Threshold Error Penalty for Fault-Tolerant Quantum Computation With Nearest Neighbor Communication

Thomas Szkopek, P. Oscar Boykin, Heng Fan, Vwani P. Roychowdhury, Eli Yablonovitch, *Fellow, IEEE*, Geoffrey Simms, Mark Gyure, and Bryan Fong

***Abstract*—The error threshold for fault-tolerant quantum computation with concatenated encoding of qubits is penalized by internal communication overhead. Many quantum computation proposals rely on nearest neighbor communication, which requires excess gate operations. For a qubit stripe with a width of $L+1$ physical qubits implementing L levels of concatenation, we find that the error threshold of 2.1×10^{-5} without any communication burden is reduced to 1.2×10^{-7} when gate errors are the dominant source of error. This $\sim$175$\times$ penalty in error threshold translates to an $\sim$13$\times$ penalty in the amplitude and timing of gate operation control pulses.**

***Index Terms*—Fault tolerance, quantum information.**

I. Introduction

A CRITICAL architectural issue for quantum computation is the internal communication of quantum information within the processor. There are a variety of proposed quantum processor implementations with different mechanisms for internal communication. For instance, the linear ion trap proposal of Cirac and Zoller [1] involves physical motion of massive ions for internal communication, as do proposals using more complex ion trap structures [2]. Alternative proposals involve using photons and cavity quantum electrodynamics (QED) for communication [3]. The cavity QED approach has been extended to the solid state [4], [5]. Even direct transport of information carrying electrons has been suggested for the solid state [6], [7].

This paper is motivated by another class of quantum computation proposals that rely upon local communication through nearest neighbor interactions [8]–[10]. For instance, communication among electron spins in semiconductors can be performed with sequential SWAP gate operations, which are generated by a controlled Heisenberg exchange between adjacent electrons. An appealing feature of the SWAP operation is that it is generated by the very same two-qubit interaction used for computational operations. Also, a substantial degree of parallelism can be employed. However, the protection of qubits with concatenated error correction requires communication between a number of physical qubits that grows exponentially with concatenation level. This exponential increase in SWAP operations might suggest that concatenated error correction will fail to reduce the logical qubit error rate. Gottesman [11] and Aharanov and Ben-Or [12] have pointed out that a threshold error exists despite an exponential increase in logical gate count with concatenation level L, although no attempt was made to quantify what that threshold might be. In this paper, we estimate that threshold.

The main result we report here is that the number of nearest neighbor communication operations is merely a constant factor over and above the necessary logical operations for error correction at each concatenation level L. Our estimated error thresholds are summarized in Table I. We analyzed in detail fault-tolerant error correction with a concatenated 7-qubit Calderbank–Shor–Steane (CSS) code [13], [14] on a linear qubit stripe with a width of $L+1$ physical qubits for L levels of concatenation and find an $\sim$175$\times$ reduction in threshold gate operation error due to nearest neighbor communication overhead. This translates to a $\sim$13$\times$ increase in accuracy of control pulse amplitude and timing in gate operations. Although nearest neighbor communication incurs a significant penalty in the requisite experimental accuracy of qubit gate operations, it is not a fundamental obstacle to fault-tolerant computation in the solid state. Our analysis is in general agreement with the recent work of Svore *et al.* [15], who also show that internal communication with local interactions incurs an error threshold penalty, although they do not fully account for all communication steps.

This paper is organized as follows. In Section II, we describe the underlying architecture of a quantum processor composed of electron spin qubits, including a description of the physical layout of electron spin qubits and their grouping into concatenated CSS logical qubits. We describe a fault-tolerant error correction protocol in Section III. Our protocol implements error recovery without direct measurement. In Section IV, we calculate the threshold error for gate operations under our error correction protocol, with various assumptions about available resources. Section V considers the relation between control pulse accuracy and gate error thresholds.

II. Layout Architecture

Given the problem of internal communication in a quantum processor, a higher dimensional architecture is preferred because it would allow qubits to be as close as possible. However, there must be access by control wires, thus limiting the packing

Manuscript received January 24, 2005; revised August 17, 2005. This work was supported by the Defense Advanced Research Projects Agency and by the Defense MicroElectronics Activity.

T. Szkopek, H. Fan, V. P. Roychowdhury, and E. Yablonovitch are with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA 90095-1594 USA.

P. O. Boykin is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6200 USA.

G. Simms, M. Gyure, and B. Fong are with HRL Laboratories, LLC, Malibu, CA 90265-4797 USA.

Digital Object Identifier 10.1109/TNANO.2005.861402

TABLE I
GATE COUNT FOR ERROR CORRECTION $N_E + N_{Ec}$ AND FOR LOGICAL CNOT OPERATIONS $N_U + N_{Uc}$ UNDER DIFFERENT ASSUMPTIONS OF INTERNAL COMMUNICATION RESOURCES AND QUANTUM ERROR CORRECTION. APPROXIMATE THRESHOLD GATE ERROR PROBABILITIES ARE GIVEN, AS WELL AS CONTROL PULSE ACCURACY THRESHOLDS (SEE TEXT FOR DETAILS)

		Error Correction Gate Count $N_E + N_{Ec}$	Two-Qubit Unitary Gate Count $N_U + N_{Uc}$	Error Probability Threshold $P_{\rm th} = 2/N^2$	Gate Accuracy Threshold $\phi_{\rm th} = 2\sqrt{P_{\rm th}} \times 180/\pi$
no communication overhead	no $\|0\rangle_L$ preparation	70	7	3.4×10^{-4}	2.1°
	$\|0\rangle_L$ preparation	298	7	2.1×10^{-5}	0.52°
remote CNOT communication	no $\|0\rangle_L$ preparation	238	35	2.7×10^{-5}	0.60°
	$\|0\rangle_L$ preparation	1090	35	1.6×10^{-6}	0.14°
SWAP communication	no $\|0\rangle_L$ preparation	1008	203	1.4×10^{-6}	0.13°
	$\|0\rangle_L$ preparation	3754	343	1.2×10^{-7}	0.034°

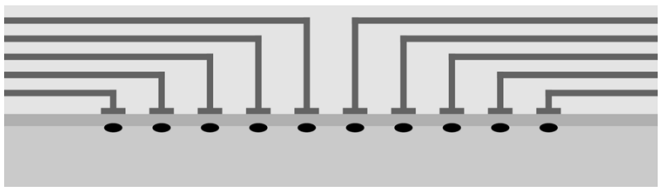

Fig. 1. Schematic representation showing how the number of available metal wire layers limits the width of a 2-D qubit array to only about 10–20 qubits.

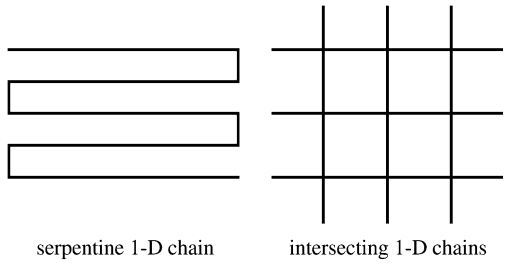

Fig. 2. Requirement for gate electrode access to qubits restricts the layout to stripes of either serpentine or intersecting geometry.

geometry. Fig. 1 shows a schematic cross section of a two-dimensional (2-D) semiconductor qubit array controlled by gate electrodes accessing qubits from the side. The number of vertical stacked control electrodes is limited to twice the number of metal wiring layers in the integrated circuit technology. The need for a reasonable fabrication yield limits the number of metallization layers to ~10, which means that the 2-D array can be at most 20 qubits wide. Fig. 1 illustrates the case for five metallization layers. In this respect, we agree with Copsey *et al.* [16], who pointed out this restriction specifically in the context of semiconductor qubits. Thus, while the qubit array might be locally 2-D, the overall architecture will consist of one-dimensional (1-D) stripes of moderate width, as illustrated in Fig. 2.

The lowest level of concatenated qubit encoding, which is $L = 1$, can be laid out along the stripe width, but all higher concatenation levels must be laid out along the stripe length and are effectively 1-D. Thus, we are led to an essentially 1-D concatenation hierarchy, which is the most challenging for internal quantum communication.

Universal sets of fault-tolerant operations are known only for CSS error-correcting codes of various size [12], [17]–[19]. In our work, we shall consider the CSS code [7], [1], [3]. Concatenation [20], where each logical qubit is composed of encoded qubits, which are in turn composed of encoded qubits and so on, can suppress logical error rate to an arbitrary degree, provided that the physical error rates remain below a threshold value. The self-similarity of concatenation naturally leads to the self-similar logical structure illustrated in Fig. 3. There are seven level $L-1$ logical qubits forming the CSS codeword that represents a single level L logical qubit $|\psi\rangle_L$. A minimum of two logical zeros, $|0\rangle_L$, and six initially arbitrary ancillae, $|a\rangle_{L-1}$, are required to perform error correction on $|\psi\rangle_L$. We consider $L+1$ parallel lines of physical qubits to implement error correction and computation with L levels of concatenation. The error correction protocol is described in detail in the next section. An important feature of the self-similar hierarchy is that, at each concatenation level, the same qubit protection block is employed

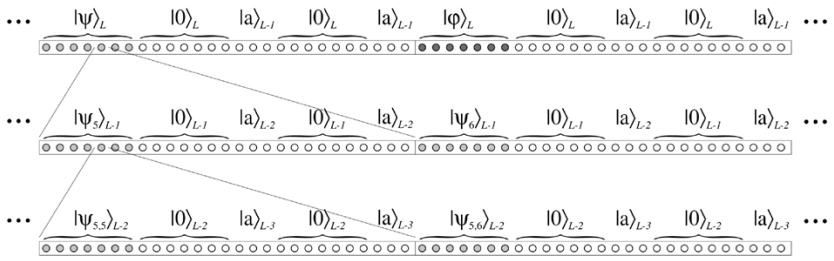

Fig. 3. Self-similar concatenated hierarchy of logical qubits on a linear array, with concatenation level L down to $L-2$, as shown. Error correction requires a minimum of two logical zeros, $|0\rangle_L$, and six ancillae, $|a\rangle_{L-1}$. Altogether, 27 level $L-1$ qubits are minimally required to protect a single level L qubit $|\psi\rangle_L$. The exponential growth with concatenation level L of *physical* nearest neighbor operations to interact $|\psi\rangle_L$ and $|\phi\rangle_L$ is apparent. We consider a layout with $L+1$ adjacent linear arrays of qubits, each organized according to the illustrated logical hierarchy.

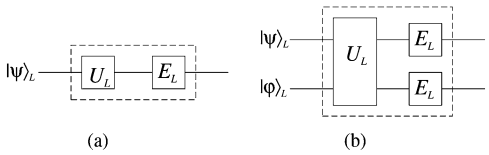

Fig. 4. Each unitary operation U_L at logical level L is followed by error correction E_L at error-correction level L.

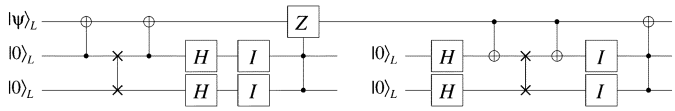

Fig. 5. Modified Steane error-correction circuit (E_L). The indicator block I computes an error syndrome and decodes the syndrome into a bit-wise error indicator used for error recovery. The logical SWAP gate, as well as the CNOT gates, requires shuffling of the constituent $L-1$ qubits (see Fig. 8). We allow only nearest neighbor operations at all logical levels in adherence to self-similarity.

(for ancillae as well as information bearing qubits). Error correction can thus take place at any logical level within an appropriate logical qubit protection block.

III. Error-Correction Protocol

For estimating error thresholds, we consider an aggressive error-correction scheme where every unitary operation U_L at concatenation level L is followed by error correction E_L at level L, as illustrated in Fig. 4.

The error-correction operation E_L can be implemented in a fault-tolerant manner with a Steane error-correction circuit [21] that is slightly modified to that shown in Fig. 5. Error correction takes place within an error-correction block, with the logical qubit $|\psi\rangle_L$ and logical zero states $|0\rangle_L$ explicitly shown. The two groups of three $L-1$ ancillae, $|a\rangle_{L-1}$, are made use of within the bit-flip *indicator* circuit, denoted by I. As can be seen in Fig. 5, the Steane error-correction circuit is particularly parsimonious in its use of gate operations and leads to particularly favorable error thresholds. The bit-flip indicator block I is essential, where, for each logical zero $|0\rangle_L$, it computes a bit-flip error syndrome into three ancillae qubits $|a\rangle_{L-1}$. The syndrome is then decoded within the indicator block I into a bit-wise error indicator that can be directly used for error recovery. Note also that only nearest neighbor operations at logic level L are employed, which is in strict adherence to self-similarity from the physical layer up to concatenation level L.

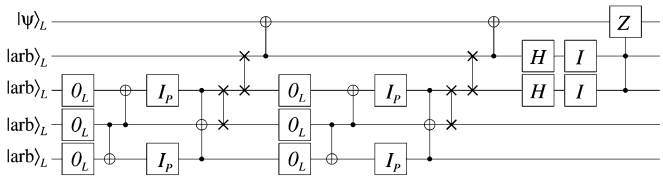

Fig. 6. Error-correction circuit (phase-error portion only) directly incorporating the preparation of requisite logical zeros. Ancillae begin in arbitrary states $|\mathrm{arb}\rangle$. Three 0_L blocks prepare logical zeros that are purified into a single $|0\rangle_L$ state for use in error correction. A modified indicator block I_P corrects for possible parity errors in the raw $|0\rangle_L$'s.

The key point about the bit-flip indicator block I is that it operates on logical zeros that have effectively measured the logical qubit error, but not the logical qubit itself, by virtue of a logical CNOT gate. As was pointed out by Boykin *et al.* [22], the identification of which operations require full quantum coherence and which operations do not is important since "quantum" operations require full protection against both phase-flip and bit-flip errors, while "classical" operations require protection against bit-flip errors only. Note from Fig. 5 that the outputs of indicator block I are used only as control bits for the error recovery operations acting upon the logical qubit. Arbitrary phase flips in the output of I have no effect on the logical qubit. Likewise, phase flips on the input of I have no effect on the logical qubit since the syndrome is encoded as bit flips on the input to I. We need only to protect against bit-flip errors in I, so that the operations within I can be thought of as essentially "classical" in nature, even though they are executed by physical qubit gates. Thus, I can, in principle, be protected with classical fault tolerance, which has been shown to be much more efficient than quantum fault tolerance [23] to ensure that the operations within I will contribute negligibly to the quantum error threshold.

Of course, the requisite logical zeros $|0\rangle_L$ that allow for efficient fault-tolerant error correction are complex entangled states which must be created with low error probability to begin with. One approach to this problem is to dedicate adjacent quantum circuitry whose sole function is to prepare and purify logical zeros, providing a steady supply at various concatenation levels specifically for this purpose. Alternatively, the preparation of logical zeros can be performed directly within the qubit error protection block. The full error-correction circuit is illustrated in Fig. 6. Purification of three $|0\rangle_L$'s, prepared by the 0_L block, results in a single $|0\rangle_L$ state for use in error correction. The 0_L zero preparation block is given in Fig. 7. Bit-flip errors are corrected with a modified indicator block I_P, which also corrects

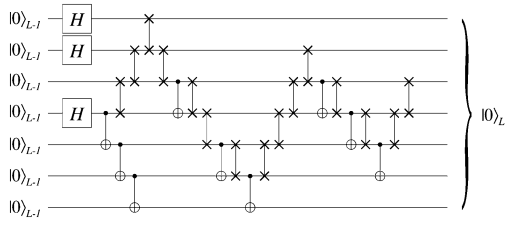

Fig. 7. Circuit 0_L for preparation of a single logical zero $|0\rangle_L$ from lower level $|0\rangle_{L-1}$'s. Only nearest neighbor operations are employed.

for a possible parity flip error corresponding to the logical zero being in the state $|1\rangle_L$ (and thus requiring a minimum of four ancillae). The qubit protection block must increase in size to accommodate $|0\rangle_L$ preparation in this case. A total of 46 qubits would be required that are arranged in the following sequence of $L-1$ qubits (compare with Fig. 6): seven qubits for storing $|\psi\rangle_L$, seven qubits for storing a $|0\rangle_L$, three ancillae $|a\rangle_{L-1}$ for I, seven qubits for storing a $|0\rangle_L$, four ancillae $|a\rangle_{L-1}$ for I_P, seven qubits for storing a $|0\rangle_L$, seven qubits for storing a $|0\rangle_L$, and four ancillae $|a\rangle_{L-1}$ for I_P.

IV. Error Threshold Penalty

The number of physical qubits for our concatenated CSS encoding required to store and protect one logical qubit is 27^L (or 46^L, including logical zero preparation). Several levels of concatenation already lead to a large number of physical qubits (although the width of the qubit stripe grows only as $L+1$). Likewise, the number of physical gate operations grows exponentially (N^L), where N is approximately the number of logical operations required at level $L-1$ in order to implement a single logical function at level L. For example, with a single level of encoding, N is simply the number of physical gate operations required to perform some function on our seven-qubit CSS code word (or multiple code words in the case of a multiqubit logical function).

The number of gate operations N will depend on the function being performed. We consider implementing a simple two-qubit unitary U_L followed by error correction E_L, as illustrated in Fig. 4(b). Error correction might require $N = N_E$ logical gate operations at level $L-1$. There will be additional logical SWAP operations at level $L-1$ that are required to move qubits around, since only nearest neighbor interactions are permitted. We let N_{Ec} be the number of required nearest neighbor SWAP communication operations, which brings the total number of level $L-1$ operations to $N = N_E + N_{Ec}$. Of course, the unitary U_L will require N_U operations at level $L-1$, as well as N_{Uc} additional communication operations at level $L-1$. The total gate operation count at level $L-1$ to implement U_L followed by E_L is simply $N = N_U + N_{Uc} + N_E + N_{Ec}$. The total *physical* gate count is again approximately $N^L = (N_U + N_{Uc} + N_E + N_{Ec})^L$, because each of the N operations at $L-1$ is simply a unitary U_{L-1} followed by error correction E_{L-1}. The self-similar hierarchy requires that N operations at $L-2$ are required for each operation at $L-1$ and so forth, including communication.

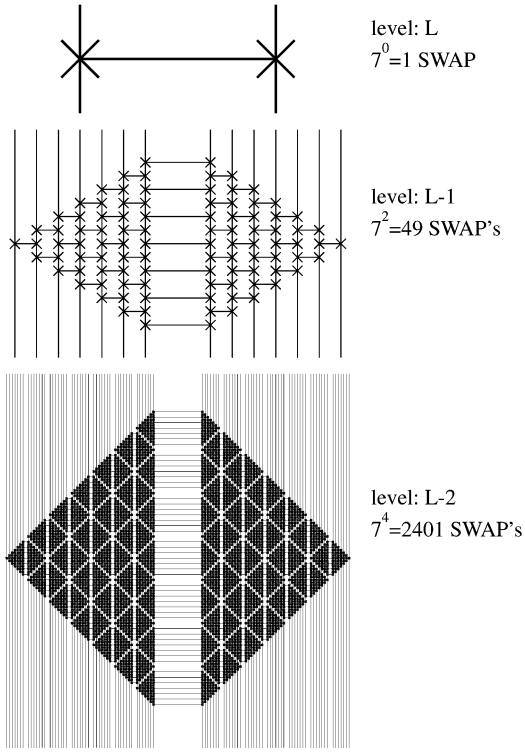

Fig. 8. Logical SWAP operation illustrated at concatenation levels L through $L-2$ with nearest neighbor interactions only. The number of level $L-1$ SWAPs required to implement a single level L SWAP between adjacent logical qubits is $N_U + N_{Uc} = 7 + 42$. There are 21 level $L-1$ SWAPs to interleave the qubits, seven level $L-1$ qubit-wise SWAPs, and 21 level $L-1$ SWAPs to undo the interleaving. Note that a single gate failure does not produce correlated errors within a logical qubit. Error correction, and swapping through the additional qubits in a qubit protection block, are omitted here for clarity.

In reality, the gate count $N_U + N_{Uc}$ varies among the various logical qubit operations possible. For instance, Hadamard at level L requires $N_U = 7$ Hadamard gates at level $L-1$ and $N_{Uc} = 0$ communication gates. In contrast, the gate operations $N_U + N_{Uc} = 7 + 42$ involved in a logical SWAP on the same qubit line are illustrated in Fig. 8 for adjacent logical qubits. Clearly, the value of N^L can be very large, although a substantial fraction of operations at each logical level can be performed in parallel. Note the fault tolerance of the logical SWAP gate: a single swap gate failure induces one error in each logical qubit, which can be recovered independently by error correction. Of course, the extra qubits involved in a qubit protection block increases the number of communication swaps N_{Uc}. As a final example, we show the partial sequence of gate operations required for the logical CNOT gate in Fig. 9. It is in implementing the CNOT gate that an additional line of qubits is used for every concatenation level, resulting in a total of $L+1$ lines of qubits. Similar sequences are used for the SWAP and CNOT gates required for the error-correction operation E_L, contributing to $N_E + N_{Ec}$.

Despite the exponential increase in physical qubits and physical gate operations with concatenation level (while the width of the stripe merely grows linearly in concatenation level), logical errors are suppressed double-exponentially with concatenation level. We let P_1 be the logical error probability on a first-level encoded state $|\psi\rangle_1$ after a two-qubit unitary followed by a single

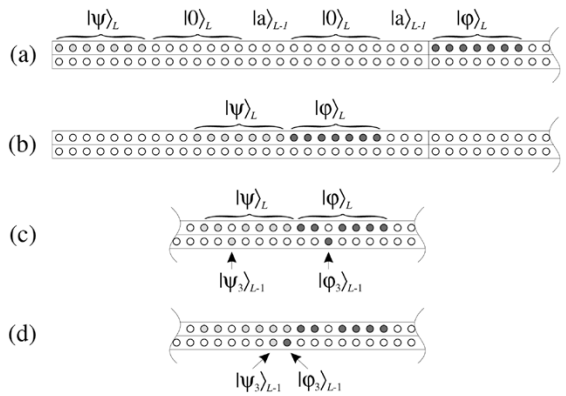

Fig. 9. Partial sequence for a logical level L CNOT operation illustrated at concatenation level $L-1$ with nearest neighbor interactions only. (a) Logical code words $|\psi\rangle_L$ and $|\varphi\rangle_L$ are (b) first brought into adjacent positions, and (c) then each of the seven constituent $L-1$ qubits are moved into an adjacent qubit row to be (d) brought together for qubit-wise interaction (only the third qubits $|\psi_3\rangle_{L-1}$ and $|\varphi_3\rangle_{L-1}$ are shown interacting). The logical qubits are brought back to their original positions for error correction after the logical CNOT. The scheme is applied recursively until physical CNOT gates are performed in the $L+1$st row. The CNOT gates for the error-correction circuit are similarly implemented. Note that a single gate failure does not produce multiple errors within a logical qubit.

error-correction cycle. By the fault-tolerant construction of U_L and E_L, the probability of a logical error is bounded above by the probability that two gate operations fail

$$P_1 \leq \binom{N}{2}\epsilon^2 \simeq \frac{N^2}{2}\epsilon^2 \tag{1}$$

where ϵ is the probability of physical gate error, which is assumed to be equal for all gates, and $N = N_U + N_{Uc} + N_E + N_{Ec}$ as before. While logical error rates shall vary slightly due to differences in $N_U + N_{Uc}$ amongst the logical gate operations with the dominant $N_E + N_{Ec}$ remaining fixed, a conservative estimate can be had by taking the gate counts for the logical CNOT gate as representative. The criterion for error correction to reduce the likelihood of qubit error is $P_1 < \epsilon$. This leads to the threshold error condition $\epsilon < 2/N^2$. Likewise, at higher levels of concatenation, we have

$$P_L \leq \binom{N}{2} P_{L-1}^2 \simeq \frac{N^2}{2} P_{L-1}^2 \tag{2}$$

which leads to $P_{L-1} < 2/N^2 = P_{\text{th}}$ being the error threshold condition for all L. The corresponding required phase accuracy for gate operations, as described in Section V, is $\phi = 2\sqrt{2}/N$. From the above relations, we arrive at the standard logical error probability for concatenated error correction

$$P_L \leq P_{\text{th}} \left(\frac{\epsilon}{P_{\text{th}}}\right)^{2^L} \tag{3}$$

but where N now includes the nearest neighbor communication overhead at a particular concatenation level. The exponent 2^L results in an overwhelming *super*-exponential in L suppression of logical errors, while the number of qubits and gate operations increases only exponentially in L.

Suppose that a quantum computation requires a sequence of T logical gate operations, then a logical error probability $P_L = 1/T$ will give the correct result with only several trials of the computation. The relation between the maximum number T of operations in a calculation and concatenation level L can be written

$$T = \frac{1}{P_L} \geq \frac{1}{P_{\text{th}}} \left(\frac{P_{\text{th}}}{\epsilon}\right)^{2^L} \tag{4}$$

or alternatively

$$L \leq \log_2 \left(\frac{\log_2(TP_{\text{th}})}{\log_2(P_{\text{th}}/\epsilon)}\right). \tag{5}$$

For instance, the error threshold might be $P_{\text{th}} = 10^{-6}$, while the physical gate operation error is an order of magnitude better at $\epsilon = P_{\text{th}}/10 = 10^{-7}$. We then have an accessible computation length $T = 10^6 \times 10^{2^L}$, which, for $L = 3$, gives $T \geq 10^{14}$. It follows that interesting calculations can be performed with only a few layers of concatenation (i.e., a qubit stripe with a width of only a few qubits) if physical error probabilities well below the error threshold can be achieved.

The problem of estimating error threshold has been reduced to counting gate operations, for which our numerical results are summarized in Table I. Note that we have neglected storage errors in our present analysis since the coherence times of electron spins in semiconductors [24] exceed the expected gate operation times by at least ~8 orders of magnitude, with further improvement expected. The top row of Table I gives the most favorable error thresholds where any qubit can interact with any other qubit without any extra communication operations. The bottom row is the least favorable case where nearest neighbor SWAP operations are used on a linear qubit array to implement all operations. The middle row represents an intermediate case, where the remote CNOT is used to perform a CNOT gate between distant qubits [25], [26]. The remote CNOT requires a shared Einstein–Podolsky–Rosen (EPR) pair, which is a resource that might be generated by independent hardware with sufficient purity that the EPR error rate contributes negligibly to the overall error rate of the remote CNOT and the error threshold. Measurement and classical communication are also required for the remote CNOT (see the Appendix).

For all three communication schemes, the gate count is given in Table I for subcases where $|0\rangle_L$'s are supplied by adjacent circuitry (e.g., a parallel qubit stripe) or where the $|0\rangle_L$'s are prepared directly within the error-correction circuit itself (as in Fig. 6), thus burdening the error threshold. In the former case, we assume that the adjacent circuitry can prepare and purify logical zeros to reach an error probability much less than the preparation circuit of the former case, thereby contributing to the error threshold negligibly. This might be achieved by successive rounds of purification.

In all cases, we assume that those portions of the circuit that can be implemented with classical fault-tolerant logic [22], albeit with qubit gates, take advantage of the greater efficiency of classical coding. The threshold error for classical fault-tolerant circuits has been estimated to be between ~1/100 to ~1/3000 depending on topology and communication resources [23]; therefore, we assume that the error rates in the classical circuits are negligible compared to the quantum circuits, so that in counting the gate operations we can neglect the operations in I

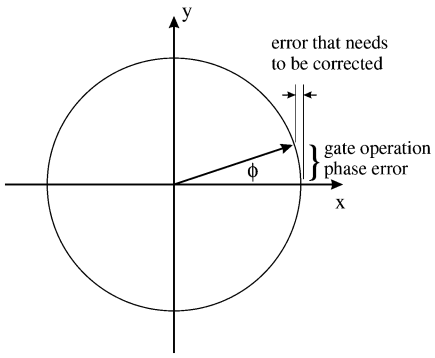

Fig. 10. Conceptual illustration of a qubit pseudospin that might miss a target x axis by an angle ϕ due to a control pulse error. The resulting probability of qubit error is $\epsilon \approx (\phi/2)^2$.

and I_P. Furthermore, the dual-control phase flip ($\Lambda_2(Z)$) and dual-control bit flip ($\Lambda_2(X) =$ Toffoli) are assumed to count merely as two-qubit interactions, since fault-tolerant classical logic can be used to generate a single classical control bit. The remaining sundry details involved in counting gate operations are left to the Appendix.

Observing the gate error thresholds in Table I, we see that SWAP communication incurs a penalty of $\sim$175$\times$ compared to the case of free communication. Communication through the remote CNOT incurs a penalty of $\sim$12$\times$ compared to the free communication case. The improvement associated with remote-CNOT communication is not as much as one might expect, since the remote CNOT requires multiple operations proportional to the size of the logical qubits. Thus, internal quantum communication reduces gate error thresholds for fault-tolerant computation by a substantial factor that we estimate to be from $\sim$12$\times$ to $\sim$175$\times$. While this certainly increases the difficulty in experimentally realizing fault-tolerant gate operations, it is by no means an impasse for solid-state quantum computation, as we discuss in Section V.

V. Error Probability and Gate Operation Accuracy

So far, we have worked entirely with error probabilities. In practice, experimental gate accuracy is more naturally specified in terms of control pulse amplitude. Consider the spin (or a qubit pseudospin), illustrated in Fig. 10. Suppose a control pulse, as is used in spin resonance, was to bring the spin into alignment with the x axis. However, an error in pulse area, phase, or timing may cause a misalignment by some small angle ϕ. The probability of error ϵ is then the probability that the spin is not projected into the $+x$ direction when a measurement is performed along the x axis. The probability of projection along the $+x$ direction is $\cos^2(\phi/2)$, so that the error probability is

$$\epsilon = \sin^2(\phi/2) \approx (\phi/2)^2. \tag{6}$$

The required gate timing and amplitude accuracy is $\phi = 2\sqrt{\epsilon}$, which is specified as a phase angle, is proportional to the *square root* of the threshold error probability. The gate accuracy thresholds are given in degrees in Table I. Of course, the $\sim$12$\times$ to $\sim$175$\times$ penalty in error probability threshold becomes only a $\sim$3.5$\times$ to $\sim$13$\times$ penalty in control pulse accuracy. In order to achieve an error probability of 10^{-7}, one would require about 1/30 of a degree accuracy in control pulse timing, which is not entirely infeasible since it would require about 1-ps phase accuracy in a clock period of about 10 ns. Recall that an error probability of 10^{-7} for a quantum processor with threshold error probability 10^{-6} and three levels of concatenation will allow a computation with $\geq 10^{14}$ operations. Thus, thinking about gate errors in terms of phase angle makes it clear that very small error probabilities are achievable.

VI. Conclusion

Internal quantum communication remains a challenging architectural problem that impacts the threshold error for fault-tolerant computation with encoded logical qubits. The communication operation overhead required to distribute information among a number of qubits that grows exponentially with concatenation level can be a significant burden. Whether one is limited to nearest neighbor communication, a communication bus (as in the original Cirac–Zoller ion trap proposal [1]), or communication by modified teleportation schemes such as the remote CNOT, there is always a communication penalty in error threshold. The minimum communication overhead cost is associated with a communication bus, where a single operation for "transmitting" and a single operation for "receiving" is possible in principle. The question of whether a sufficiently robust communication bus is available for solid-state qubits remains open. Ballistic transport of electron spins through mesoscopic wires is predicted to give error rates of $\sim$0.6 for GaAs [7], far above our stated threshold requirements even for the free communication case. Much more promising is the combination of cavity QED techniques with confined electron spins [4] or superconducting circuits [5], where an electromagnetic bus can couple a number of qubits. The error rates of such a bus, the reconfigurability of its links, and its parallelism (i.e., how many qubits can be transported simultaneously or through the same link) must all be carefully considered in determining what benefits, if any, we can expect over nearest neighbor architectures. Nonetheless, we expect that communication overhead can be mitigated to a large extent by circuit optimization. Recent work [27] on laying out Shor's factorization algorithm on a linear chain of qubits under the restriction of nearest neighbor interaction has shown that circuit optimization can greatly reduce the number of logical qubit SWAPs required.

Appendix I
Threshold Error Calculations

We provide a brief summary here of the counting of gate operations, which then leads to the threshold error. Error correction at concatenation level L with the circuit E_L requires the use of both single-qubit unitaries and two-qubit unitaries at levels L down to the physical layer. Interestingly, the quantum portions of the circuit E_L (see Figs. 5 or 6) consists of gate operations that are directly fault-tolerant, where qubit-wise (or transversal) operations are sufficient. These operations include CNOT, SWAP, and Hadamard rotation (H). The control bits of the dual control gates are classical, so a full quantum Toffoli is not required. Of course, indirectly fault-tolerant gates such as

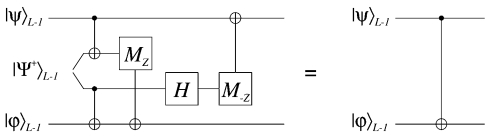

Fig. 11. Remote-CNOT gate, modified from [26], requires a shared EPR pair $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$, measurement M_Z, and classical communication to implement a CNOT operation between distant qubits.

the Toffoli ($\Lambda_2(X)$) or $\pi/8$ rotation ($Z^{1/4}$) are required for universal computation. We do not calculate the error threshold for indirectly fault-tolerant gates here.

A. Free Communication

First, we consider the idealized case where communication is achieved without any extra operations, in other words, any two qubits can interact directly at any time. In this case, $N_{Uc} = N_{Ec} = 0$, and we need only count the number of computationally useful gates. A directly fault-tolerant two-qubit unitary will require $N_U = 7$ operations. The error correction gate count without logical zero preparation is

$$N_E = 4 \times 7\mathrm{CNOT} + 4 \times 7\mathrm{H} + 7\Lambda_2(X) + 7\Lambda_2(Z) = 70 \quad (7)$$

where the $L-1$ gate type and count are indicated. With logical zero preparation, we have

$$\begin{aligned} N_E &= 70 + 12 \times 0_L + 4 \times 7\Lambda_2(X) + 8 \times 7\mathrm{CNOT} \\ &= 70 + 12 \times (3\mathrm{H} + 9\mathrm{CNOT}) + 84 \\ &= 298 \end{aligned} \quad (8)$$

where again $L-1$ gate type and count was indicated.

B. Remote-CNOT Communication

Next, we consider the intermediate communication case involving remote-CNOT operation, which we abbreviate as reCNOT. The reCNOT circuit is indicated in Fig. 11. For simplicity, we assume that the classical communication and EPR preparation introduce negligible errors compared to the other gate operations involved. We see that a reCNOT between two level $L-1$ qubits requires five level $L-1$ operations, so that a reCNOT between two level L qubits requires $N_U + N_{Uc} = 5 \times 7$ level $L-1$ operations. The error correction gate count without logical zero preparation becomes

$$\begin{aligned} N_E &= 4 \times 7\mathrm{reCNOT} + 4 \times 7\mathrm{H} + 7\Lambda_2(X) + 7\Lambda_2(Z) \\ &= 140 + 28 + 35 + 35 \\ &= 238 \end{aligned} \quad (9)$$

where $\Lambda_2(X)$ and $\Lambda_2(Z)$ are counted as reCNOT operations (recall that they can be implemented with single classical control bits). With logical zero preparation, we have

$$\begin{aligned} N_E &= 238 + 12 \times 0_L + 4 \times 7\Lambda_2(X) + 8 \times 7\mathrm{reCNOT} \\ &= 238 + 12 \times (3\mathrm{H} + 3\mathrm{CNOT} + 6\mathrm{reCNOT}) \\ &\quad + 140 + 280 \\ &= 238 + 432 + 140 + 280 \\ &= 1090 \end{aligned} \quad (10)$$

where we have made use of both nearest neighbor CNOT and reCNOT in the logical zero preparation.

C. SWAP Communication

Finally, we consider communication by SWAP gates. Without logical zero preparation, a level L qubit protection block is 27 $L-1$ qubits long. Applying CNOT between two level L qubits as in Fig. 9 requires $N_U + N_{Uc} = 203$ level $L-1$ operations on each logical qubit argument. The error-correction operation requires

$$\begin{aligned} N_E &= 4 \times (7\mathrm{CNOT} + 112\mathrm{SWAP}) \\ &\quad + 4 \times 7\mathrm{H} + 2 \times (7\mathrm{SWAP} + 84\mathrm{SWAP}) \\ &\quad + (7\Lambda_2(X) + 154\mathrm{SWAP}) + (7\Lambda_2(Z) + 154\mathrm{SWAP}) \\ &= 1008 \end{aligned} \quad (11)$$

where we note that 112 communication SWAPs are required for applying CNOT between $|\psi\rangle_L$ with an adjacent $|0\rangle_L$, and 84 communication SWAPs are required for logical swapping of a $|0\rangle_L$ with another $|0\rangle_L$ taking account of the extra ancillae $|a\rangle_{L-1}$ in the way.

When logical zero generation is included, the qubit protection block increases in size to 46 qubits. Applying CNOT between two level L qubits now requires $N_U + N_{Uc} = 343$ level $L-1$ operations because of the increased size of the qubit protection block. The error correction operation requires

$$\begin{aligned} N_E &= 1008 + 12 \times 0_L + 2 \times (7\mathrm{SWAP} + 84\mathrm{SWAP}) \\ &\quad + 4 \times (7\mathrm{SWAP} + 98\mathrm{SWAP}) + 4 \times (7\mathrm{CNOT} \\ &\quad + 112\mathrm{SWAP}) + 4 \times (7\mathrm{CNOT} + 168\mathrm{SWAP}) \\ &\quad + 4 \times (7\Lambda_2(X) + 154\mathrm{SWAP}) \\ &= 3754 \end{aligned} \quad (12)$$

where we note that each logical $|0\rangle_L$ generation requires 27 level $L-1$ operations (Fig. 7), and the SWAP communication accounts for all extra ancillae $|a\rangle_{L-1}$ in the way.

ACKNOWLEDGMENT

The authors would like to thank I. Chuang for bringing the issue of internal communication in quantum computation to our attention. The authors also thank D. Gottesman for pointing out an error in an earlier manuscript.

REFERENCES

[1] J. I. Cirac and P. Zoller, "Quantum computations with cold trapped ions," *Phys. Rev. Lett.*, vol. 74, pp. 4091–4094, 1995.

[2] D. Kielpinski, C. R. Monroe, and D. J. Wineland, "Architecture for a large-scale ion-trap quantum computer," *Nature*, vol. 417, pp. 709–711, 2002.

[3] J. I. Cirac, P. Zoller, H. Kimble, and H. Mabuchi, "Quantum state transfer and entanglement distribution among distant nodes in a quantum network," *Phys. Rev. Lett.*, vol. 78, pp. 3221–3224, 1997.

[4] A. Imamoglu, D. D. Awschalom, G. Burkard, D. P. DiVincenzo, D. Loss, M. Sherwin, and A. Small, "Quantum information processing using quantum dot spins and cavity qed," *Phys. Rev. Lett.*, vol. 83, pp. 4204–4207, 1999.

[5] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R. S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, "Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics," *Nature*, vol. 431, pp. 162–167, 2004.

[6] G. Burkard, D. Loss, and E. V. Sukhorukov, "Noise of entangled electrons: Bunching and antibunching," *Phys. Rev. B*, vol. 61, pp. R16303–R16306, 2000.
[7] P. Recher, E. V. Sukhorukov, and D. Loss, "Quantum dot as spin filter and spin memory," *Phys. Rev. Lett.*, vol. 85, pp. 1962–1965, 2000.
[8] D. Loss and D. P. DiVincenzo, "Quantum computation with quantum dots," *Phys. Rev. A*, vol. 57, pp. 120–126, 1998.
[9] B. E. Kane, "A silicon-based nuclear spin quantum computer," *Nature*, vol. 393, pp. 133–137, 1998.
[10] R. Vrijen, E. Yablonovitch, K. Wang, H. W. Jiang, A. Balandin, V. Roychowdhury, T. Mor, and D. DiVincenzo, "Electron spin resonance transistors for quantum computing in silicon-germanium heterostructures," *Phys. Rev. A*, vol. 62, pp. 012 306-1–012 306-10, 2000.
[11] D. Gottesman, "Fault-tolerant quantum computation with local gates," *J. Mod. Opt.*, vol. 47, pp. 333–345, 2000.
[12] D. Aharanov and M. Ben-Or. Fault-Tolerant Quantum Computation with Constant Error Rate. [Online] Available: xxx.lanl.gov/abs/quant-ph/9 906 129
[13] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Phys. Rev. A*, vol. 54, pp. 1098–1105, 1996.
[14] A. M. Steane, "Error correcting codes in quantum theory," *Phys. Rev. Lett.*, vol. 77, pp. 793–797, 1996.
[15] K. M. Svore, B. M. Terhal, and D. P. DiVincenzo. Local Fault-Tolerant Quantum Computation. [Online] Available: xxx.lanl.gov/abs/quant-ph/0 410 047
[16] D. Copsey, M. Oskin, F. Impens, T. Metodiev, A. Cross, F. T. Chong, I. L. Chuang, and J. Kubiatowicz, "Toward a scalable, silicon-based quantum computing architecture," *IEEE J. Sel. Topics Quantum Electron*, vol. 9, no. 6, pp. 1552–1569, Nov./Dec. 2003.
[17] P. W. Shor, "Fault-tolerant quantum computation," in *Proc. 37th Annu. Symp. Foundations of Computer Science*, 1996, pp. 56–65.
[18] P. O. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan, "On universal fault-tolerant quantum computing: A novel basis and a new constructive proof of universality for shor's basis," in *Proc. 40th Ann. Symp. Foundations of Computer Science*, 1999, pp. 486–494.
[19] A. M. Steane and B. Ibinson. Fault-Tolerant Logical Gate Networks for CSS Codes. [Online] Available: xxx.lanl.gov/abs/quant-ph/0 311 014
[20] E. Knill and R. Laflamme. Concatenated Quantum Codes. [Online] Available: xxx.lanl.gov/abs/quant-ph/9 608 012
[21] A. M. Steane, "Active stabilization, quantum computation, and quantum state synthesis," *Phys. Rev. Lett.*, vol. 78, pp. 2252–2255, 1997.
[22] P. O. Boykin, T. Mor, V. Roychowdhury, and F. Vatan, "Fault tolerant computation on ensemble quantum computers," in *Proc. Int. Conf. Dependable Systems and Networks*, 2004, pp. 157–166.
[23] P. O. Boykin and V. P. Roychowdhury, "*Proc. Int. Conf. Dependable Systems and Networks*,", (DSN 2005), pp. 444–453. [Online]. Available: xxx.lanl.gov/abs/cs/0 504 010.
[24] A. M. Tyryshkin, S. A. Lyon, A. V. Astashkin, and A. M. Raitsimring, "Electron spin relaxation times of phosphorus donors in silicon," *Phys. Rev. B*, vol. 68, p. 193 207, 2003.
[25] A. Sørenson and K. Mølmer, "Error-free quantum communication through noisy channels," *Phys. Rev. A*, vol. 58, pp. 2745–2749, 1998.
[26] D. Gottesman. The Heisenberg Representation of Quantum Computers. [Online] Available: xxx.lanl.gov/abs/quant-ph/9 807 006
[27] A. G. Fowler, S. J. Devitt, and L. C. L. Hollenberg. Implementation of Shor's Algorithm on a Linear Nearest Neighbor Qubit Array. [Online] Available: xxx.lanl.gov/abs/quant-ph/0 402 196

Thomas Szkopek received the B.A.Sc. and M.A.Sc. degrees from the University of Toronto, Toronto, ON, Canada, in 1999 and 2001, and he is currently working toward the Ph.D. degree at the University of California, Los Angeles.

He is studying under the direction of Prof. E. Yablonovitch, where he is undertaking research in single-electron spin measurement.

P. Oscar Boykin received the Ph.D. degree in physics from the University of California at Los Angeles (UCLA), in 2002.

He is currently an Assistant Professor of electrical and computer engineering at the University of Florida, Gainesville. His research interests include quantum cryptography, quantum information and computation, peer-to-peer computing, and neural coding.

Heng Fan, photograph and biography not available at the time of publication.

Vwani P. Roychowdhury received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1989.

From 1991 to 1996, he was a faculty member with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, where he was promoted to Associate Professor in 1995. In 1996, he joined the University of California, Los Angeles, where he is currently a Professor of electrical engineering. He also serves on the faculty of the Biomedical Engineering Interdepartmental Program. His research interests include models of computation, quantum and nanoelectronic computation, quantum information processing, fault-tolerant computation, combinatorics and information theory, advanced statistical processing, and adaptive algorithms.

Eli Yablonovitch (F'92) received the Ph.D. degree in applied physics from Harvard University, Cambridge, MA, in 1972.

He was with Bell Telephone Laboratories for two years and then became a Professor of applied physics at Harvard. In 1979, he joined Exxon to do research on photovoltaic solar energy. Then in 1984, he joined Bell Communications Research, where he was a Distinguished Member of Staff and a Director of Solid-State Physics Research. In 1992, he joined the University of California, Los Angeles, where he is now The Northrop Grumman Opto-Electronics Chair, Professor of Electrical Engineering. His work has covered a broad variety of topics: nonlinear optics, laser–plasma interaction, infrared laser chemistry, photovoltaic energy conversion, strained quantum-well lasers, and chemical modification of semiconductor surfaces. Currently, his main interests are in optoelectronics, high-speed optical communications, high-efficiency light-emitting diodes and nano-cavity lasers, photonic crystals at optical and microwave frequencies, quantum computing, and quantum communication.

Prof. Yablonovitch is a Fellow of the Optical Society of America and the American Physical Society, a Life Member of Eta Kappa Nu, and a member of the National Academy of Engineering and the National Academy of Sciences. He was the recipient of the Adolf Lomb Medal, the W. Streifer Scientific Achievement Award, the R. W. Wood Prize, and the Julius Springer Prize.

Geoffrey Simms received the B.S. degree in physics from the California State Polytechnic University, Pomona, in 1995, and the M.S. degree in physics from he University of Illinois at Urbana-Champaign in 1997.

His research has been in computational general relativity, computational electromagnetics, strain modeling in semiconductor epitaxy, and quantum information devices and systems. He has been with the Computational Physics Department, HRL Laboratories, LLC, Malibu, CA, since 1998.

Mark Gyure received the Ph.D. degree in theoretical condensed matter physics from the University of Colorado at Boulder, in 1990.

He is currently a Senior Research Physicist and Manager of the Computational Physics Department, HRL Laboratories, LLC, Malibu, CA. He joined HRL Laboratories in 1994 following post-doctoral research with Boston University in the area of computational statistical physics. At HRL Laboratories, his research has been in computational materials and device physics including the theory of semiconductor growth, computational methods in statistical and quantum mechanics, and devices and architectures for solid-state implementations of quantum computing.

Bryan Fong, photograph and biography not available at the time of publication.