Scalable quantum computing in the presence of large detected-error rates

E. Knill*
National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305, USA
(Received 23 December 2003; revised manuscript received 9 December 2004; published 14 April 2005)

The theoretically tolerable erasure error rate for scalable quantum computing is shown to be well above 0.1, given standard scalability assumptions. This bound is obtained by implementing computations with generic stabilizer code teleportation steps that combine the necessary operations with error correction. An interesting consequence of the technique is that the only errors that affect the maximum tolerable error rate are storage and Bell measurement errors. If storage errors are negligible, then any detected Bell measurement error below 1/2 is permissible. For practical computation with high detected error rates, the implementation overheads need to be improved.

DOI: 10.1103/PhysRevA.71.042322 PACS number(s): 03.67.Lx, 03.67.Pp, 89.70.+c

I. INTRODUCTION

One of the most significant obstacles to scalable quantum computing is physical noise that can quickly destroy the information contained in the computational state. It is now known that, provided the physical noise is sufficiently low and local in space and time, scalable quantum computing is possible by means of fault-tolerant encodings of quantum information [1–9]. Therefore, in studying a proposed physical implementation, a key question is whether the noise in the implementation is sufficiently low for scalability to be possible in principle. More importantly, is it feasible in practice? Answers to these questions depend significantly on the specific noise in the implementation, as well as on the way in which the quantum operations necessary for computing are realized. Nevertheless, the current consensus is that the error should be below an error rate of 10^{-4} per operation [10,11]. There have been numerous suggestions that 10^{-4} is a pessimistic estimate of the error threshold (below which scalable quantum computing is possible) and certainly does not apply uniformly to all needed quantum operations [12–15]. Recent research [15–18] makes it clear that thresholds are at least an order of magnitude higher under reasonable assumptions.

If the nature of the errors is constrained, then the maximum tolerable error rate can be much higher. A notable example of this is in efficient linear optics quantum computing [19] where, by design, the errors are dominated by unintentional but detected measurements of σ_z. For this error model, any error rate below 0.5 is tolerable [20–22]. That the tolerable error rate is so high is due to the great advantages of being able to detect errors before attempting correction. Here it is shown that these advantages also apply to the erasure error model. In this model, errors are detected but otherwise unknown. Another way of thinking about this error model is to imagine that the only error is loss of qubits, and whether or not a qubit is present can easily be determined without affecting the qubit's state. The main result of this paper is that the maximum tolerable erasure error rate can be as high as 0.292 per operation, given otherwise standard (though not necessarily practical) scalability assumptions. Interestingly, the only operations that have to meet this error probability are those needed for Bell measurements and for storing a qubit for the duration of one Bell measurement. Other operations need have only a nonzero probability of success. If qubit memory is perfect, any probability of erasure below 1/2 during a Bell measurement can be tolerated. Although the techniques presented here are theoretically efficient, the resource requirements at high detected error rates are impractical. Further work is required to approach feasibility in practice.

To establish a bound on threshold erasure error rates, the techniques used in Ref. [19] are adapted to the general setting. Basically, all computational operations with error correction are combined into a single, very flexible teleportation step. The ideas that make this possible can be found in Refs. [15,23–26]. Which operation is applied and the means for error correction are determined by which state is prepared for use in the teleportation step. This is where error detection can be used to advantage: The prepared state can be guaranteed to be error-free at the time it is brought into the computation. Specifically, one can configure the computation so that all states that may be needed are manufactured in large quantities at a state factory, with any states for which errors are detected being discarded before use. Although this is potentially inefficient, it means that for theoretical scalability, only errors in the implementation of the teleportation process itself are relevant.

II. ERROR MODELS

Threshold error rates depend in subtle ways on the details of the error model adopted and how it is tied to the universal gate set used for computing. For pure detected-error models, each gate or other operation is error free if no error is detected. The model focused on in this paper is the erasure error model, in which a detected error implies complete loss of the state of the qubits involved. Note that detected errors can be converted to erasure errors by depolarizing qubits on which an error has been detected. The error rate is determined by the probabilities of detected error for the various operations. For simplicity, all operations are assumed to take

*Electronic address: knill@boulder.nist.gov

the same amount of time ("one time-step") and are synchronously clocked on the qubits. It is assumed that classical computations are instantaneous and error free, and that operations can be applied to any pair of qubits without communication delays. In other words, there is no communication or classical computation latency. Classical computation latency can be a problem if the algorithms for calculating the necessary error-correction steps depending on measurement outcomes are too complex. Quantum communication latency occurs when distant qubits need to be coupled. If classical computation latency is negligible, quantum communication latency can be significantly reduced by using teleportation methods [27].

It is assumed that the probabilities of detected error are strictly less than 1. The following detected-error probabilities will be used. (1) e_m, the probability of detected error for the "memory" operation, which involves storing the state of a qubit for one time-step. (2) e_b, the probability of detected error in implementing a Bell measurement on two qubits. Here, a Bell measurement is considered to be an elementary operation and assumed to require one time-step. Interestingly, probabilities of detected errors for other operations do not affect the threshold if the methods described below are used. It is assumed that errors are independent between different operations.

III. STABILIZER CODES

For an introduction to the theory of stabilizer codes, see Ref. [28]. Here we give a brief review to establish the basics. A stabilizer codes for qubits is defined as a common eigenspace of a set of commuting products of Pauli operators. Let n be the length of the code, that is, the number of qubits used. It is convenient to specify a product of Pauli operators (Pauli product) by a pair of length-n binary (row) vectors $\mathbf{s}=(\mathbf{a},\mathbf{b})$, $\mathbf{a}=(a_i)_{i=1}^n$, $\mathbf{b}=(b_i)_{i=1}^n$. For example, $(a_1,a_2,a_3)=(1,0,1)$ and $(b_1,b_2,b_3)=(0,1,1)$. For brevity, one can omit commas and use square brackets as follows: $(a_1,a_2,a_3)=[a_1a_2a_3]=(1,0,1)=[101]$. The product of Pauli operators associated with $\mathbf{s}$ is given by

$$P(\mathbf{s}) = P(\mathbf{a},\mathbf{b}) = \prod_{j=1}^{n} \begin{cases} \sigma_x^{(j)} & \text{if } a_j = 1 \text{ and } b_j = 0, \\ \sigma_z^{(j)} & \text{if } a_j = 0 \text{ and } b_j = 1, \\ \sigma_y^{(j)} = i\sigma_x^{(j)}\sigma_z^{(j)} & \text{if } a_j = 1 \text{ and } b_j = 1, \end{cases} \tag{1}$$

where $\sigma_x,\sigma_y,\sigma_z$ are the Pauli matrices and parenthesized superscripts denote the label of the qubit on which the given operator acts. Let $\mathbf{t}=(\mathbf{c},\mathbf{d})$ be another pair of length-n binary vectors. Computations involving vectors and matrices are performed modulo 2. $P(\mathbf{s})$ commutes with $P(\mathbf{t})$ if $\mathbf{a}\cdot\mathbf{d}^T-\mathbf{b}\cdot\mathbf{c}^T=0$, and anticommutes otherwise. Explicitly, $P(\mathbf{s})P(\mathbf{t})=(-1)^{\mathbf{a}\cdot\mathbf{d}^T-\mathbf{b}\cdot\mathbf{c}^T}P(\mathbf{t})P(\mathbf{s})$. One can consider pairs of n-dimensional row vectors such as $\mathbf{s}$ as $2n$-dimensional row vectors. To maintain the association with qubit positions, it is convenient to merge the two n-dimensional vectors. That is, by definition, $\mathbf{s}=(\mathbf{a},\mathbf{b})=[a_1b_1a_2b_2\cdots]$. With the example above, one can write $\mathbf{s}=[100111]$. To make the association with Pauli matrices, the abbreviations $I=00$, $X=10$, $Z=01$, and $Y=11$ are convenient. Thus $[100111]=[XZY]$.

Let $\mathbb{S}$ be the $2n\times 2n$ block-diagonal matrix

$$\mathbb{S} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ -1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & -1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{2}$$

Then $P(\mathbf{s})$ commutes with $P(\mathbf{t})$ if $\mathbf{s}\mathbb{S}\mathbf{t}^T=0$ and anticommutes otherwise. Specifically, $P(\mathbf{s})P(\mathbf{t})=(-1)^{\mathbf{s}\mathbb{S}\mathbf{t}^T}P(\mathbf{t})P(\mathbf{s})$.

Consider l independent (modulo 2) $2n$-dimensional binary vectors $\mathbf{s}_i$ describing Pauli products as explained above. Let Q be the *check matrix* whose rows are the $\mathbf{s}_i$. If the $P(\mathbf{s}_i)$ commute, then the state space of n qubits decomposes into 2^l disjoint common eigenspaces of the $P(\mathbf{s}_i)$, each of dimension 2^{n-l}. The eigenspaces are characterized by their syndromes, that is, by the eigenvalues of the $P(\mathbf{s}_i)$. Each syndrome is described by an l-dimensional binary (row) vector $\mathbf{e}$ and relates to the eigenvalues as follows: If $|\psi\rangle$ is in the eigenspace with syndrome $\mathbf{e}=(e_i)_{i=1}^l$, then $P(\mathbf{s}_i)|\psi\rangle=(-1)^{e_i}|\psi\rangle$. The projection operator onto the eigenspace with syndrome $\mathbf{e}$ is given by

$$\Pi(Q,\mathbf{e}) = \prod_i \frac{1}{2}[I + (-1)^{e_i}P(\mathbf{s}_i)]. \tag{3}$$

If it is necessary to emphasize the dependence of the syndrome on Q, it will be referred to as the Q syndrome. The eigenspaces are the stabilizer codes associated with Q. For $\mathbf{s}$ in the row span of Q, $P(\mathbf{s})$ stabilizes the states of each of Q's stabilizer codes up to a phase. Such Pauli products form the stabilizer of the codes. Write $\Pi(Q)=\Pi(Q,\mathbf{0})$ and consider this to be the fundamental stabilizer code associated with Q. The word "fundamental" will be omitted whenever possible. Furthermore, $\Pi(Q)$ is used to refer both to the projection operator and to the code as a subspace: The intended meaning will be clear from the context.

It is important to understand the effects of Pauli products on states in a stabilizer code. One can verify that $P(\mathbf{s})\Pi(Q,\mathbf{e})P(\mathbf{s})^\dagger=\Pi(Q,\mathbf{e}')$, where $\mathbf{e}'=\mathbf{e}+\mathbf{s}\mathbb{S}Q^T$. This implies that if $|\psi\rangle$ has syndrome $\mathbf{e}$, so that $\Pi(Q,\mathbf{e})|\psi\rangle=|\psi\rangle$, then $P(\mathbf{s})|\psi\rangle$ has syndrome $\mathbf{e}+\mathbf{s}\mathbb{S}Q^T$. To see this, compute $P(\mathbf{s})|\psi\rangle=P(\mathbf{s})\Pi(Q,\mathbf{e})|\psi\rangle=\Pi(Q,\mathbf{e}+\mathbf{s}\mathbb{S}Q^T)P(\mathbf{s})|\psi\rangle$.

Let $C=C(Q)$ be the row span of Q. C is a classical binary code. If Q' has the same row span as Q, then $C(Q')=C(Q)$, and the set of stabilizer codes associated with Q' is the same as that associated with Q. For understanding the error-correcting properties of stabilizer codes, one has to look at $C^\perp$, the set of vectors $\mathbf{x}$ such that $\mathbf{x}\mathbb{S}Q^T=0$ or, equivalently, such that $P(\mathbf{x})$ commutes with all of the $P(\mathbf{s}_i)$. For $\mathbf{t}\in C^\perp$ but not in C, $P(\mathbf{t})$ preserves each stabilizer code associated with Q but acts nontrivially in each code. Consequently, the quantum minimum distance of these codes is the minimum distance of the set $C^\perp\setminus C$. Here, minimum distance is defined as the weight of the smallest-weight (nonzero) vector in $C^{\perp}\backslash C$. The weight of $\mathbf{x}$ is the number of qubits on which $P(\mathbf{x})$ acts nontrivially.

When working with stabilizer codes and syndrome measurements, it is helpful to be able to determine the new stabilizer of a state after making a syndrome measurement for a different code. Let Q be as above. Suppose that the initial state $|\psi\rangle$ is an arbitrary state of $\Pi(Q,\mathbf{e})$ and that one measures the R syndrome with outcome $\mathbf{f}$. What Pauli products are guaranteed to stabilize the resulting state $|\phi\rangle$ up to a phase? $P(\mathbf{r})$ stabilizes $|\phi\rangle$ if $\mathbf{r}$ is in $C(R)$ or in $C(Q)\cap C(R)^{\perp}$. The latter set consists of the Pauli operators that are guaranteed to stabilize the initial state and commute with the measurement. In general, the only Pauli products guaranteed to stabilize $|\phi\rangle$ are products of the above. One can construct an independent set of such products from Q and R by the usual linear-algebra methods modulo 2.

It is not the case that minimum distance completely determines whether $\Pi(Q)$ is a stabilizer code with good error-correction properties for typical error models. That is, provided that the number of low-weight elements of $C^{\perp}\backslash C$ is sufficiently small, it is still possible to correct most errors. Suppose that the $(n-l)$-qubit state $|\psi\rangle$ is encoded as $|\psi\rangle_{\mathsf{L}}$ in $\Pi(Q)$. For any error model, the effect of the errors on $|\psi\rangle_{\mathsf{L}}$ can be thought of as a probabilistic mixture of the states $A_k|\psi\rangle_{\mathsf{L}}$, where (A_k) are the operators in the operator sum representation of the errors and satisfy $\sum_k A_k^{\dagger}A_k=I$. The probability of $A_k|\psi\rangle_{\mathsf{L}}$ is ${}^{\mathsf{L}}\langle\psi|A_k^{\dagger}A_k|\psi\rangle_{\mathsf{L}}$. Because Pauli products form a complete operator basis, $A_k=\sum_{\mathbf{s}}\alpha_{k\mathbf{s}}P(\mathbf{s})$. To correct the errors one can measure the Q syndrome of the noisy state. Suppose that the measured syndrome is $\mathbf{e}$. Then the state $A_k|\psi\rangle_{\mathsf{L}}$ is projected to $\sum_{\mathbf{s}:\mathbf{s}\mathbb{S}Q^T=\mathbf{e}}\alpha_{k\mathbf{s}}P(\mathbf{s})|\psi\rangle_{\mathsf{L}}$. The sum is over a set $C^{\perp}+\mathbf{s}_0$. A good code for the error model has the property that, with high probability, the dominant amplitudes among the $\alpha_{k\mathbf{s}}$ with $\mathbf{s}\in C^{\perp}+\mathbf{s}_0$ satisfy the condition that $\mathbf{s}$ is in the same set $C+\mathbf{s}'\subseteq C^{\perp}+\mathbf{s}_0$, independent of which A_k occurred. If that is true, then a decoding algorithm can determine the dominant amplitude's coset $C+\mathbf{s}'$ and apply $P(\mathbf{s}')^{\dagger}$ to restore $|\psi\rangle_{\mathsf{L}}$. A practical code also has the property that there is an efficient decoding algorithm that has a high probability of successfully inferring $C+\mathbf{s}'$.

The discussion of the previous paragraph assumes that nothing is known about the error locations. Suppose that it is known that the errors occurred on a given set S of m qubits. If the errors are erasures, without loss of generality, reset the erased qubit to 0 (replacing it with a fresh qubit if necessary). Suppose that after this, the measured syndrome is $\mathbf{e}$. The possible Pauli products appearing in the new state $|\phi\rangle=\sum_{\mathbf{s}:\mathbf{s}\mathbb{S}Q^T=\mathbf{e}}\alpha_{k\mathbf{s}}P(\mathbf{s})|\psi\rangle_{\mathsf{L}}$ satisfy the condition that $\mathbf{s}$ has nonzero entries only for qubits in S and $\mathbf{s}\in C^{\perp}+\mathbf{s}_0$ for some $\mathbf{s}_0$. Suppose that $C^{\perp}\backslash C$ contains no $\mathbf{s}$ with nonzero entries only for qubits in S. Then all $\mathbf{s}$ appearing in the sum for $|\phi\rangle$ are in the same set $C+\mathbf{s}'$ for some $\mathbf{s}'$. Applying $P(\mathbf{s}')$ corrects the error. Note that a suitable $\mathbf{s}'$ can be computed efficiently given $\mathbf{e}$ and S: It suffices to solve $\mathbf{s}\mathbb{S}Q^T=e$ subject to the condition that $\mathbf{s}$ is zero for positions associated with qubits outside of S. This is a set of linear equations modulo 2.

Because of the argument of the previous paragraph, an erasure code for S is defined as a code C such that $C^{\perp}\backslash C$ contains no $\mathbf{s}$ with nonzero entries only for qubits in S. It follows that a code of minimum distance d is an erasure code for all S of cardinality at most $d-1$. A useful property of erasure codes when all errors are detected is that if an error combination cannot be corrected, then this is known. This is because given S it is possible to determine whether the code is an erasure code for S. If an error combination cannot be corrected, this becomes a detected error for the encoded information. In particular, for the erasure error model, the encoded information is also subject to erasure errors, hopefully at a much lower rate.

In addition to being able to correct errors with high probability, a good stabilizer code should be able to encode a large number of qubits. For the present purposes, analysis is simplified by encoding one qubit at a time. However, efficiency can be improved substantially by encoding more and the basic techniques that are used are still applicable. Let Q be a matrix with $n-1$ rows defining a two-dimensional stabilizer code. A qubit can be encoded in a way consistent with the stabilizer formalism by choosing two row vectors $\mathbf{t}_x$ and $\mathbf{t}_z$ with the property that $\mathbf{t}_x\mathbb{S}Q^T=\mathbf{0}$, $\mathbf{t}_z\mathbb{S}Q^T=\mathbf{0}$, and $\mathbf{t}_x\mathbb{S}\mathbf{t}_z^T=1$. Then $P(\mathbf{t}_x)$ and $P(\mathbf{t}_z)$ relate to each other as X and Z and can therefore serve as encoded X and Z observables. Note that if Q is extended by $\mathbf{t}_x$, $\mathbf{t}_z$ or $\mathbf{t}_y=\mathbf{t}_x+\mathbf{t}_z$, then one-dimensional stabilizer codes are obtained whose states are encoded X, Z, and Y eigenstates.

IV. ERROR CORRECTION BY TELEPORTATION

Let Q be the $l\times 2n$ binary check matrix with entries defining a stabilizer code on n qubits for encoding $k=n-l$ qubits with good error-detecting or -correcting properties. Consider an n qubit *input block* carrying k qubits encoded in the stabilizer code for Q, where the block has been affected by errors. An effective way of detecting or correcting errors is to teleport each of the n qubits of the input block using two blocks of n qubits that form an *encoded Bell pair*. That is, both blocks have syndrome $\mathbf{0}$ with respect to Q and corresponding qubits encoded in the two blocks are in the state $(|00\rangle+|11\rangle)/\sqrt{2}$. The state of the two blocks is defined by the following preparation procedure: Start with n pairs of qubits in the standard Bell state $(|00\rangle+|11\rangle)/\sqrt{2}$. The two blocks are formed from the first and second members of each pair, respectively. Use a Q-syndrome measurement on the n second members of each pair to project them into one of the joint eigenspaces of Q. Finally, apply identical Pauli matrices to both members of pairs in such a way as to reset the syndromes to $\mathbf{0}$. To teleport, apply the usual protocol to corresponding qubits in the three blocks. In the absence of errors, this copies the encoded input state to the second block of the encoded Bell pair. The following argument shows that the errors are revealed by parities of the teleportation measurement outcomes.

The standard quantum teleportation protocol begins with an arbitrary state $|\psi\rangle_1$ in qubit 1 and the Bell state $(|00\rangle_{23}+|11\rangle_{23})/\sqrt{2}$ in qubits 2,3. The global initial state can be viewed as $|\psi\rangle$ encoded in the stabilizer code generated by IXX and IZZ, whose check matrix has rows $\mathbf{b}_1=[001010]$ and $\mathbf{b}_2=[000101]$. Let $B^{(23)}$ be the check matrix whose rows

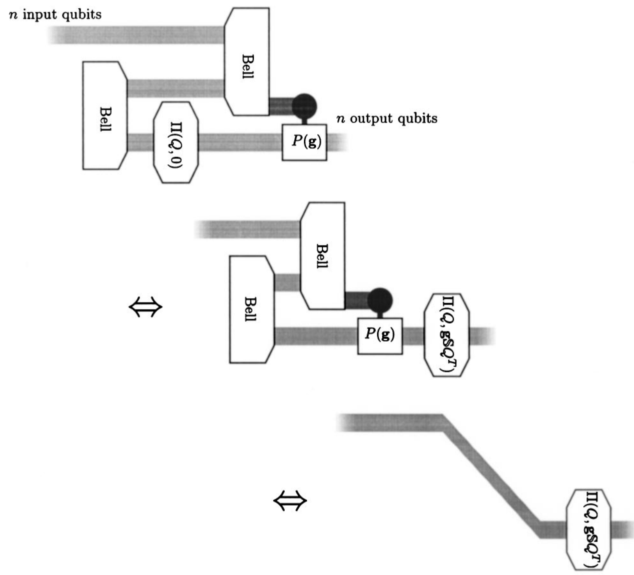

FIG. 1. Teleporting with an encoded entangled state is equivalent to a syndrome measurement. The gray lines are the time lines of blocks of n qubits. The boxes denote various operations. The Bell-state preparation on corresponding pairs of qubits in two blocks is depicted with a box angled to the right and labeled "Bell." The state used for teleportation in the top diagram is obtained after Bell-state preparation by projecting one of the blocks with $\Pi(Q,0)$. (The actual preparation procedure is different but has the same output.) Projection operators are shown with boxes angled both ways with the operator written in the box. Bell measurement of corresponding pairs of qubits in two blocks is depicted with a box angled to the left and labeled "Bell." A Bell measurement on qubits 1 and 2 may be implemented by applying a CNOT from qubit 1 to 2, performing an σ_x measurement on qubit 1 and a σ_z measurement on qubit 2. The top diagram is the actual network implemented. The other two are logically equivalent. The Bell measurement outcome **g** is correlated with the effective projection in the bottom diagram. If the input state has a particular syndrome, then only **g** for which the projection is onto the subspace with this syndrome have nonzero probabilities.

are the $\mathbf{b}_i$. To teleport, one makes a Bell-basis measurement on the first two qubits. This is equivalent to making a $B^{(12)}$-syndrome measurement, where $B^{(12)}$ has as rows [101000]=[XXI] and [010100]=[ZZI]. This is identical to $B^{(23)}$ with qubits 2,3 exchanged for qubits 1,2. Depending on the syndrome **e** that results from the measurement, one applies correcting Pauli matrices to qubit 3 to restore $|\psi\rangle$ in qubit 3.

Consider the teleportation of n qubits in a block as described above. The protocol is such that the $2n$ binary measurement outcomes linearly (with respect to computation modulo 2) determine the Pauli product correction to be applied to the second block of the encoded Bell pair. Let **g** be the binary representation of the Pauli product correction. The syndrome of the input block constrains **g** as shown in Fig. 1. The principle is as explained in Ref. [24] for unitary gates, but generalized to measurements. In this case, a stabilizer projection on the destination qubits before teleportation is equivalent to a projection after teleportation, where the syndrome associated with the projection is modified by the correction Pauli product used at the end of teleportation. The expression $\mathbf{g}\mathbb{S}Q^T$ must match the syndrome of the input block. Consequently, the syndrome of the input block can be deduced from **g**, a function of the teleportation Bell measurement. Errors can be detected or corrected accordingly. Compared to the syndrome extraction methods of Steane [15], error-correcting teleportation involves only one step instead of at least two but requires preparing more complex states.

It is necessary to consider the effects of errors in the prepared encoded Bell pair. Errors on the second block propagate forward and must be handled by future teleportations. Because of the Bell measurement, Pauli product errors on the first block have an effect equivalent to the same errors on the input block. Thus, using the inferred syndrome for detection or correction of errors deals with errors in both blocks, as long as their combination is within the capabilities of the code.

Error correction or detection by teleportation handles leakage errors in the same way as other errors. If a qubit "leaked," the outcome of its Bell measurement becomes undetermined. The Bell measurement can be filled in arbitrarily, because for the purpose of interpreting the syndrome, the effect is the same as if a Pauli error occurred depending on how the measurement result is filled in.

V. COMBINING OPERATIONS WITH ERROR CORRECTION

Operations can now be integrated into the error-correction process using the techniques described in Ref. [24]. The basic idea is to apply the desired encoded operation to the destination qubits of the entangled state (or states) to be used for teleportation. If the operation is in the Clifford group, the teleportation protocol results in the desired operation being applied, except that the Pauli products needed to correct the state are modified. (The Clifford group consists of the unitary operators that normalize the group generated by the Pauli products.) To achieve universality, an additional operation that has the property of conjugating Pauli products to elements of the Clifford group is required. One such operation is the 45° rotation $e^{-i\sigma_x\pi/8}$. Again, it is applied in encoded form to the destination qubits of the prepared state used for teleportation. After teleportation, the necessary correction may be a Clifford-group element not of the form of a Pauli product. If that is the case, this Clifford-group element is applied in the next teleportation step.

Note that for applying Clifford-group elements such as the controlled-not, the teleportation step has to act on two encoded qubits and the error-correction aspects of the teleportation step for the two qubits have to be implemented on both at the same time.

VI. MEASUREMENT

The same teleportation step used for computing can also be used for measurement, except that in this case the destination qubits are redundant. That is, if one prepares an encoded state on the first block of qubits with the encoded qubits in logical zero, the Bell measurement will reveal not just the syndrome of the code, but also the measurement outcome. Errors can be classically corrected to reveal the true measurement outcome.

VII. OMITTING POST-TELEPORTATION CORRECTIONS

The thresholds to be established require that all post-teleportation corrections are omitted. For Clifford gates implemented by teleportation, the only such correction involves appyling Pauli products to restore the syndrome. That these Pauli products can be omitted by book-keeping methods was observed by Steane [15]. This involves keeping track of a Pauli frame, which is the Pauli product that correctly restores the state to the zero syndrome code. Note that the Pauli frame is usually retroactive and does not account for newly introduced errors. See also Ref. [16]. Assume that the only non-Clifford gate is the 45° rotation given above. Whenever this is implemented by teleportation, it may be necessary to apply a 90° rotation to compensate for a possible direction error induced by the Pauli frame. Rather than applying this rotation explicitly, it can be absorbed into the next teleportation step. Since the state needed for this next step must be ready immediately after the Bell measurement of the current step, this requires anticipating possible compensations by preparing two states ahead of time, where the second serves to implement the additional 90° rotation if necessary. Note that the next gate must be a Clifford gate since there is no reason to apply a second 45° rotation. Together with the possible 90° compensating rotation, this is still a Clifford gate.

VIII. THRESHOLDS

The sequential implementation of the scheme in the context of a computation is shown in Fig. 2. In order to obtain a lower bound on the erasure error threshold and discussing the depolarizing error threshold, observe that the efficiency with which the needed states are prepared has no effect on the threshold. It is necessary only to determine for what error rates it is possible to make the error per step in the encoded (logical) qubits arbitrarily small. Consider the erasure error model. In this case, the error model for the logical qubits is also independent erasures. If the rate of logical erasures is sufficiently small, then according to the known threshold theorems, we can use the logical qubits to efficiently implement arbitrarily accurate quantum computations. Efficiency in these theorems requires only that the cost of each elementary step of the computation is bounded by a constant independent of the length of the computation. Here, this constant depends on the length of the code needed to achieve an encoded error below the general threshold, for which there are known lower bounds. The effort required to prepare the states for the teleportation steps only adds to the constant. In particular, the states can be prepared naively, by attempting to implement a quantum network that prepares them, and discarding any unsuccessful attempts. It is known that any one-qubit state encoded in a stabilizer code of length n requires at most $O(n^2)$ quantum gates [29]. Because each suc-

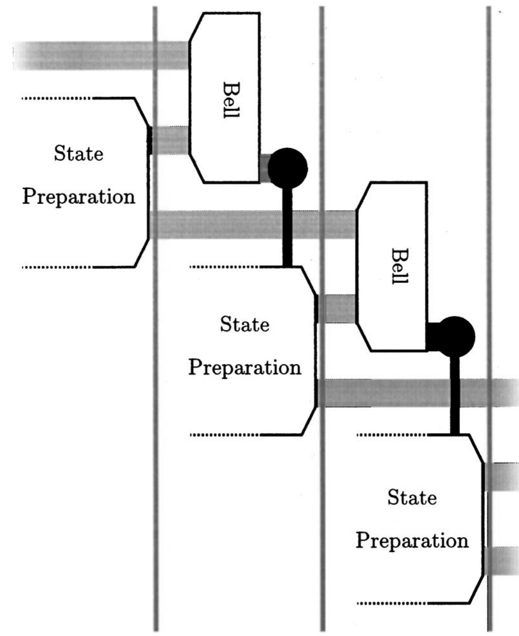

FIG. 2. Sequence of computational steps. Each step consists of a teleportation with integrated operations and error correction. The steps required for correction determine the prepared state used in the next step. That is, a state that has the appropriate operations preapplied to the destination qubit is used. Any such state is assumed to be available at the "state preparation factory" at the next step. Zero communication and classical computation delays are assumed. In particular the classical feed forward following the Bell measurement and selection of the desired prepared state take no additional time. The state preparations are implemented so that the states are ready at selection time. The prepared states have been checked for errors, with no errors detected just before they are used. Errors in the computation are therefore due only to the Bell measurement and the storage time required for the second block of each prepared state. The storage time is the duration of the Bell measurement.

ceeds only with probability $s<1$, the expected number of attempts for each state to be prepared is bounded by $e^{O(n^2)}$. Although this is superexponential, it contributes "only" a constant overhead to the implementation of each encoded operation. In conclusion, it is possible to invoke the general threshold theorems to show that state preparation overhead can be ignored for the purpose of establishing lower bounds on the threshold by the methods used here. Nevertheless, a self-contained proof not relying on the general threshold theorems for scalability is given in Sec. IX.

Suppose that we use a one-qubit erasure code for which the probability of an uncorrectable erasure is $f(e)$, given that the probability of erasure of each qubit is e. The error probability of a quantum computation using the scheme of Fig. 2 is determined by the probability that the Pauli product correction or Pauli frame change deduced from the syndrome derived from the Bell measurement outcomes is correct. Note that the inferred Pauli product correction does not strictly speaking correct the output state as it exists at the end of the Bell measurement. Rather, it is retroactive and would have restored the output state if it had been applied to the

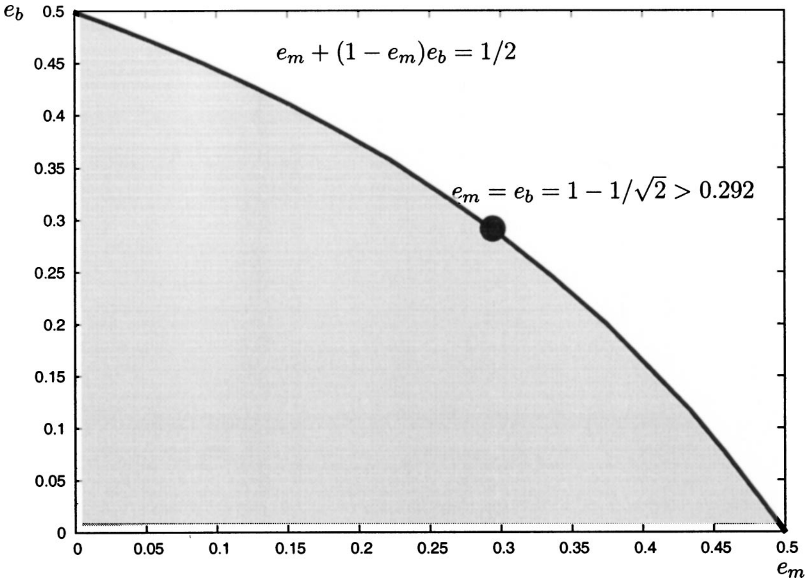

FIG. 3. Region for which scalable computing with the detected-error model is possible. Tolerable error rates are in the gray region, strictly below the upper boundary. If memory and Bell measurement error are equal, then any error rate below 0.292 is tolerable. The point on the boundary corresponding to this value is shown. If memory errors are negligible, then Bell-measurement error rates close to $1/2$ are tolerable in principle.

output state when it emerged from the state preparation (see Fig. 2). Subsequent errors accumulated while waiting for the current Bell measurement to complete are taken into account at the next teleportation. Thus, the errors that affect the syndrome determined by the Bell measurements come from the Bell measurement itself and from the (detected) storage error accumulated on the input qubits during the previous Bell measurement after they emerged from the previous teleportation's state preparation. The total probability of detected error is given by $e_m+(1-e_m)e_b$, where the first term comes from detected error in the storage period of the input qubits and the second comes from the Bell measurement, which may be applied only to qubits with no previously detected error. The probability of erasure of the encoded qubit is therefore $f[e_m+(1-e_m)e_b]$. Because the prepared states are error-free at the instant when they are used, this erasure probability also applies to each qubit independently in teleported two-qubit operations. It remains to determine for what error rates there exist one-qubit erasure stabilizer codes with arbitrarily small probability of encoded erasure. If the supremum of these error rates is $e_{\max}$, then the threshold for e_m and e_b is determined by the curve $e_{\max}=e_m+(1-e_m)e_b$. The value of $e_{\max}$ is determined in Ref. [30] and is given by $e_{\max}=1/2$. Figure 3 shows the region for e_m and e_b where scalable computing with the detected-error model is possible in principle.

IX. STATE-PREPARATION INEFFICIENCY

It is possible to prove directly that state preparation can be done with polynomial overhead instead of relying on the general threshold theorems. This is done here to make the paper more self-contained and to show that two levels of concatenation suffice for the erasure error model. Let n be the final length of the (concatenated) code for each logical qubit. The code is constructed by concatenating two erasure codes of length l_1 and l_2 with $l_1 l_2=n$ and $l_1<\sqrt{n}$. State preparation through the first level of encoding is handled by the naive method of repeated attempts. At the second level of encoding, the methods of the previous sections are used to improve the probability of successful state preparation.

To see that one level of encoding with the naive state-preparation method is insufficient, consider the following: The goal is to implement a computation of N elementary operations with polynomial overhead. In order for the computation to have a probability $1-\epsilon$ of success with logical qubits and no further error correction, the logical qubits must be subject to an error rate of ϵ/N per operation. (Erasure errors add probabilistically.) Given that e_m and e_b are in the scalability region, the error rate as a function of code length n for the best codes goes as e^{-cn} (asymptotically) for some constant c depending on e_m and e_b [31]. One should therefore choose $n>\ln(N/\epsilon)/c$. Typically, the state-preparation networks for these codes have at least $c'n^2$ gates for some constant $c'>0$. The naive state preparation therefore requires resources bounded by $e^{c''n^2}$ for some constant $c''>0$. Substituting the lower bound for n gives a superpolynomial function of N/ϵ.

To eliminate the superpolynomial overhead, choose a first level code that reduces the error rate to e^{-cl_1}. Choose a second level code that can correct any combination of at most $l_2/6$ erasures (such stabilizer codes exist by using random coding [32,33]). The concatenation of the two codes results in logical qubits with an error rate bounded by $\binom{l_2}{l_2/6}(e^{-cl_1})^{l_2/6}\leq e^{-c'n}$ for some constant $c'>0$ and sufficiently large l_1,l_2. Choose $n>\ln(N/\epsilon)/c'$. The naive state-preparation method for computations with qubits encoded in the first-level code requires an overhead of $e^{c''l_1^2}\leq e^{dn}$ for some constant d. Computations at the next level are implemented using the teleportation techniques discussed above. The probability of success of one step is at least $1-e^{-c'''l_1}$ for some constant c''' (different from c' because a step involves both storage and Bell measurements). The probability of success of a state-preparation network for states needed by the second level code is at least $1-c'l_2^2e^{-c'''l_1}>1/2$ for sufficiently large l_1. The total overhead for the concatenated state preparation is bounded by $2c'l_2^2e^{dn}<e^{d'\ln(N/\epsilon)}=\mathrm{poly}(N/\epsilon)$.

An obvious way to improve the efficiency of this state-preparation scheme is to choose the size of the first level code to better balance the overheads between the two levels. In particular, the first level code need have length only of order $\log(n)$ for the probability of success of the second-level state preparation to exceed a constant. This reduces the overhead to polylogarithmic in N/ϵ, comparable to overheads in the standard threshold theorems based on concatenation.

X. DISCUSSION

The work reported here shows that the maximum tolerable error rates depend strongly on the error model. If the errors are constrained, then they can be tolerated much better than depolarizing errors. In particular, if errors can be detected, tolerable error rates for computation are well above 0.1, depending on what are considered elementary operations and the relationship between storage errors, gate errors, and measurement times. This is strong motivation to build in error detection when engineering quantum devices and designing error-correction strategies.

The use of teleportation demonstrates yet again the now well-known versatility of this basic quantum communication protocol. It is worth noting that frequent use of teleportation in a computation implicitly solves the leakage problem. This is the problem where qubits are lost from the computation without the event being detected, either by physical loss of the underlying particles, or by the particle's state leaving the qubit-defining subspace. In every teleportation step, the destination qubits are fresh, and any previously leaked qubits contribute only to errors in the Bell measurements. These errors can be treated just like other errors.

After the work presented here was completed, the power of error-correcting teleportation and error detection has been shown to extend to general independent error models, leading to very high thresholds for fault-tolerant quantum computing. See Ref. [16].

ACKNOWLEDGMENTS

Portions of this work were done at Los Alamos National Laboratory. This work was supported by the Department of Energy (DOE Contract No. W-7405-ENG-36) and the U.S. National Security Agency. This paper is a contribution of the National Institute of Standards and Technology, an agency of the U.S. government, and is not subject to U.S. copyright.

[1] P. W. Shor, in *Proceedings of the 37th Symposium on the Foundations of Computer Science (FOCS)* (IEEE press, Los Alamitos, CA, 1996), pp. 56–65.

[2] A. Y. Kitaev, in *Quantum Communication and Computing and Measurement*, edited by O. H. *et al.* (Plenum, New York, 1997).

[3] E. Knill and R. Laflamme, Technical Report No. LAUR-96-2808, Los Alamos National Laboratory, 1996; e-print quant-ph/9608012.

[4] D. Aharonov and M. Ben-Or, in *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation (STOC)* (ACM Press, New York, 1996), pp. 176–188.

[5] D. Aharonov and M. Ben-Or, e-print quant-ph/9906129.

[6] E. Knill, R. Laflamme, and W. Zurek, Proc. R. Soc. London, Ser. A **454**, 365 (1998).

[7] E. Knill, R. Laflamme, and W. H. Zurek, Science **279**, 342 (1998).

[8] D. Gottesman, Phys. Rev. A **57**, 127 (1998).

[9] J. Preskill, Proc. R. Soc. London, Ser. A **454**, 385 (1998).

[10] D. Gottesman and J. Preskill, unpublished analysis of the accuracy threshold.

[11] T. E. Panel, Technical Report No. LAUR-02-6900, Los Alamos National Laboratory, produced for ARDA.

[12] C. Zalka, e-print quant-ph/9612028.

[13] D. Aharonov, talk at QIP, IBM, Jan 16, 2002 (unpublished).

[14] W. Dür and H.-J. Briegel, Phys. Rev. Lett. **90**, 067901 (2003).

[15] A. M. Steane, Phys. Rev. A **68**, 042322 (2003).

[16] E. Knill, e-print quant-ph/0410199.

[17] E. Knill, e-print quant-ph/0402171.

[18] B. W. Reichardt, e-print quant-ph/0406025.

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

[20] E. Knill, R. Laflamme, and G. Milburn, Technical Report No. LAUR-00-3477, Los Alamos National Laboratory; e-print quant-ph/0006120.

[21] E. Knill, Technical Report No. LAUR-01-5768, Los Alamos National Laboratory, foils of a series of lectures on efficient linear optics quantum computation.

[22] A detected σ_z measurement threshold of 0.5 is proven in Ref. [20]. It is possible to prove that the threshold is 1 using the techniques discussed in this paper. A pictorial proof is in transparencies used at several talks in 2002 and 2003 and is available by request from the author.

[23] A. Steane, Nature (London) **399**, 124 (1999).

[24] D. Gottesman and I. L. Chuang, Nature (London) **402**, 390 (1999).

[25] A. M. Steane, e-print quant-ph/0202036.

[26] A. M. Steane and B. Ibinson, e-print quant-ph/0311014.

[27] G. K. Brennen, D. Song, and C. J. Williams, Phys. Rev. A **67**, 050302 (2003).

[28] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2001).

[29] R. Cleve and D. Gottesman, Phys. Rev. A **56**, 76 (1997).

[30] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Phys. Rev. Lett. **78**, 3217 (1997).

[31] The constant c is given by $\gamma(1-2e)$ where e is the total erasure probability per qubit that needs to be corrected and γ is order unity.

[32] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[33] A. Calderbank, E. Rains, P. Shor, and N. Sloane, Phys. Rev. Lett. **78**, 405 (1997).