Concatenated coding in the presence of dephasing

Iain Gourlay and John F. Snowdon
Department of Physics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom

(Received 3 August 1999; published 18 July 2000)

We investigate the use of concatenated coding to protect against dephasing in the absence of other types of error in order to carry out large quantum computations. This analysis is based on a well-known three-bit quantum code. Fault tolerant methods for carrying out gate operations, ancilla preparation, and syndrome identification are discussed and the maximum (or threshold) error rate which can be tolerated (if quantum coherence is to be maintained for arbitrarily long computations) is estimated. The methods for performing fault tolerant gate operations are compared to the methods appropriate for the seven-bit code and it is concluded that the three-bit code is not likely to be useful for large-scale quantum computation.

PACS number(s): 03.67.Lx, 89.80.+h

I. INTRODUCTION

If quantum computing is to become a reality, the most significant obstacle from both a theoretical and experimental perspective is that of errors introduced by inaccurate gate operations and interaction with the environment. In classical computation, error correction has been studied extensively and powerful methods of dealing with errors have been devised. Ideally, it would be possible to extend these results and devise error correction schemes for quantum computation. This problem has been studied extensively and initial theoretical results were not encouraging [1,2]. However, since then significant progress has been made in the field of quantum error correction [3–8]. Recent results indicate that under certain (restrictive) conditions, arbitrarily long computations could, in principle, be performed reliably if the error probability per quantum gate is below a certain value, the threshold error rate [4–8].

Quantum error correction codes exploit the fact that if a qubit (two-level quantum system) experiences an error, the Pauli spin matrices in conjunction with the identity from a complete basis set with which to describe the error. Hence the qubit can undergo a bit-flip error ($\hat{\sigma}_x$), a sign-flip error ($\hat{\sigma}_z$), a bit-flip and a sign-flip error ($\hat{\sigma}_y$), or no error (the identity).

For many physical systems, the dominant type of error is dephasing, whereby interaction with the environment destroys quantum coherence. The environment simply refers to all external degrees of freedom with which the qubit may interact, including any applied fields which may be used to implement logic operations. For example, suppose that spin-1/2 particles are used as qubits and gate operations are performed by applying magnetic fields (note that there are many other possible representations of qubits but the same sort of argument still applies). In a fully quantum description it is clear that the qubit and the field become entangled. If the ideal output of a single bit gate for a particular input is

$$|\Psi_{\text{qubit}}\rangle=\alpha|0\rangle+\beta|1\rangle, \tag{1}$$

then the actual state of the system (qubit+environment) will be of the form

$$|\Psi_{\text{qubit+environment}}\rangle=a|0\rangle|E_0\rangle+b|1\rangle|E_1\rangle, \tag{2}$$

where $|E_0\rangle$ and $|E_1\rangle$ are not necessarily orthogonal. If $a\approx\alpha$ and $b\approx\beta$ then the reduced density operator describing the qubit state (obtained by tracing over environmental degrees of freedom) is

$$\hat{\rho}\approx|\alpha|^2|0\rangle\langle 0|+|\beta|^2|1\rangle\langle 1|+\alpha\beta^*\langle E_1|E_0\rangle|0\rangle\langle 1| \\ +\alpha^*\beta\langle E_0|E_1\rangle|1\rangle\langle 0|. \tag{3}$$

Note that if $\langle E_0|E_1\rangle$ is real then the density operator in Eq. (3) can be written (in matrix notation) as

$$\hat{\rho}\rightarrow\frac{(1+\langle E_0|E_1\rangle)}{2}\begin{pmatrix}|\alpha|^2 & \alpha\beta^*\\ \alpha^*\beta & |\beta|^2\end{pmatrix} \\ +\frac{(1-\langle E_0|E_1\rangle)}{2}\begin{pmatrix}|\alpha|^2 & -\alpha\beta^*\\ -\alpha^*\beta & |\beta|^2\end{pmatrix}. \tag{4}$$

Hence if $\langle E_0|E_1\rangle$ is real then we can describe this situation by saying that with probability $P=(1+\langle E_0|E_1\rangle)/2$ the qubit is error free after the gate operation and with probability $1-P$ the qubit suffers a sign-flip error. Dephasing can also occur on a stored qubit (i.e., when no gate operations are being performed) due to coupling to external degrees of freedom. This is the type of error we consider and the following section describes a simple and well-known error correction code (see [3] for an alternative description), which can be used when storage errors of this type dominate.

The existence of a threshold value for the maximum tolerable error depends on the use of *fault tolerant* computation and *concatenated coding*. In Sec. III, the error model used is described and all assumptions are detailed. The concepts of fault tolerance and concatenated coding are discussed in Sec. IV, where fault tolerant methods for performing gate operations for the dephasing error correction code are described in detail. In Sec. V some significant differences between fault tolerant computation using the 3-bit code and fault tolerant computation using the much studied seven-bit code [7] are discussed. This is followed by an approximation of the threshold error rate for computation using the three-bit code.

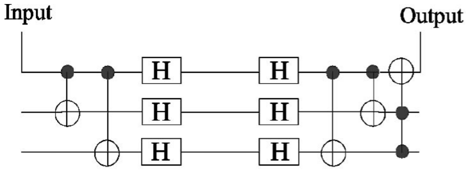

FIG. 1. Quantum error correction circuit for dephasing errors. The first two CNOT operations and the first three H operations encode the qubit. The qubit is then stored (during which time errors occur), followed by the correction and decoding part of the circuit.

II. THREE-BIT QUANTUM CODE

The previous section explained the manner in which dephasing may arise as a result of gate operations. Quantum coherence can also occur on a stored qubit due to interactions with the environment, so that a qubit beginning in the state (1) evolves as follows:

$$\hat{\rho}\approx|\alpha|^2|0\rangle\langle 0|+|\beta|^2|1\rangle\langle 1|+\alpha\beta^* f(t)|0\rangle\langle 1| + \alpha^*\beta f^*(t)|1\rangle\langle 0|, \tag{5}$$

where typically (in the Markov regime), $f(t)=e^{t/\tau_d}$ [9]. The decoherence time τ_d is a good indication of the time scale over which decoherence occurs. Here, entanglement with external degrees of freedom could result in a state of the form given in Eq. (3), so that $f(t)=\langle E_1|E_0\rangle$. The probability that a sign-flip error occurs is then

$$\varepsilon_s=\frac{1-e^{-(t_s/\tau_d)}}{2} \tag{6}$$

where t_s is the storage time.

Suppose we choose to work in the basis $\{|\overline{0}\rangle,|\overline{1}\rangle$ (from this point referred to as the x basis) where

$$|\overline{0}\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle), \quad |\overline{1}\rangle=\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle). \tag{7}$$

Transforming between the $\{|0\rangle,|1\rangle$ (the z basis) and the x basis is done by performing the single qubit Hadamard (H) gate. In the new basis, the errors are bit flips rather than sign flips and this can be exploited to devise a simple error correction procedure. We encode our qubit in the basis $\{|0_C\rangle,|1_C\rangle$, where

$$|0_C\rangle=|\overline{000}, \quad |1_C\rangle=|\overline{111}\rangle. \tag{8}$$

Suppose a single bit-flip error occurs during storage, e.g.,

$$\alpha|0_C\rangle+\beta|1_C\rangle\rightarrow\alpha|\overline{001}\rangle+\beta|\overline{110}\rangle. \tag{9}$$

Errors of this type can be detected and corrected. The complete encoding, decoding, and correction circuit is shown in Fig. 1. Note that all gates operating between bare qubits (where a bare qubit is simply an unencoded single qubit) are drawn in the z basis (this applies to all the figures). This circuit consists of three parts: the first part consists of two controlled-NOT (CNOT) gates and three single qubit Hadamard (H) gates, which encode the input state. Hence if the input state is $\alpha|0\rangle+\beta|1\rangle$, then after the encoding part of the circuit the state is $\alpha|0_C\rangle+\beta|1_C\rangle$. In the second part of the circuit, the encoded state is stored, with no gate operations being performed. It is assumed that errors can occur during this time, such that each bare qubit is flipped in the x basis with probability ε_s and remains error free with probability $1-\varepsilon_s$. The third part of the circuit is the correcting and/or decoding part. The H gates and the CNOT gates decode the state and disentangle the three qubits. If a single error has occurred in the first qubit (the initial input state) during storage, then the other two qubits are in the state $|1\rangle$ prior to the final gate in the circuit, a controlled-controlled-NOT or Toffoli gate. This gate then corrects the first qubit, leaving the output error free.

Note that the circuit in Fig. 1 fails if two or more errors occur during storage. To second order, the probability that two qubits experience errors during storage is $3\varepsilon_s^2$. Hence, if the storage time is long enough so that storage errors dominate over errors in encoding, decoding, and correction and in addition errors on different qubits are not correlated then the probability of error after correction is given by

$$P(\text{error})=3\varepsilon_s^2+O(\varepsilon_s^3), \tag{10}$$

which is a significant improvement on ε_s if $\varepsilon_s\ll 1$.

If gate errors dominate then the circuit shown in Fig. 1 is not effective in reducing errors since the information is vulnerable during encoding and decoding. Ideally the qubit would be encoded only once and thereafter all operations performed on the encoded data. Particular care must be taken to prevent errors spreading in an encoded qubit in such a way as to render the information unrecoverable, i.e., operations must be fault tolerant. This point is clarified in Sec. IV.

III. ERROR MODEL

The purpose of this section is to outline the error model and specify the assumptions made. It is assumed that single qubit gate errors can be modeled by assuming that after each gate, there is a probability ε_g that a bit-flip error occurs in the x basis. The error only affects the qubit involved in the gate. For two-bit gates, it is assumed that both qubits can be affected and all three possible errors (i.e., either of the qubits flip or both of them flip) are equally likely. Following the example of Zalka [6], the error probability is chosen so that the overall probability of a given bit experiencing an error is the same as in the case of the single qubit gate. Hence, each of the errors has a probability $\varepsilon_g/2$ of occurring. Storage errors are neglected for the purposes of calculating the threshold value, with the following justification.

(1) It is assumed that it is possible to carry out operations with an arbitrary degree of parallelism. This may be unrealistic for many possible implementations but the question of how much parallelism is possible is an important one in assessing the value of a given approach to implementing large-scale quantum computation.

(2) It is assumed that the time interval between gate operations on a given (bare) qubit can be made small compared

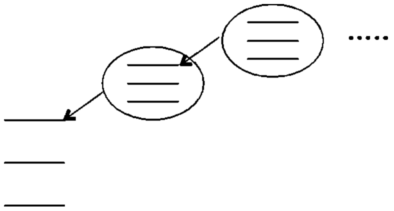

FIG. 2. Pictorial representation of the concept of code concatenation. The upper qubit in the encoded state, indicated by the first arrow, viewing left to right, is itself an encoded qubit (as are the other two qubits in the encoded state). Similarly, the upper qubit in this encoded state (indicated by the second arrow) is an encoded state. This hierarchical system can have as many levels as are necessary to protect the qubit sufficiently from unrecoverable errors.

to the gate operation time. Hence, storage errors are likely to be much smaller than gate errors.

Note that the error model describes stochastic uncorrelated errors as a result of single qubit operations, with correlated errors only occurring due to two-bit (or three-bit) gates. As pointed out by Steane [7], this may be physically unrealistic since in a real system there is likely to be some correlation of errors on different qubits. For example, in the cold trapped-ion method (initially proposed by Cirac and Zoller) [10,11], the individual qubits are coupled via the vibrational motion of the whole ion string and in this case errors are likely to be correlated. However, it is assumed here that an implementation is used which has the property that any such correlation is small enough to be negligible. In addition, the model assumes that there are no systematic errors. However, systematic errors are not likely to be such a serious problem, since their existence could be detected by test runs of a simple computation. In addition, it is likely that their effect would not accumulate in a long computation as the effect of random errors would [6]. With this error model in mind, the concepts of fault tolerance and concatenated coding are now discussed.

IV. FAULT TOLERANCE AND CONCATENATED CODING

Before discussing fault tolerance, it seems appropriate to describe concatenated coding, since this allows for a clear definition of fault tolerance. Concatenated coding, as it applies to the three-bit code, can be explained as follows. As described previously we can encode a single qubit into three qubits, so that $|0\rangle \rightarrow |0_C\rangle = |\overline{000}\rangle$ and $|1\rangle \rightarrow |\overline{111}\rangle$. This can be extended further by encoding each of the three qubits in the encoded state, so that

$$|0_C\rangle \rightarrow |0_C 0_C 0_C\rangle, \quad |1_C\rangle \rightarrow |1_C 1_C 1_C\rangle. \tag{11}$$

The concept of concatenated coding is pictorially represented in Fig. 2. We consider a bare qubit to be the zeroth level of concatenation. When the qubit is encoded (in this case into three qubits), this corresponds to the first level of concatenation. When each of the qubits in the codeword are themselves encoded into three qubits, this corresponds to the second level of concatenation. It is clear how this procedure is generalized to any level of concatenation desired.

It turns out that this is, in principle, a very powerful method of encoding in the following sense: given the error model discussed in Sec. III (or a similar error model, allowing for $\hat{\sigma}_x$ and $\hat{\sigma}_y$ errors but using a code which also corrects for these errors) and using fault tolerant methods (see below) it is possible to perform arbitrarily long computations with a bounded overall error probability *if the error probability per gate is small enough*. The basic idea is that every time the level of hierarchy is increased by one, each bare qubit in the code undergoes more operations and is consequently more likely to experience an error. On the other hand, more errors can be tolerated before an unrecoverable error occurs. If the probability of an unrecoverable error on a qubit encoded on the Lth level is δ, then *as long as errors on bare qubits belonging to the same encoded qubit are uncorrelated* (see the definition of fault tolerance below) the probability of error on the $(L+1)$th level is $O(\delta^2)$ since two of the three Lth level qubits would need to experience unrecoverable errors. If the gate error rate is too large then the additional gate operations needed result in an increase in the overall error probability as the level of hierarchy is increased. However, if the gate error probability is smaller than a particular threshold value, then the increase in error probability due to performing more gate operations is outweighed by the decrease in error probability due to the increase in level of hierarchy.

In devising methods for performing operations on encoded qubits, it is of critical importance to control the spread of errors. To clarify this point, suppose we wish to perform an encoded Hadamard operation, i.e.,

$$|0_C\rangle \overset{H}{\rightarrow} \frac{1}{\sqrt{2}}(|0_C\rangle + |1_C\rangle), \quad |1_C\rangle \overset{H}{\rightarrow} \frac{1}{\sqrt{2}}(|0_C\rangle - |1_C\rangle). \tag{12}$$

The simplest method to perform this gate is to decode the data, perform the Hadamard operation, and then re-encode the data. The problem with this method is that it is not fault tolerant, where we define fault tolerance as follows.

Suppose performing a gate operation involving a qubit (or qubits) encoded on the Lth level of concatenation introduces an unrecoverable error to the qubit (qubits) with probability $O(\varepsilon)$. The operation is fault tolerant if and only if the probability of an unrecoverable error for the same operation performed on qubits encoded on the $(L+1)$th level is no worse than $O(\varepsilon^2)$.

It is clear that performing the Hadamard operation in the manner described above does not satisfy this condition. There are two points worth making (in relation to the above example) regarding this point. Firstly, it is necessary to avoid decoding information in order to perform logic operations, since an error on a bare qubit is unrecoverable. Secondly, note that in the above example gates are operating *between qubits in the same codeword*. Generally this is a bad idea as it allows errors to spread so that an error can affect more than one qubit in the codeword.

For this error correction code, there is another source of unrecoverable errors, which is clarified by the following example of a method (which is not fault tolerant) for performing the operation P, where

$$P|0_C\rangle=|0_C\rangle, \quad P|1_C\rangle=i|1_C\rangle. \tag{13}$$

The operation can be carried out by performing the operation R on the first qubit in the encoded state, where

$$\begin{aligned} R|0\rangle&=\frac{(e^{i\pi/4}|0\rangle+e^{-i\pi/4}|1\rangle)}{\sqrt{2}}, \\ R|1\rangle&=\frac{(e^{-i\pi/4}|0\rangle+e^{i\pi/4}|1\rangle)}{\sqrt{2}}. \end{aligned} \tag{14}$$

Suppose there is a bit-flip error in the first qubit before the operation is performed. In that case, the state evolves as follows:

$$\begin{aligned} \alpha|0_C\rangle+\beta|1_C\rangle &\xrightarrow{\text{ERROR}} \alpha|\overline{100}\rangle+\beta|\overline{011}\rangle \\ &\xrightarrow{P} i\alpha|\overline{100}\rangle+\beta|\overline{011}\rangle. \end{aligned} \tag{15}$$

This error cannot be corrected, since the code is only capable of correcting bit-flip errors (in the x basis), not phase errors. This is a critical problem in constructing fault tolerant gates for the three-bit code rather than a code which corrects for arbitrary single qubit errors.

In order to be able to perform arbitrary quantum computations fault tolerantly, it is necessary to be able to perform a complete set of operations, i.e., a set of operations that can be used to efficiently approximate any quantum operation (with the qubits remaining in states spanned by the codewords). The set of operations considered here is $\{Q,P,\text{CNOT},T\}$ although there are a number of other possibilities [8] (another possibility is to replace Q with H, a more appropriate choice for the seven-bit code discussed in Sec. V), where P is described above, Q is introduced below, CNOT is the controlled-NOT operation and T is the Toffoli gate, i.e.,

$$|a\rangle|b\rangle|c\rangle \xrightarrow{T} |a\rangle|b\rangle|ab\oplus c\rangle. \tag{16}$$

In constructing the complete set, it is assumed that we can make measurements on bare qubits in the z basis, the x basis, and the y basis $[\{(|0\rangle+i|1\rangle)/\sqrt{2},(|0\rangle-i|1\rangle)/\sqrt{2}]$. In addition, errors introduced by measurements are assumed to be negligible. This only slightly affects the threshold estimation, since only a few measurements are made compared to the number of gate operations.

Having introduced the x, y, and z bases, it seems appropriate to introduce the following eigenbases (spanning the eight-dimensional Hilbert space of three qubits), which are useful in describing fault tolerant methods for the gates in the universal set introduced above. Let the Z basis consist of the vectors $|0_C\rangle,|1_C\rangle$ and the erroneous versions of these obtained when a single bit-flip (in the x basis) occurs. We refer to the erroneous versions as $|0_{C,i}\rangle$ and $|1_{C,i}\rangle$, where the i is 1, 2, or 3 and denotes which of the three qubits has experienced an error. Whenever we refer to a measurement in the Z basis, this means a measurement, represented by an operator M_Z, satisfying the following eigenvalue equations:

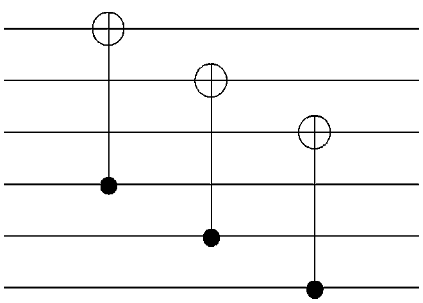

FIG. 3. A fault tolerant encoded CNOT operation. Note that the operation exploits the fact that a CNOT operation on a bare qubit is the same in the $\{|\overline{0}\rangle,|\overline{1}\rangle$ basis as in the $\{|0\rangle,|1\rangle\}$ basis, except that the control and target bits are interchanged.

$$\begin{aligned} M_Z|0_C\rangle=(-1)|0_C\rangle, &\quad M_Z|0_{C,i}\rangle=(-1)|0_{C,i}\rangle, \\ M_Z|1_C\rangle=(+1)|1_C\rangle, &\quad M_Z|1_{C,i}\rangle=(+1)|1_{C,i}\rangle. \end{aligned} \tag{17}$$

Hence if a measurement is made to distinguish between $|0_C\rangle$ and $|1_C\rangle$, the result is not affected by a single error.

The X and Y bases are defined in a similar way. The X basis consists of the vectors $(|0_C\rangle+|1_C\rangle)/\sqrt{2}=|0_X\rangle$ and $(|0_C\rangle-|1_C\rangle)/\sqrt{2}=|1_X\rangle$ and the erroneous versions of these (in the same sense as for the Z basis). A measurement in the X basis refers to a measurement represented by an operator M_X, such that $|0_X\rangle$ and the erroneous versions have eigenvalue -1, whereas $|1_X\rangle$ and the erroneous versions have an eigenvalue $+1$. Finally, the Y basis consists of the vectors $(|0_C\rangle+i|1_C\rangle)/\sqrt{2}=|0_Y\rangle$ and $(|0_C\rangle-i|1_C\rangle)/\sqrt{2}=|1_Y\rangle$ and the erroneous versions of these. A measurement in the Y basis refers to a measurement represented by an operator M_Y, such that $|0_Y\rangle$ and the erroneous versions have eigenvalue -1, whereas $|1_Y\rangle$ and the erroneous versions have eigenvalue $+1$.

In discussing encoded fault tolerant operations, the CNOT operation is considered first as it is the simplest (of the four operations in the set) to perform fault tolerantly. In performing this operation, we can exploit the following trick: If a CNOT operation is performed in the x basis between bare qubits, then this is also a CNOT operation in the x basis, the only difference being that the control and target bits are interchanged (i.e., the control bit in the z basis is the target bit in the x basis). With this in mind, the CNOT gate can be performed transversally on encoded qubits as shown in Fig. 3. Note that two independent errors must occur in order to produce an unrecoverable error when this gate is performed, since the CNOT gate does not produce phase errors (in the x basis) from bit-flip errors [unlike R, see Eq. (14) and the accompanying discussion].

The P operation can be performed fault tolerantly using the following method described by Gottesman [12]. Note that this implementation is rather complex (see Fig. 4 for the complete construction) and care must be taken to ensure fault

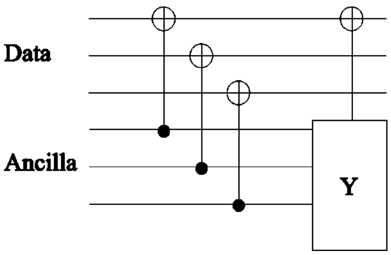

FIG. 4. Fault tolerant P gate. The box labeled Y is the circuit for measuring in the Y basis (shown in Fig. 8). A NOT gate is then performed on one of the data bits, conditioned on the result of the Y measurement. Hence if $|0_Y\rangle$ is observed, then the NOT operation is performed. Note that the ancilla is initially in the state $|0_C\rangle$.

tolerance, given the additional constraint (over codes that correct arbitrary single bit errors) that bit-flip error (in the z basis) cannot be tolerated.

First, prepare an encoded ancilla bit in the state $|0_C\rangle$. Next, perform a CNOT gate, with the data as the control bit and the ancilla as the target bit. If the initial data state is $|\Psi\rangle_{\text{data}} = \alpha|0_C\rangle + \beta|1_C\rangle$, then the overall state is now

$$\begin{aligned}|\Psi\rangle_{\text{data+anc.}} &= \alpha|0_C\rangle_{\text{data}}|0_C\rangle_{\text{anc.}} + \beta|1_C\rangle_{\text{data}}|1_C\rangle_{\text{anc.}}\\ &= P^\dagger|\Psi\rangle_{\text{data}}\frac{(|0_C\rangle_{\text{anc}} + i|1_C\rangle_{\text{anc.}})}{\sqrt{2}}\\ &\quad + P|\Psi\rangle_{\text{data}}\frac{(|0_C\rangle_{\text{anc.}} - i|1_C\rangle_{\text{anc.}})}{\sqrt{2}}, \qquad (18)\end{aligned}$$

where $P^\dagger$ is the Hermitean conjugate of P. Hence, if a fault tolerant measurement can be made in the Y basis then the P gate can also be performed fault tolerantly with the following procedure completing the operation.

We measure the ancilla bit in the Y basis, performing Z on the data if the result is $|0_Y\rangle$, where

$$Z|0_C\rangle = |0_C\rangle, \quad Z|1_C\rangle = -|1_C\rangle. \qquad (19)$$

The Z gate is carried out simply by performing a NOT operation on the first bare qubit (or any one of the qubits in the encoded bit). Note that performing this operation is fault tolerant, since although a single error before the gate results in the sign of the state $|0_C\rangle$ being flipped rather than the sign of $|1_C\rangle$, this is in fact equivalent, since an overall phase factor is physically meaningless. Hence the problem of performing P fault tolerantly is reduced to finding a way to perform fault tolerant measurements in the Y basis (see below).

The operation Q, where

$$Q|0_C\rangle = \frac{1}{\sqrt{2}}(|0_C\rangle - i|1_C\rangle), |1_C\rangle = \frac{1}{\sqrt{2}}(-i|0_C\rangle + |1_C\rangle) \qquad (20)$$

can be carried out as follows [12]. Prepare an ancilla in the state

$$|0_X\rangle = \frac{1}{\sqrt{2}}(|0_C\rangle + |1_C\rangle). \qquad (21)$$

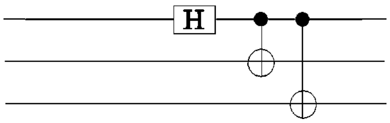

FIG. 5. Circuit for preparing the cat state $(|000\rangle + |111\rangle)/\sqrt{2}$. The initial state (prior to the circuit) is $|000\rangle$.

Next perform a CNOT gate, with the ancilla as the control bit and the data as the target. If the initial state of the data is $|\Psi\rangle_{\text{data}}$ then the new state is

$$\frac{1}{\sqrt{2}}(Q|\Psi_{\text{data}}\rangle|0_Y\rangle + Q^\dagger|\Psi_{\text{data}}\rangle|1_Y\rangle). \qquad (22)$$

Finally a fault tolerant measurement is made in the Y basis and a NOT operation performed on the data if the state $|1_Y\rangle$ is observed. All that remains to complete the universal set of gates is to describe fault tolerant methods for performing the Toffoli gate, making measurements in the X and Y bases and preparing the ancilla state $|0_X\rangle$.

We use the construction suggested by Shor [13] for the Toffoli gate. Note that it is necessary to prepare the ancilla state

$$|\text{ANC1}\rangle = \frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}|i_C\rangle|j_C\rangle|ij_C\rangle. \qquad (23)$$

Noting that

$$\frac{1}{\sqrt{8}}\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}|i_C\rangle|j_C\rangle|k_C\rangle = \frac{1}{\sqrt{2}}(|\text{ANC1}\rangle + |\text{ANC2}\rangle), \qquad (24)$$

where

$$|\text{ANC2}\rangle = \frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}|i_C\rangle|j_C\rangle|\text{NOT}(ij)_C\rangle, \qquad (25)$$

it is possible to prepare the state given by Eq. (23) by making a fault tolerant measurement in the $\{|\text{ANC1}\rangle, |\text{ANC2}\rangle$ basis and performing the NOT operation on the third encoded qubit if the observed state is $|\text{ANC2}\rangle$. This measurement is carried out as follows.

First the following cat state is prepared:

$$|\text{CAT}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle). \qquad (26)$$

The circuit required to construct this state is shown in Fig. 5. A controlled-controlled-Z gate is now performed, with the cat state and the first encoded ancilla [i_C in Eq. (24)] as the control bits and the second ancilla bit [j_C in Eq. (24)] as the target. This operation can be carried out by Hadamard transforming the first ancilla (that is a Hadamard transform on the

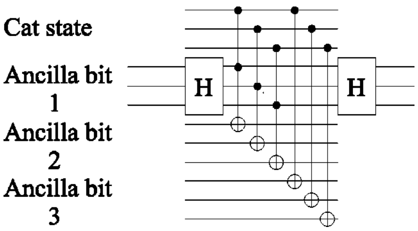

FIG. 6. The controlled-controlled-Z gate used in the preparation of $|\mathrm{ANC1}\rangle$, where the cat state and encoded ancilla bit 1 are the control bits, followed by the controlled-Z gate, where the bits in ancilla bit 3 are flipped in the $\{|0\rangle,|1\rangle\}$ basis, conditioned on the cat state qubits (in the same basis). The gates labeled H denoted the encoded H operation, where $H=PQP$.

logical bit, not a bitwise Hadamard transform on the bare qubits) then using bitwise Toffoli gates and finally Hadamard transforming the first ancilla again. Since this operation is not performed using operations from the fault tolerant set described, we must be careful to ensure that it is fault tolerant. The gate can be regarded as follows: First an encoded Hadamard transform is performed on the first ancilla, followed by an encoded CNOT gate (with the second ancilla as the control bit and the first ancilla as the target), which only operates when the cat state bits are in the state $|1\rangle$. Finally another encoded Hadamard transform is performed on the first ancilla. A single phase error on one of the ancilla bits (causing it to flip in the x basis) is then correctable in the usual way. The only other type of error that can occur is a phase error on the cat state of the form $|111\rangle \rightarrow -|111\rangle$. This type of error is recoverable if more cat states are used (see below). In that case the procedure is fault tolerant.

The next step consists of performing a controlled-Z gate with the cat state as the control and the third ancilla as the target. This gate is performed using bitwise CNOT gates. As before, a single error affecting only one of the ancilla bits is recoverable in the usual way, while the only other type of error is a cat state error of the form $|111\rangle \rightarrow -|111\rangle$. This procedure (controlled-controlled-Z then controlled-Z) is shown in Fig. 6. The evolution resulting from these operations is shown in Eq. (27),

$$
\begin{aligned}
&|\mathrm{CAT}\rangle \otimes \frac{1}{\sqrt{8}} \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} |i_C\rangle |j_C\rangle |k_C\rangle \\
&\quad \rightarrow \frac{1}{\sqrt{2}} \Big(|000\rangle \otimes \frac{1}{\sqrt{8}} \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} |i_C\rangle |j_C\rangle |k_C\rangle \\
&\qquad + |111\rangle \otimes \frac{1}{\sqrt{8}} \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} (-1)^{ij\oplus k} |i_C\rangle |j_C\rangle |k_C\rangle \Big) \\
&\quad = |\mathrm{CAT}\rangle |\mathrm{ANC1}\rangle + \frac{1}{\sqrt{2}} (|000\rangle - |111\rangle) \mathrm{ANC2}\rangle. \qquad (27)
\end{aligned}
$$

Hence by performing a measurement in the basis $\{|\mathrm{CAT}\rangle,|\mathrm{CAT}'\rangle\}$, where $|\mathrm{CAT}'\rangle = (|000\rangle - |111\rangle)/\sqrt{2}$ the ancilla state is collapsed into either $|\mathrm{ANC1}\rangle$ or $|\mathrm{ANC2}\rangle$. This measurement can be performed destructively (in that the cat state is destroyed) by performing the bitwise Hadamard transform on the individual qubits in the cat state and then measuring each of them in the z basis. If an even number of qubits is observed in the state $|1\rangle$ then the ancilla is in the state $|\mathrm{ANC1}\rangle$. Otherwise it is in the state $|\mathrm{ANC2}\rangle$.

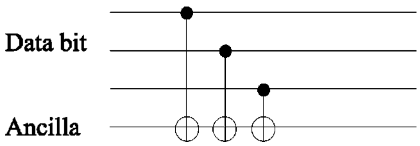

FIG. 7. Measurement in the $\{|0_X\rangle,|1_X\rangle\}$ basis. The ancilla bit begins in the state $|0\rangle$. If the ancilla is observed in the state $|0\rangle$ at the output then the data bit has been collapsed onto $|0_X\rangle$, while if the ancilla is observed in the state $|1\rangle$ then the data bit has been collapsed onto the state $|1_X\rangle$.

Note that a single error in this procedure can result in an overall error in the identification of the ancilla state. This occurs, for errors of the form, $|111\rangle \rightarrow -|111\rangle$ on the cat state, the only part of the above procedure that is not fault tolerant as described above. However, the procedure can be made fault tolerant by performing the same operations with two more cat states and taking a majority vote. The important point in both the controlled-Z and the controlled-controlled-Z gates is that uncorrectable single errors are confined to the cat states. In this case, two independent errors would need to occur to cause an error in identification of the ancilla state.

As shown by Shor, once the state $|\mathrm{ANC1}\rangle$ has been prepared, the Toffoli gate can be carried out using the operations already discussed. For more details see, for example, [4]. In order to prepare the state $|0_X\rangle$, it is sufficient to make a nondestructive measurement in the X basis and then perform a Z gate if the result is $|1_X\rangle$. This measurement can be performed using the circuit shown in Fig. 7. This is simply a parity check in the z basis, since $|0_X\rangle$ is a superposition of all the states containing an even number of $1s$ and $|1_X\rangle$ is a superposition of all the states containing an odd number of $1s$. It may seem somewhat surprising that this method of performing the X-basis measurement is indeed fault tolerant, since the ancilla bit interacts with all the data bits and it might be expected that phase errors could thereby propagate and cause an unrecoverable error. This turns out not to be the case. The possible errors can be classified as follows.

(1) A phase error (in the z basis) occurs on one of the data bits prior to the X-basis measurement. This results in the same (correctable) error in the output state after the measurement.

(2) An error occurs in the first CNOT gate. If this affects the first data bit only or the ancilla bit only, then the result is a phase error on the first qubit after the measurement is completed. If the error affects both the ancilla and the data, then no overall error occurs.

(3) An error occurs in the second CNOT gate. If this affects the data only, then the result is a phase error on the second data bit after the measurement is completed. If it affects both the data and the ancilla, then the result is a phase error on the first data qubit after the measurement is completed. If it affects the ancilla only, then letting the initial state be

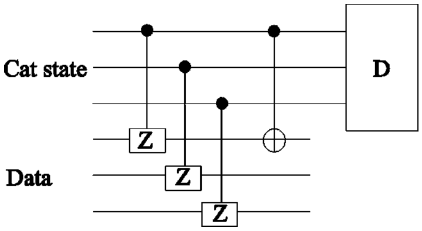

FIG. 8. Circuit for performing a measurement in the Y basis. The box labeled D denotes the disentanglement circuit, which disentangles the cat state, so that $(|000\rangle \pm i|111\rangle)/\sqrt{2} \rightarrow (|0\rangle \pm i|1\rangle)00\rangle/\sqrt{2}$. After this circuit, the first (uppermost in the diagram) cat state bit is measured in the y basis. The procedure is then repeated, with the difference that the CNOT operation in the above diagram is performed between the second qubits in the cat state and the data. The procedure is then carried out a third time, with the CNOT operation being performed between the third qubits in the cat state and data. A majority vote is then taken on the measurement results to obtain the overall measurement result.

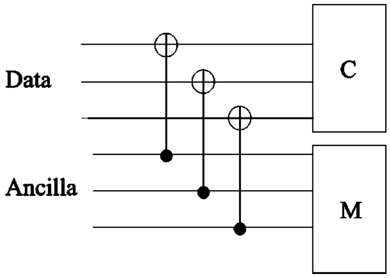

FIG. 9. Fault tolerant error correction circuit. The ancilla is prepared in the initial state if $(|0_C\rangle + |1_C\rangle)/\sqrt{2}$. The box labeled M represents the syndrome measurement, where the ancilla bits are measured in the x basis. If all the ancilla bits are observed in the same state then the data is assumed to be error free. If one ancilla bit is observed in a different state from the other two, then the corresponding data qubit is assumed to have experienced an error. The box labeled C represents the correction of any such error, conditioned on the result of the syndrome measurement.

$(\alpha|0_X\rangle + \beta|1_X\rangle)|0\rangle$, the state prior to the measurement of the ancilla at the end of the circuit is $\alpha|0_{X3}\rangle|0\rangle - \beta|1_{X3}\rangle|1\rangle$, where the 3 indicates a phase error on the third qubit. The additional phase shift (between the $|0_{X3}\rangle|0\rangle$ and $|1_{X3}\rangle|1\rangle$ states) is harmless as the ancilla is measured at this point.

(4) An error occurs in the third CNOT gate. If this affects the data, then the result is an error in the data after the measurement is complete. An ancilla error (with or without a data error) has no affect, as it results only in the type of harmless phase error discussed in (3) above.

The effects of single errors described above can easily be verified by the reader and demonstrate the fault tolerance of this procedure. Note that this measurement can be used to produce the ancilla state needed in the error correction circuit (see below) from the state $|0_C\rangle$. Measurements in the Y basis can be carried out fault tolerantly using the following method.

First prepare the cat state described by Eq. (26). Next perform a bitwise controlled-Z gate with each bare qubit in the cat state as a control bit and the corresponding qubits in the state to be measured (let this be $|\psi_y\rangle$) as a target bit. Next, a CNOT gate is performed with one bit in the cat state as a control bit and the corresponding bit in $|\psi_y\rangle$ as the target. As a consequence the overall system evolves (if no errors occur) as shown below,

$$\begin{aligned} |\psi_y\rangle|\text{CAT}\rangle &= (a|0_Y\rangle + b|1_Y\rangle)|\text{CAT}\rangle \\ &\rightarrow \frac{a}{\sqrt{2}}|0_Y\rangle(|000\rangle + i|111\rangle) \\ &\quad + \frac{b}{\sqrt{2}}|1_Y\rangle(|000\rangle - i|111\rangle). \end{aligned} \tag{28}$$

By disentangling the qubits in the cat state and measuring qubit 1 in the y basis, the state $|\psi_y\rangle$ is collapsed onto either $|0_Y\rangle$ or $|1_Y\rangle$. In order to ensure fault tolerance, this process is repeated three times prior to any measurements being made, using different qubits in the final CNOT operation each time. A majority vote is then taken to determine whether the measurement result is $|0_Y\rangle$ or $|1_Y\rangle$. This is shown in Fig. 8.

It is important to note that, due partly to the problems introduced when only phase errors of the type discussed can be corrected and partly to the structure of the code, the set of universal gates consists of a fairly complex set of operations compared to some of the other codes. The fault tolerant P gate construction in Fig. 4 emphasizes this point. By comparison, for the seven-bit code discussed in [7], the same gate can be carried out simply by performing, bitwise, the operation $P^\dagger$. This is as a result of the particular structure of this code, which is the simplest CSS (Calderbank, Shor, and Steane) code. A discussion of this category of codes is given in [3].

The final ingredient needed for fault tolerant computation is of course a fault tolerant error correction circuit. The circuit for this is quite simple (compared to the correction procedure for other codes) and is shown in Fig. 9.

V. COMPARISON WITH SEVEN-BIT CODE

In this section the three-bit code is compared to the seven-bit code (described below) and it is noted that performing encoded operations requires many more primitive operations in the case of the three-bit code. The seven-bit code is capable of correcting arbitrary single qubit errors and the codewords are

$$\begin{aligned} |0_{C(7)}\rangle = \frac{1}{\sqrt{8}}(&|0000000\rangle + |1010101\rangle \\ &+ |0110011\rangle + |1100110\rangle \\ &+ |0001111\rangle + |1011010\rangle \\ &+ |0111100\rangle + |1101001\rangle), \end{aligned} \tag{29}$$

$$|1_{C(7)}\rangle=\frac{1}{\sqrt{8}}(|1111111\rangle+|0101010\rangle$$
$$+|1001100\rangle+|0011001\rangle$$
$$+|1110000\rangle+|0100101\rangle$$
$$+|1000011\rangle+|0010110\rangle).$$

The codeword representing a logical zero is a superposition of all the even parity (classical) Hamming codewords while the codeword representing a logical one is a superposition of all the odd-parity Hamming codewords [3].

As mentioned in Sec. IV, the most appropriate universal set of gates for the seven-bit code is $\{H,P,\mathrm{CNOT},T\}$. It is possible to carry out the Hadamard operation simply by performing this operation on each individual qubit making up the encoded state. The CNOT gate can be carried out bitwise, simply by performing CNOT gates between corresponding bits in the encoded control and target bits. The P gate can also be performed bitwise although in this case, the operation performed on each individual qubit is actually $P^{\dagger}$. The Toffoli gate can be performed in the same way as for the three-bit code. In this case the state $|\mathrm{ANC1}\rangle$ is produced using a seven-bit cat state, as discussed in [4].

It is clear that the seven-bit code is significantly more efficient than the three-bit code in terms of number of gates required to perform encoded operations due to the possibility of performing all one- and two-bit gates in the universal set given bitwise. For example, the P gate requires only seven single-bit operations in the case of the seven-bit gate, while for the three-bit code, six (best case) or seven (worst case) single-bit operations and 27 two-bit operations are needed, even the absence of any error correction. However, the three-bit code has the advantage that less qubits are needed to encode information. In addition, the error correction procedure is much simpler. There follows an estimation of the threshold value, i.e., the error rate that can be tolerated if arbitrarily long computations are to be possible with bounded error using concatenated coding.

VI. THRESHOLD ESTIMATION

The threshold error rate is estimated using the error model described in Sec. III. We estimate the threshold as follows.

(1) Consider a qubit encoded on the first level of concatenation. An encoded P operation is performed on the qubit and this is followed by an error correction cycle. In addition, error correction is performed on the ancilla, prior to the Y-basis measurement and after each of the first two measurement trials in the Y-basis measurement. The P gate is chosen as it includes the Y measurement, the most complicated single procedure used to construct any of the one-bit gates in the universal set.

(2) An approximate expression is obtained for the probability that an unrecoverable error has occurred after the computation described in (1). Let this probability be P_u.

(3) The probability of error per gate is then P_u. This is compared to ε_g (the error rate per gate for a bare qubit) with the requirement that $P_u<\varepsilon_g$. Hence, the threshold is estimated by finding the value of ε_g which satisfies $P_u=\varepsilon_g$. This is the estimated threshold value.

The most complex part of this procedure is (2), estimating P_u. This is estimated as follows.

(i) We estimate the probability that the P gate is performed without error and an unrecoverable error occurs in the final error correction circuit. This is estimated to $O(\varepsilon_g^2)$. In performing this calculation, it is noted that at least two independent errors must occur for an overall error to occur but that not all sets of two errors correspond to an unrecoverable error.

(ii) Similarly, the probability that an unrecoverable error occurs in one of the other error correction circuits (acting on the ancilla) is estimated and it is assumed that all such errors result in an unrecoverable error in the data [although this is not necessarily the case, see comment at the end of (iv)].

(iii) Similarly, the probability that an unrecoverable error results due to an uncorrected error in the ancilla, prior to one of the error correction circuits and a further error during the correction circuit, is estimated to $O(\varepsilon_g^2)$.

(iv) The probability that an unrecoverable error occurs during a single measurement trial in the Y-basis measurement (i.e., between error correction cycles) is estimated to $O(\varepsilon_g^2)$. In order to simplify the calculation, it is assumed that if two independent errors occur during such a measurement trial then the result is an unrecoverable error (with the exception that two errors affecting a single cat state but not affecting the data are neglected as they clearly do not result in an unrecoverable error). Since certain pairs of errors do not result in an unrecoverable error, this results in a slightly pessimistic threshold estimate.

The contributions from (i), (ii), (iii), and (iv) are added giving an approximate expression for P_u. The expression obtained in this way is

$$P_u\approx\varepsilon_g^2(421.75). \tag{30}$$

Hence, we obtain

$$\varepsilon_{\mathrm{th}}\approx\frac{1}{421.75}\approx2.4\times10^{-3}. \tag{31}$$

This is better than the threshold value estimates obtained for the seven-bit code in [4–6].

VII. CONCLUSIONS

According to the above results it seems (at least superficially) that it would make sense to use the three-bit code for quantum computations using physical systems where dephasing is the dominant source of errors. The encoding procedure is simpler (simply Hadamard transform three qubits all initially in the state $|0\rangle$) and the error correction procedure is also simpler. As shown in Sec. V, it is clear that performing operations on encoded qubits is simpler in the case of the seven-bit code than for the three-bit code. However, it is possible that there are more efficient methods of performing a universal set of fault tolerant operations for the three-bit code than the methods proposed here. If this is the case then the threshold value may be larger than the estimated value in the preceding section. However, there is a compelling argument in favor of choosing the seven-bit code even in systems where dephasing errors dominate. This argument is as follows.

Suppose that in addition to phase-flip ($\hat{\sigma}_z$) errors in the z basis, $\hat{\sigma}_x$ and $\hat{\sigma}_y$ errors also occur. Let the probability of either a $\hat{\sigma}_x$ or $\hat{\sigma}_y$ error occurring as the result of a quantum gate be $\delta \ll \varepsilon_g$. For convenience it is assumed that the error probability is the same for single-bit and two-bit gates. In addition, suppose that any such error results in an unrecoverable error. Since this is usually the case, this only introduces a slight error. What value (roughly) can δ be if such errors are to be negligible relative to $\hat{\sigma}_z$ errors? Clearly this depends on the level of concatenation: the more levels of concatenation, the more likely an unrecoverable $\hat{\sigma}_x$ or $\hat{\sigma}_y$ error is to occur. We can get a fairly good impression of the constraints on δ by considering the following case. Suppose a P gate is performed on a qubit encoded on the first level of concatenation, incorporating the same error correction steps as in the threshold estimate in the previous section. We impose the condition that the probability of a $\hat{\sigma}_x$ or $\hat{\sigma}_y$ error occurring is much less than ε_g. Performing P requires (including ancilla preparation and all the error correction circuits) at most 64 primitive operations. Hence the requirement becomes $67\delta \ll \varepsilon_g$. Note that this is just for a gate at the first level of concatenation. For a large-scale computation, higher level concatenation would be needed to combat dephasing errors and the constraints on gate errors would become even more stringent.

The above argument strongly suggests that it would not be appropriate to use the three-bit code for large-scale quantum computations. However, noting the relative simplicity of the error correction circuit, it may be useful for storage and/or transmission of quantum information, particularly if the states being stored were quite simple to prepare (although this in itself is a significant restriction). In addition, it may be a less daunting task to experimentally perform a very simple fault tolerant computation and error correction using the three-bit code rather than the seven-bit code (assuming dephasing was the dominant source of errors). For example, preparing the state $|0_C\rangle$, storing the data for some time interval and performing fault tolerant error correction requires far less operations for the three-bit code than for the seven-bit code.

In conclusion, it seems that the structure of the seven-bit code gives it a significant advantage over the three-bit code: many operations can be performed bitwise for the seven-bit code whereas for the three-bit code only the CNOT operation can be performed in this way. In addition, the uncorrected ($\hat{\sigma}_x$ or $\hat{\sigma}_y$) errors would accumulate rapidly over the course of a large computation. However, the three-bit code may lend itself more readily to experiments which are likely to be feasible in the short term for physical implementations where dephasing errors dominate.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Ekert for useful discussions on concatenated coding.

[1] M. B. Plenio and P. L. Knight, Phys. Rev. A **53**, 2986 (1996).

[2] M. B. Plenio and P. L. Knight, Proc. R. Soc. London, Ser. A **453**, 2017 (1997).

[3] A. M. Steane, *Introduction to Quantum Computation and Information Theory* edited by Hoi-Kwong Lo, Sandu Popescu, and Tim Spillet (World Scientific, Singapore, 1998), p. 184.

[4] J. Preskill, *Introduction to Quantum Computation and Information Theory*, edited by Hoi-Kwong Lo, Sandu Popescu, and Tim Spillet (World Scientific, Singapore, 1998), p. 213.

[5] E. Knill, R. Laflamme, and W. Zurek, e-print quant-ph/9702058.

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

[7] A. M. Steane, Fortschr. Phys. **46**, 443 (1998).

[8] A. M. Steane, e-print quant-ph/9809054.

[9] G. Mahler, V. A. Weberruss, *Quantum Networks* (Springer, New York, 1995).

[10] J. I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).

[11] B. E. King, C. S. Wood, C. J. Myatt, Q. A. Turchette, D. Leibfried, W. M. Itano, C. Monroe, and D. J. Wineland, Phys. Rev. Lett. **81**, 1525 (1998).

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

[13] P. W. Shor, *37th Symposium on Foundations of Computing* (IEEE Computer Society Press, Vermont, 1996), p. 56.