LA-UR

96-2718

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE: ASSUMPTIONS FOR FAULT TOLERANT QUANTUM COMPUTING

AUTHOR(S): E. Knill

RECEIVED

SEP 09 1996

OSTI

SUBMITTED TO: External Distribution - Hard Copy

MASTER

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive royalty-free license to publish or reproduce the published form of this contribution or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Los Alamos

Los Alamos National Laboratory
Los Alamos New Mexico 87545

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

RB

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Assumptions for Fault Tolerant Quantum Computing

E. Knill, R. Laflamme

knill@lanl.gov, Mail Stop B265
laflamme@lanl.gov, Mail Stop B288
Los Alamos National Laboratory
Los Alamos, NM 87545.

June 1996

Abstract

Assumptions useful for fault tolerant quantum computing are stated and briefly discussed. We focus on assumptions related to properties of the computational system. The strongest form of the assumptions seems to be sufficient for achieving highly fault tolerant quantum computation. We discuss weakenings which are also likely to suffice.

Note: This report is preliminary. Please let us know of any suggestions, omissions or errors so that we can correct them before distributing this work more widely.

1 Introduction

Several recent advances in the theory of quantum computing make it increasingly likely that quantum algorithms can eventually be implemented on real devices. These advances include the error-correcting codes of Shor [13], Calderbank and Shor [3] and Steane [16], the application of concatenated codes to quantum channels and memories of Knill and Laflamme [9, 10], and the fault tolerant computing methods of Shor [15]. The validity of these methods depends on several assumptions on the functioning of the primitive devices and operations in a quantum computer. These assumptions are critical for fully understanding the technical obstacles that need to be overcome before quantum computing becomes a reality. In this note we

state and partially formalize the most important assumptions. Weakenings which might be more accessible to implementation are given. The strongest forms of the assumptions appear to be sufficient to achieve excellent fault tolerance properties and suffice to establish the claims of [15]. The weaker versions also seem to be sufficient but may require more complicated proofs of (or “optimistic” arguments for) the required fault tolerance properties.

2 Assumptions for Fault Tolerant Computing

2.1 Perfect Classical Computing

Assumption 2.1. *Perfect classical computing: An error free classical computer which can control the state of the quantum computation is available.*

It is increasingly clear that classical computations are extremely useful for quantum algorithms. Classical methods are used for the off-line computation of classical results [12, 13], for error correction [14, 3, 16, 2, 8], implementing otherwise inaccessible operations [15, 10], operational control and efficiency [7]. This is compatible with the view of a practical quantum computer as a quantum random access machine (QRAM), where a standard classical RAM has access to both classical and quantum registers, with the ability to initialize quantum registers in classical states, to apply local unitary operations to quantum registers and to measure them in the classical basis. In fault tolerant computing, the primitive operations are further restricted to ensure preservation of fault tolerance.

A critical aspect of all fault tolerant quantum computing proposals to date is the use of measurement followed by conditional operations to maintain fault tolerance. If the classical computer which controls the quantum state makes errors during those operations, this will affect the entire state of the quantum computation. Although some errors such as incorrect measurements can be thought of as errors in the underlying quantum gates rather than the classical controller, others can impact the state of the computations much more directly than would individual errors in the qubits or in the primitive unitary operations. Such errors limit the length of the computation (both quantum and classical).

Fortunately, methods for controlling errors in classical computers are well understood, and any practical classical computation can be implemented essentially without error. Of course, as long as the final answer of interest to the user is unique, it suffices to obtain it with some sufficiently small error probability. Simple repetition serves to confirm it to any level of confidence

desired. Thus the weak version of this assumption requires only that the total probability of error in the classical component of the computation is sufficiently small ($< .5$ for a yes/no problem).

2.2 Parallel quantum operations

Assumption 2.2. *Parallel Quantum Operations: A set of primitive unitary operators on independent qubits (or other primitive component systems) can be implemented in parallel.*

There are two reasons why parallelism is useful for fault tolerant computing. First, one of the key features of existing proposals is that many operations are implemented independently on disjoint sets of qubits, both for error correction and for fault tolerant operations. Second, errors are introduced into each qubit at a constant rate in time (e.g. by decoherence), regardless of other operations. Thus there is a typical time scale within which most qubits must be considered at least once by the classical controller, else the state of the computation is lost. This means that no matter how many qubits are required by the computation, the classical controller must be able to address each of them within this typical time scale. To avoid scaling problems, it is convenient to assume that the necessary operations can be applied in parallel. However, it suffices for the controller to be able to work at a rate sufficiently fast to operate on the total number of qubits within a small fraction of the decoherence time scale.

2.3 No Leakage

Assumption 2.3. *No leakage: The representation of the component systems of the computation (e.g. qubits) is effectively complete. That is there is no amplitude leakage from the systems to a direct summand not involved in the computation.*

The no leakage assumption is normally hidden in the formalism used to describe the state of a quantum computer. The state is assumed to exist on a tensor product of qubits, $Q^{\otimes n}$. This idealized assumption almost never holds in practice. In general, the true Hilbert space of each qubit involves large direct summands. In effect, $Q^{\otimes n}$ should be replaced by $(Q + \mathcal{R})^{\otimes n} + \mathcal{M}$. This matters only if the internal evolution of the system and effects such as relaxation allow transitions from Q to $\mathcal{R}$ or (in the worst case) to $\mathcal{M}$.

There are two conceptually simple methods for dealing with the problem of leakage from Q to $\mathcal{R}$. The first is to detect leakage when it occurs and

return the amplitude of the state to Q . The technical problem of detecting leakage may be highly non-trivial. For example, if the polarization of a photon is used to represent a qubit, one source of leakage is photon loss. In this case, restoring amplitude involves non-destructively detecting the absence of the photon and replacing it if it is not present.

The second method involves explicitly representing all the relevant leaked states and either using them in the quantum codes or operating on them directly during the critical operations. The complexity of this method depends on the exact nature of the leaked states and the transition amplitudes.

In practice, the no leakage assumption is satisfied if the leakage amplitudes are negligible, or if leakage is explicitly managed by the error correction procedures.

2.4 Locality of Operations

Assumption 2.4. *Locality of operations: The effective operation associated with the implementation of a primitive operation on a set of qubits (the target qubits) has no effect on other qubits.*

Suppose that the state of the computation is supported on $Q^{\otimes n}$ and we wish to apply an operation U to one or two of the qubits. The actual operation can be represented either as a superoperator $\mathcal{A}$ [11] or as a sum of error operators $\mathcal{E}$ [9] acting on $Q^{\otimes n}$. In either case, the locality assumption requires that the operation is of the form $\mathcal{A} \otimes I$, with $\mathcal{A}$ acting only on the target qubits.

Since each qubit is acted on whether or not it is the target of an operation, the assumption needs to be further clarified. If a qubit is not the explicit target of an operation in the current round, it is acted on by an implicit “memory” operation. Thus each global operation is a set of parallel operations acting on independent qubits. Formally, we partition the qubits into sets $X_1, \dots, X_r$, where each X_i normally contains only one or two qubits. The intended parallel operation is a tensor product of unitary operators $U_1 \otimes \dots \otimes U_r$, with U_i acting on the qubits in X_i . The memory operation is the identity. The locality assumption requires that the actual superoperator applied is also a tensor product $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_r$, with $\mathcal{A}_i$ acting on the qubits in X_i .

If we use the error basis representation, then each $\mathcal{A}_i$ can be written in the form $\sum_j |e_j\rangle E_j U_i$, where the E_j are in a standard error operator basis (e.g. tensor products of bit flips, sign flips or their products), and where the $|e_i\rangle$ are non-orthogonal and non-normalized states in the environment.

A first weakening of the locality assumption is obtained by assuming that correlated errors have very low amplitude compared to the independent ones. Thus the true operation is well approximated by a tensor product of the required form. A much greater weakening of the locality assumption requires that the error component of the operation behaves as if independent. That is the amplitude of errors involving k qubits should decrease exponentially with k . This still requires that there is essentially no unintended coupling of the qubits. Effects such as wave propagation are disallowed and if present could be very destructive.

A formal version of weak locality requires a good measure of the effective error rate. One such measure which has good behavior under composition (even with the weakest assumptions on dependencies) can be based on the following rather technical definition: If E is a tensor product of error operators on $Q^{\otimes n}$, let $|E|$ (the *weight* of E) be the number of non-identity factors of E . The *support* of E , $\text{supp} E$, is the set of qubits on which E acts with a non-identity error operator. A filter $\mathcal{F}$ is a family of subsets of the qubits which is upwards closed. That is, if $X \in \mathcal{F}$ and $Y \supseteq X$, then $Y \in \mathcal{F}$. The p -weight of $\mathcal{F}$ is given by the sum $|\mathcal{F}|_p = \sum_{X \in \mathcal{F}} p^{|X|}$. An operator of the form $\mathcal{E} = \sum_i |e_i\rangle E_i$ has error rate bounded by p if for each filter $\mathcal{F}$, $\sum_{\text{supp} E_i \in \mathcal{F}} ||e_i|| \leq |\mathcal{F}|_p$. For analyses using only the weakest assumptions one can add the notion of loss. $\mathcal{E}$ has error rate p and loss q if $\mathcal{E} = \mathcal{E}_p + \mathcal{E}_q$, where the maximum amplitude of $\mathcal{E}_q|\psi\rangle$ is bounded by q and $\mathcal{E}_p$ has error rate p .

Note that the error rate measure defined above is only useful if the total sum of the amplitudes is near one. Other definitions may prove to be more generally applicable. Alternatively, one can use an "optimistic" approach and assume that all the errors are dissipated (i.e. they have orthogonal environments with no possibility of further interference). This is a strong assumption which can be used to complement locality and independence (see below). Although this approach is very unrealistic and fails to yield worst case results, dissipation may be a good approximation in practice. The combination of nearly random error and active efforts to dissipate error as rapidly as possible may effectively enforce the assumption. The extent to which this argument holds can be determined by experiment, once non-trivial quantum devices are technically feasible.

2.5 Independence of Operations

Assumption 2.5. *Independent operations: There are no unintended dependencies between the error amplitudes of distinct operations.*

For parallel operations this has already been expressed by the locality assumption. The independence assumption is therefore significant primarily for sequential operations. In the error basis representations, it means that if the current state of the computation is described as $\mathcal{E}|\psi\rangle$, where $\mathcal{E}$ is an error operator and $|\psi\rangle$ is the intended state, then the error operator associated with the next operation does not depend on the summand of $\mathcal{E}|\psi\rangle$. Equivalently, the environments associated with the next errors live in independent systems from those associated with $\mathcal{E}$. Or, in terms of superoperators, two sequential operations can be represented as the composition of two superoperators (in which case the assumption is implicit in the formalism).

There are of course intentional dependencies, particularly in recovery operations performed for error correction. In this case an operation is chosen depending on the outcome of a measurement. A similar situation arises in implementing certain operations fault tolerantly [15, 10]. For analysis it is convenient to limit the extent to which intentional dependencies are exploited as much as possible.

The purpose of the independence assumption in proving fault tolerance properties is to allow arbitrary conditioning on different types of errors. Without an independence assumption, the associated amplitudes can be difficult to estimate. Of course the independence assumption does not need to hold in the strictest sense, provided the bounds required for the fault tolerance calculations still hold. Most of these bounds depend only on conditioned amplitudes and do not require exact independence otherwise. Thus a weak version of the independence assumption requires only that the error amplitude of an operation is bounded by a known (small) value regardless of which summand of the previous error operator representation of the state it is applied to. An intermediate version of the assumption requires that dependencies only occur locally, due to previous errors in the qubits targeted by an operation. Of course, error amplitudes must still be appropriately bounded for each possible event.

2.6 Dissipated Error

Assumption 2.6. *Dissipated error: Error amplitudes are orthogonal.*

This is a very strong assumption and if made allows using essentially probabilistic arguments about error events. The extent to which the assumption is physical depends critically on how the different types of errors are represented. The assumption can be viewed as making an implicit Markov approximation for how errors can occur, where the types of “quantum jumps”

that can occur is given by the error basis used in the representation. From a theoretical perspective, it can be useful to make this assumption to confirm that a given method works, at least in principle. However, we do not believe that the assumption is sufficiently realistic to give strong error correction or fault tolerance results. Another difficulty is that in many cases (such as the fault tolerant methods of Shor [15]), the dissipation assumption can not be maintained due to the fact that some operations that need to be applied do not commute with the traditional unitary error bases.

2.7 Perfect Classical Measurement

Assumption 2.7. *Perfect measurement in the classical basis: The outcome of a measurement in the classical basis has the correct distribution and side effects on systems not measured.*

This is a convenient assumption that can easily be weakened by taking advantage of the locality, independence and classical computing assumptions. In essence the outcome of a measurement should be equivalent to the (independent and local) errors introduced by a memory operation followed by the correct superoperator representing the measurement. Furthermore, the future evolution of the systems being measured should have no further effect on the remaining systems (i.e. a measured system's state is dissipated). This is already implied by the locality and independence assumptions, provided that no further actions are performed which might cause interference from residual coherence due to an incomplete measurement. Except for the final measurement at the end of a computation, measurements are performed for the purpose of error correction, to implement an otherwise inaccessible operation, or for efficiency. These measurements determine some of the further actions taken by the classical controller.

2.8 Error Rates for Accessible Operations

Assumption 2.8. *Accessible operations: A set of primitive operations can be implemented with sufficient accuracy.*

In the simplest case, the set of primitive operations includes any two qubit unitary operation. However, it has been shown that there are small sets of unitary operations which suffice [1, 6, 15]. For example, the Hadamard transform together with controlled not, some conditional phase shifts and the Toffoli gate are sufficient according to [15]. A very simple set is given

by the Hadamard transform and the controlled phase shift of $|1\rangle$ by i [10]. Useful alternatives include various controlled phase shifts, controlled-not, generation of $|0\rangle + |1\rangle$ and measurement of $|0\rangle + |1\rangle$ versus $|0\rangle - |1\rangle$ [10]. One of the issues not yet resolved is the approximation complexity of arbitrary unitary two qubit operations by these small generating sets.

The accuracy requirement on the set of primitive operations depends on the method for fault tolerant computing. Accuracy can be measured either in terms of fidelity amplitude (worst case on pure states, worst case on entangled states, or for the uniformly entangled state [8, 11]), or by the maximum error amplitude of the error component of the true operator in an error basis representation. For qubits (or any constant arity), all these measures of accuracy are related by constants. The true operator need not always be the same in each application of a specific gate, but the error amplitudes must be uniformly bounded for all applications and consistent with the locality and independence assumptions.

3 Discussion

We have given the critical assumptions that are needed to obtain strong error bounds for fault tolerant computing methods and have suggested suitable weakenings in each case which might suffice in practice. From an experimental point of view, the no-leakage, locality, independence and operational accuracy assumptions are the most important. The question is to what extent these assumptions are experimentally satisfiable. It is a notable fact that none of the current proposals for qubits and quantum gates satisfy the strong forms of the assumptions. Even the weak forms are not generally satisfied. For example, in the ion trap device [5], operations are necessarily dependent via the phonon mode. This may be fixed by either encoding the phonon mode (which seems difficult, however see [4]), or by dissipating residual phonons after each operation. This same dependency makes difficult the parallel implementation of two-qubit operations. The extent of leakage depends on the transition amplitudes to levels not involved in the computation. Note that there are additional levels that are used explicitly for implementing certain gates. Leakage into those levels is a potential problem. Many of these difficulties are intrinsic to the proposed methods for using ion traps and are present even if everything works perfectly.

4 Acknowledgements

This work was performed under the auspices of the U.S. Department of Energy under Contract No. W-7405-ENG-36.

References

- [1] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter. Elementary gates for quantum computation. Submitted to Physical Review A, 1995.
- [2] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed state entanglement and quantum error-correcting codes. quant-ph/9604024, 1996.
- [3] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. quant-ph/9512032, 1995.
- [4] J. Cirac, T. Pellizzari, and P. Zoller. Enforcing coherent evolution in dissipative quantum dynamics. available on quant-ph, 1996.
- [5] J. Cirac and P. Zoller. Quantum computations with cold trapped ions. *Physical Review Letters*, 74:4091-, 1995.
- [6] D. Deutsch, A. Barenco, and A. Ekert. Universality in quantum computation. Submitted to Proc. R. Soc. Lond., 1995.
- [7] R. B. Griffiths and C-S Niu. Semiclassical fourier transform for quantum computation. quant-ph/9511007, 1995.
- [8] E. Knill and R. Laflamme. A theory of quantum error correcting codes. Technical Report LAUR-96-1300, Los Alamos National Laboratory, 1995.
- [9] E. Knill and R. Laflamme. Concatenated quantum codes. In progress, 1996.
- [10] E. Knill, R. Laflamme, and W. Zurek. Concatenated codes for fault tolerant computing. Preliminary report, 1996.
- [11] B. Schumacher. Sending entanglement through noisy quantum channels. Preprint, 1996.

- [12] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35'th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE Press, 1994.
- [13] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. Full version, 1995.
- [14] P. W. Shor. Scheme for reducing decoherence in quantum computer memory. Preprint, 1995.
- [15] P. W. Shor. Fault-tolerant quantum computation. quant-ph/9605011, submitted to FOCS 1996, 1996.
- [16] A. Steane. Multiple particle interference and quantum error correction. Submitted to Proc. Roy. Soc. Lond. A, 1995.