kitaev, in *Quantum Communication, Computing, and Measurement*, edited by A.S.H.O. Hirota and C.M. Caves (Plenum Press, New York, 1997), pp. 181–188.
[9] E. Knill, R. Laflamme, and W.H. Zurek, Proc. R. Soc. London, Ser. A **454**, 365 (1998).
[10] J. Preskill, Proc. R. Soc., Math. Physic. Eng. Sci. **454**, 385 (1998).
[11] D. Aharonov and M. Ben-Or, in Proceedings of the 37th Annual Symposium on the Foundations of Computer Science (FOCS) 1996 (unpublished), pp. 46–55.
[12] G. Vidal and R. Tarrach, Phys. Rev. A **59**, 141 (1999).
[13] M.A. Nielsen, C.M. Dawson, J.L. Dodd, A. Gilchrist, D. Mortimer, T.J. Osborne, M.J. Bremner, A.W. Harrow, and A. Hines, e-print quant-ph/0208077.
[14] M.A. Nielsen, e-print quant-ph/0011036.
[15] H. Barnum, e-print quant-ph/9910072.
[16] G. Vidal, D. Jonathan, and M.A. Nielsen, Phys. Rev. A **62**, 012304 (2000).
[17] V. Vedral, M.B. Plenio, M.A. Rippin, and P.L. Knight, Phys. Rev. Lett. **78**, 2275 (1997).
[18] H. Barnum, M.A. Nielsen, and B.W. Schumacher, Phys. Rev. A **57**, 4153 (1998).
[19] C.H. Bennett, D.P. DiVincenzo, C.A. Fuchs, T. Mor, E. Rains, P.W. Shor, J.A. Smolin, and W.K. Wootters, Phys. Rev. A **59**, 1070 (1999).
[20] J.I. Cirac, W. Dür, B. Kraus, and M. Lewenstein, Phys. Rev. Lett. **86**, 544 (2001).
[21] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).
[22] A.C. Yao, in Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science, 1993 (unpublished), pp. 352–361.
[23] R.T. Rockafeller, *Convex Analysis* (Princeton University Press, Princeton, 1970).
[24] R. Jozsa and N. Linden, e-print quant-ph/0201143.
[25] S.L. Braunstein, C.M. Caves, R. Jozsa, N. Linden, S. Popescu, and R. Schack, Phys. Rev. Lett. **83**, 1054 (1999).
[26] N.C. Menicucci and C.M. Caves, Phys. Rev. Lett. **88**, 167901 (2002).
[27] N. Linden and S. Popescu, Phys. Rev. Lett. **87**, 047901 (2001).
[28] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
[29] J.L. Brylinski and R. Brylinski, in *Mathematics of Quantum Computation*, edited by R.K. Brylinski and G. Chen (Chapman and Hall, London/CRC Press, Boca Raton, 2002), Chap. II.
[30] M.J. Bremner, C.M. Dawson, J.L. Dodd, A. Gilchrist, A.W. Harrow, D. Mortimer, M.A. Nielsen, and T.J. Osborne, e-print quant-ph/0207072.
[31] I.L. Chuang and M.A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).
[32] J.F. Poyatos, J.I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).
[33] A. Jamiolkowski, Rev. Mod. Phys. **3**, 275 (1972).
[34] M. Steiner, e-print quant-ph/0304009.
[35] F. Verstraete and H. Verschelde, Phys. Rev. Lett. **90**, 097901 (2003).