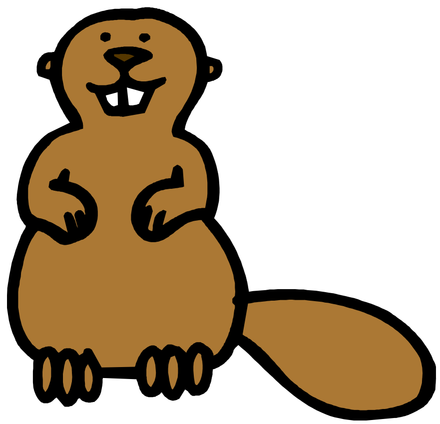
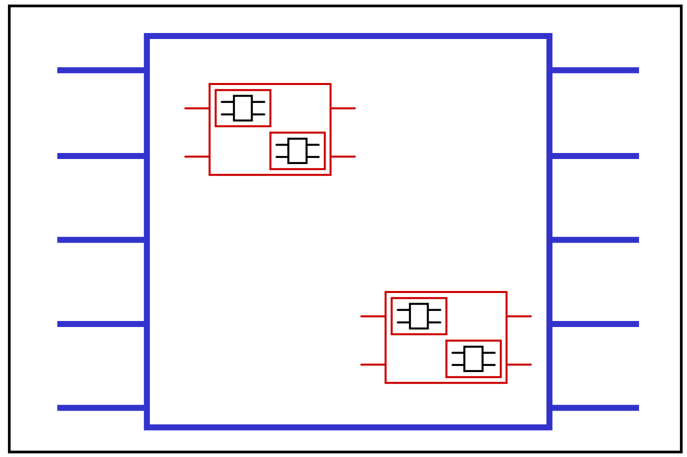
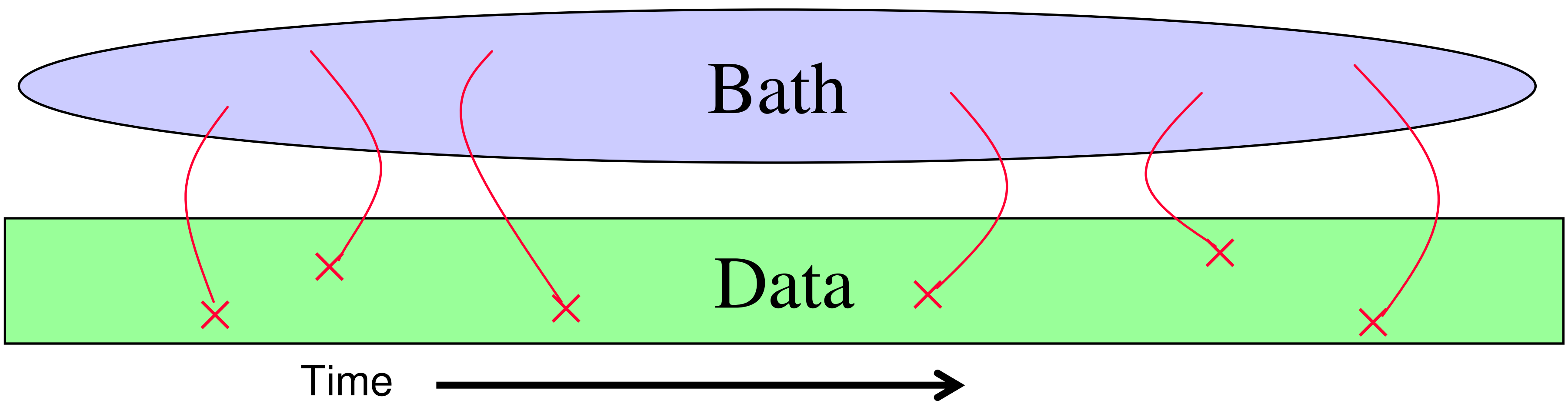
Fault-tolerant quantum computation







Institute for Quantum Information
Sponsored by the NSF

John Preskill, Caltech
WQIS, 23 April 2009

Quantum fault tolerance

The goal of fault-tolerant quantum computing is to operate a large-scale (quantum) computer reliably, even though the components of the computer are noisy.

Reliability can be enhanced by encoding the computer's state in the blocks of a quantum error-correcting code. Each "logical" qubit is stored nonlocally, shared by many physical qubits, and can be protected if the noise is sufficiently weak and also sufficiently weakly correlated in space and time.

Two central questions are:

1) For what noise models does fault-tolerant quantum computing work effectively?

2) For a given noise model, what is the overhead cost of simulating an ideal quantum computation with noisy hardware?

Quantum fault tolerance

To *really* operate a large-scale quantum computer, many implementation-specific systems engineering issues will need to be addressed.

I am a theoretical physicist, not an engineer, yet I have devoted much of my research effort since 1995 to quantum fault tolerance, because I believe that this subject raises questions and stimulates insights that are of broad and fundamental interest in quantum information science.

And .. whatever the applications turn out to be, the quest for a large-scale quantum computer is (in my opinion) one of the grand scientific challenges of the 21^{st} century. Will we be able to overcome the debilitating effects of decoherence and realize subtle interference phenomena in systems with many degrees of freedom? If so, these systems are bound to behave in ways that will surprise and delight us.

Quantum fault tolerance

Quantum fault tolerance overlaps strongly with other topics that are being discussed at this workshop.

In topological quantum computing, quantum gates are protected from noise at the *physical* level. Here, too, the key notion is that by encoding quantum information nonlocally (in the fusion spaces of many nonabelian anyons) it can be protected against damage due to local noise. [Freedman talk]

Methods from quantum control theory (e.g., dynamical decoupling, noiseless subsystems) can also reduce the damaging effects of noise, though these methods do not suffice by themselves to ensure scalability. [Whaley talk]

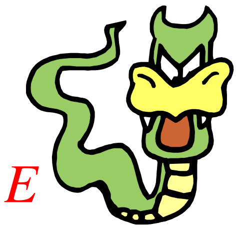
The elements of quantum fault tolerance guide the design of quantum computer architectures. [Chuang talk]

Fault-tolerant quantum computation

1. Quantum error-correcting codes
2. Fault tolerance
3. Quantum accuracy threshold theorem
4. Noise models
5. Some useful ideas:
    a) gate teleportation
    b) subsystem codes
    c) topological codes
    d) message passing
    e) codes for biased noise
    f) protected devices
6. Questions

Errors

The most general type of error acting on n qubits can be expressed as a unitary transformation acting on the qubits and their environment:



$$U : |\psi\rangle \otimes |0\rangle_E \to \sum_a E_a |\psi\rangle \otimes |a\rangle_E$$

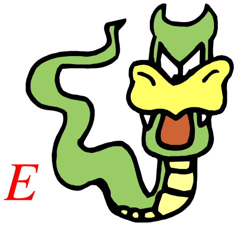
The states $|a\rangle_E$ of the environment are neither normalized nor mutually orthogonal. The operators $\{E_a\}$ are a basis for operators acting on n qubits, conveniently chosen to be "Pauli operators": $\{I, X, Y, Z\}^{\otimes n}$,

where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The errors could be "unitary errors" if $|a\rangle_E = C_a |0\rangle_E$ or decoherence errors if the states of the environment are mutually orthogonal.

Errors



$$U : |\psi\rangle \otimes |0\rangle_E \rightarrow \sum_a E_a |\psi\rangle \otimes |a\rangle_E$$

Our objective is to recover the (unknown) state $|\psi\rangle$ of the quantum computer. We can't expect to succeed for arbitrary errors, but we might succeed if the errors are of a restricted type. In fact, since the interactions with the environment are *local*, it is reasonable to expect that the errors are not too strongly correlated.

Define the "weight" w of a Pauli operator to be the number of qubits on which it acts nontrivially; that is $X, Y,$ or Z is applied to w of the qubits, and I is applied to n-w qubits. If errors are rare and weakly correlated, then Pauli operators E_a with large weight have small amplitude $\| |a\rangle_E \|$.

Error recovery

We would like to devise a recovery procedure that acts on the data and an *ancilla*:

$$V : E_a |\psi\rangle \otimes |0\rangle_A \rightarrow |\psi\rangle \otimes |a\rangle_A$$

which works for any $E_a \in \{\text{Pauli operators of weight} \le t\}$.

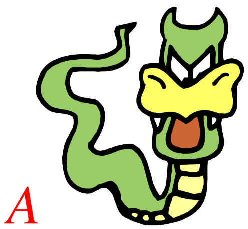
Then we say that we can "correct t errors" in the block of n qubits. Information about the error that occurred gets transferred to the ancilla and can be discarded:

$$|\psi\rangle \otimes |0\rangle_E \otimes |0\rangle_A \xrightarrow{\text{error}} \sum_a E_a |\psi\rangle \otimes |a\rangle_E \otimes |0\rangle_A$$

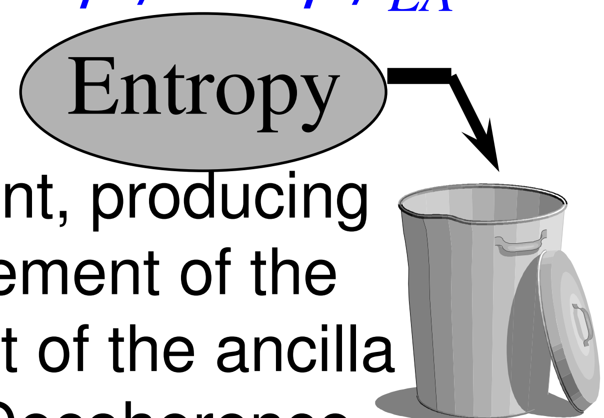
$$\xrightarrow{\text{recover}} \sum_a |\psi\rangle \otimes |a\rangle_A \otimes |a\rangle_E = |\psi\rangle \otimes |\varphi\rangle_{EA}$$

Error recovery

$$|\psi\rangle \otimes |0\rangle_E \otimes |0\rangle_A \xrightarrow{\text{error}} \sum_a E_a |\psi\rangle \otimes |a\rangle_E \otimes |0\rangle_A$$



$$\xrightarrow{\text{recover}} \sum_a |\psi\rangle \otimes |a\rangle_A \otimes |a\rangle_E = |\psi\rangle \otimes |\varphi\rangle_{EA}$$



Errors entangle the data with the environment, producing *decoherence*. Recovery transforms entanglement of the data with the environment into entanglement of the ancilla with the environment, "purifying" the data. Decoherence is thus reversed. Entropy introduced in the data is transferred to the ancilla and can be discarded --- we "refrigerate" the data at the expense of "heating" the ancilla. If we wish to erase the ancilla (cool it to $T \approx 0$, so that we can use it again) we need to pay a power bill.

Quantum error-correcting code

We won't be able to correct all errors of weight up to t for arbitrary states $|\psi\rangle \in \mathfrak{H}_{n \text{ qubits}}$. But perhaps we can succeed for states contained in a *code subspace* of the full Hilbert space,

$$\mathfrak{H}_{\text{code}} \in \mathfrak{H}_{n \text{ qubits}}.$$

If the code subspace has dimension 2^k, then we say that k encoded qubits are embedded in the block of n qubits.

How can such a code be constructed? It will *suffice* if

$$\left\{ E_a \mathfrak{H}_{\text{code}}, \quad E_a \in \left\{ \text{Pauli operators of weight} \le t \right\} \right\}$$

are mutually orthogonal ("nondegenerate code").

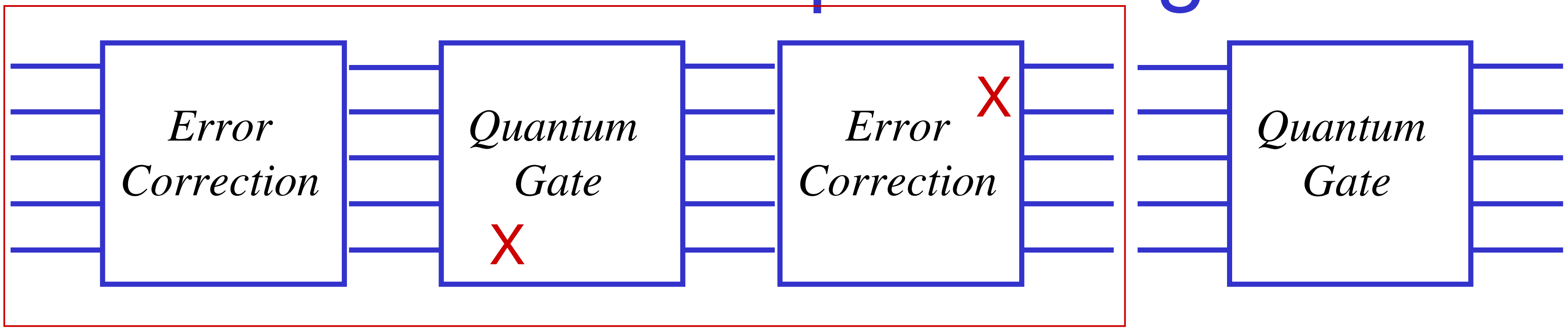
If so, then it is possible in principle to perform an (incomplete) orthogonal measurement that determines the error $\boldsymbol{E_a}$ (without revealing any information about the encoded state). We recover by applying the unitary transformation $\boldsymbol{E_a^{-1}}$.

Fault tolerance

In principle, quantum error-correcting codes allow us to recover from the damage due to errors with low weight. But the recovery operation is itself a quantum computation. Will the recovery really work if the quantum gates that we use to recover from error are themselves noisy?

Furthermore, we need to do more than just *store* a quantum state with high fidelity; we also need to *process* the information protected by the code. How do we devise a universal set of quantum gates that act on the encoded quantum states, without inflicting irreversible damage on the data?

Fault-tolerant quantum gates



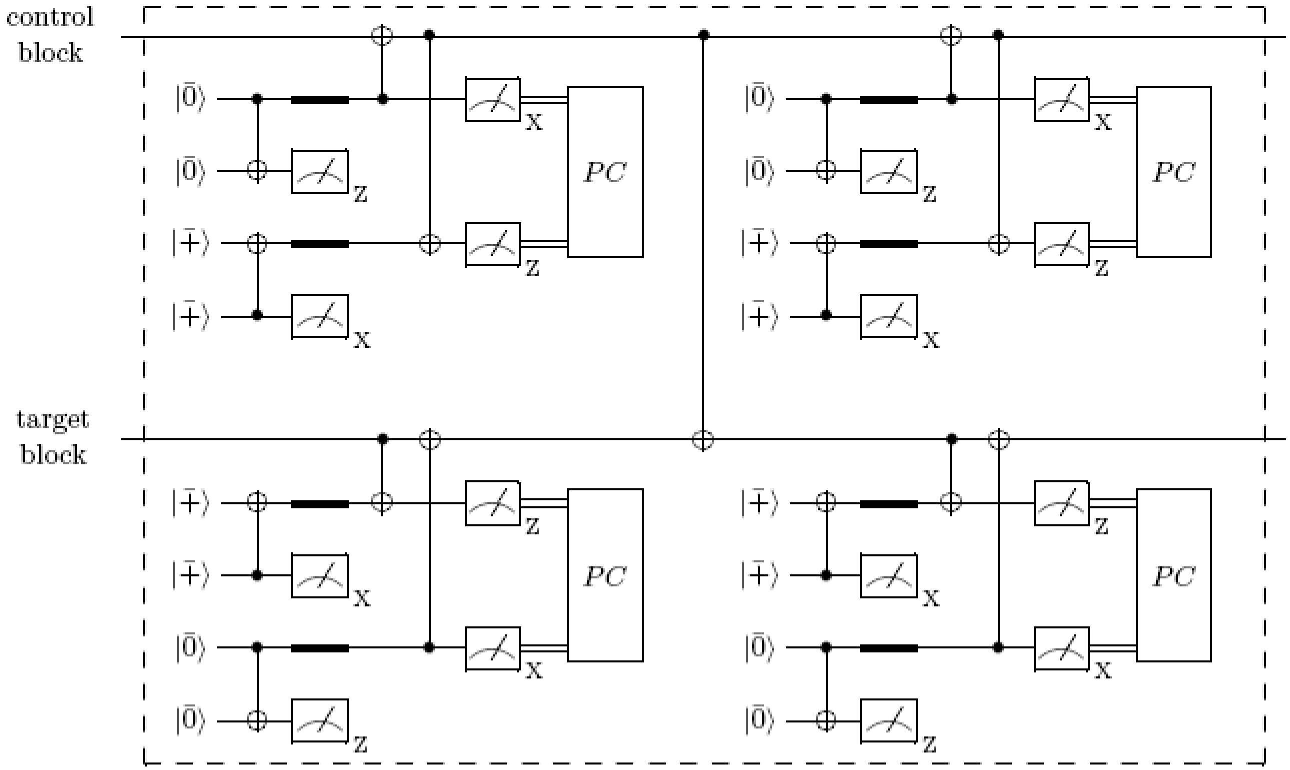
Each gate is followed by an error correction step. The circuit simulation fails only if two faults occur in some “extended rectangle.”

If we simulate an ideal circuit with L quantum gates, and faults occur independently with probability ε at each circuit location, then the probability of failure is

$$P_{\mathrm{fail}} \leq L A_{\mathrm{max}} \varepsilon^2$$

where A_{max} is an upper bound on the number of (malignant) pairs of circuit locations in each extended rectangle. Therefore, by using a quantum code that corrects one error and fault-tolerant quantum gates, we can improve the circuit size that can be simulated reliably to L=O(ε^{-2}), compared to L=O(ε^{-1}) for an unprotected quantum circuit.

Example: CNOT extended rectangle for a 7-qubit code



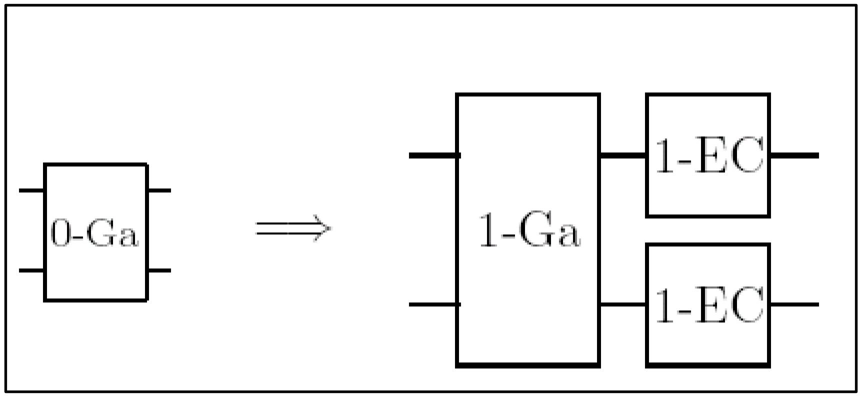
4 ×142 + 7 = 575 locations

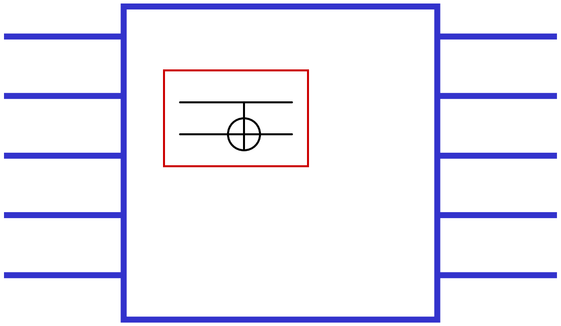
165,025 *pairs* of locations

35,235 *malignant pairs* of locations

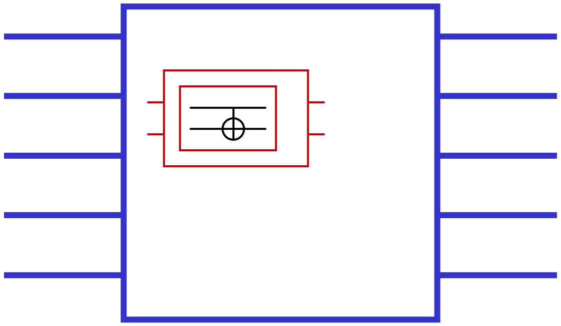
Recursive simulation

In a fault-tolerant simulation, each (level-0) ideal gate is replaced by a *1-gadget*: a (level-1) gate gadget followed by (level-1) error correction on each output block. In a level-k simulation, this replacement is repeated k times --- the ideal gate is replaced by a k-*gadget*.

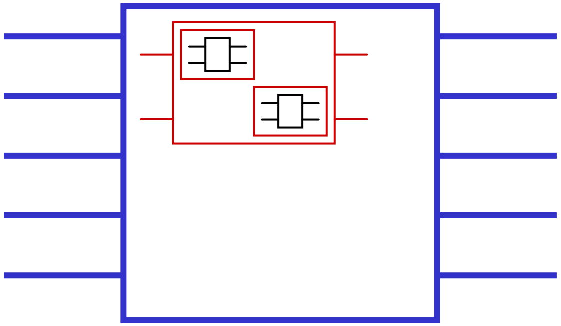




A *1-gadget* is built from quantum gates.



A *2-gadget* is built from 1-gadgets.



A *3-gadget* is built from 2-gadgets.

The effective noise for the level-1 gadget has a "renormalized" strength: $\varepsilon^{(1)} \le \varepsilon^2 / \varepsilon_0 = \varepsilon_0 (\varepsilon / \varepsilon_0)^2$

At level k the renormalized noise strength is: $\varepsilon^{(k)} < \varepsilon_0 (\varepsilon / \varepsilon_0)^{2^k}$

Accuracy Threshold

Quantum Accuracy Threshold Theorem: Consider a quantum computer subject to local stochastic noise with strength ε. There exists a constant $\varepsilon_0 > 0$ such that for a fixed $\varepsilon < \varepsilon_0$ and fixed $\delta > 0$, any circuit of size L can be simulated by a circuit of size L^* with accuracy greater than 1-δ, where, for some constant c,

$$L^* = O\left[L\left(\log L\right)^c \right]$$

Aharonov, Ben-Or
Kitaev
Laflamme, Knill, Zurek
Aliferis, Gottesman, Preskill
Reichardt

assuming:

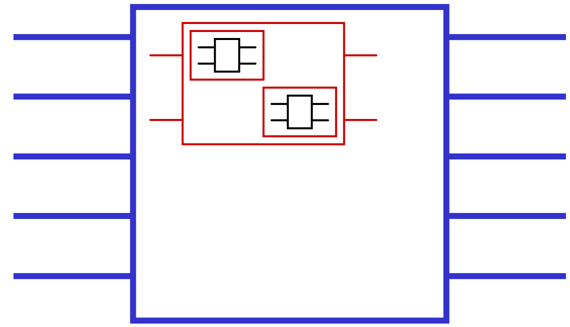
parallelism, fresh qubits (*necessary* assumptions)

nonlocal gates, fast measurements, fast and accurate classical processing, no leakage (*convenient* assumptions).

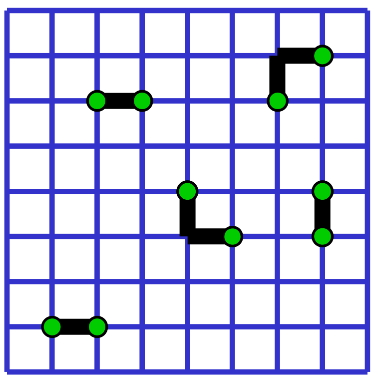
The numerical value of the *accuracy threshold* ε_0 is of practical interest …

Accuracy Threshold

Accuracy threshold theorems have been proved for three types of fault-tolerant schemes:

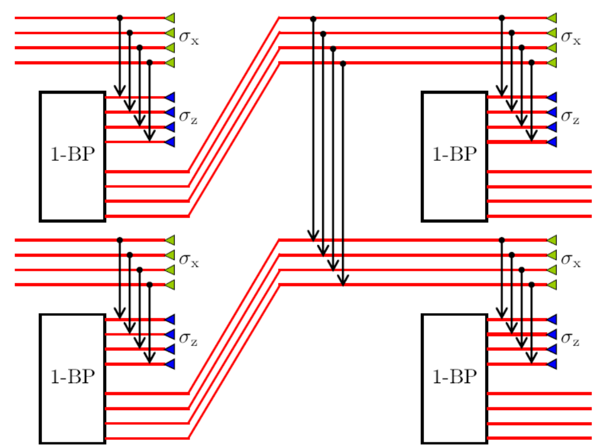


Recursive: hierarchy of gadgets within gadgets, with logical error rate decreasing rapidly with level.



Topological: check operators are local on a two-dimensional surface, and detect the *boundary* points of error *chains*. Logical error rate decays exponentially with block's linear size.

Teleported: Encoded Bell pairs are prepared recursively, but used only at the top level. The (quantum) depth blowup of the simulation is a constant factor.



Accuracy Threshold

Estimates of the numerical value of the quantum accuracy threshold estimates have been based on three types of analysis.

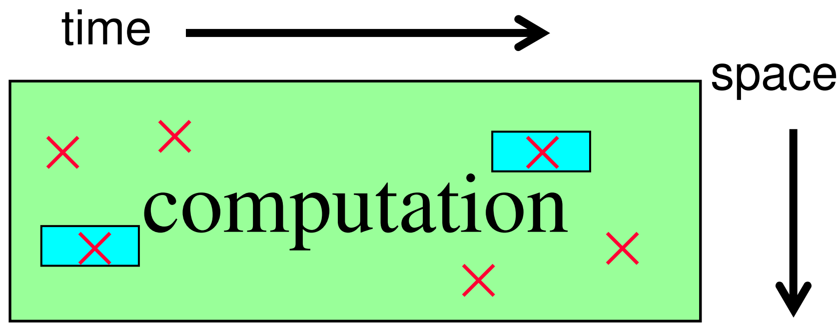
Numerical simulation: Simulate a stochastic noise model and estimate the probability of gadget failure. Gives the most optimistic threshold estimates, but may not be fully trustworthy, and in any case can be applied only to simple noise models.

Rigorous proof: Prove that quantum computing is scalable for a class of noise models. Gives the most pessimistic threshold estimates, but trustworthy and applicable to noise models not easily amenable to simulation.

Hybrid methods: Analytic estimate based on assumptions that some effects can be safely neglected. Typically yields intermediate values for the accuracy threshold.

Noise models

Two types of noise models are most commonly considered in rigorous estimates of the accuracy threshold.

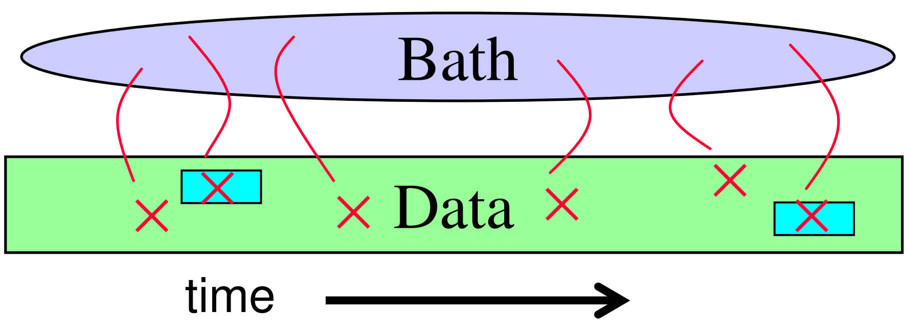


In the local stochastic noise model, "fault paths" are assigned *probabilities*. For any set of r gates in the circuit, the probability that all r of the gates have faults is no larger than ε^r.

The threshold theorem shows that fault-tolerance works for $\varepsilon < \varepsilon_0$. Though not fully realistic, these models provide a useful caricature of noise in actual devices, and can be compared with simulations.

In more realistic Hamiltonian noise models, fault paths can add *coherently*. The joint dynamics of the system and "bath" is determined by a Hamiltonian

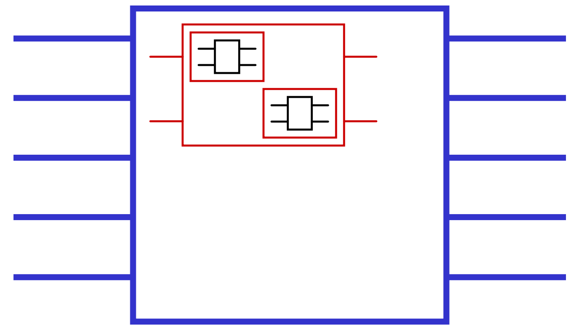
$$H = H_{System} + H_{Bath} + H_{System-Bath}$$



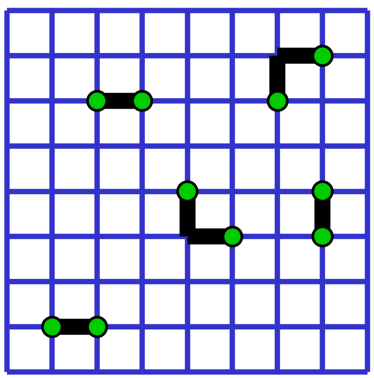
that acts *locally* on the system. Fault tolerance works if the system-bath coupling responsible for the noise is sufficiently weak.

Accuracy Threshold

Some threshold estimates for stochastic noise:

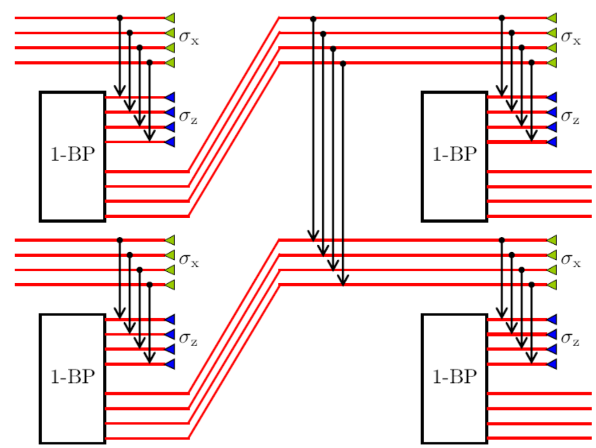


Recursive: $\varepsilon_0 > 1.94 \times 10^{-4}$ proven for local stochastic noise using "Bacon-Shor codes.
-- Aliferis, Cross



Topological: $\varepsilon_0 \sim 7.5 \times 10^{-3}$ estimated for independent *depolarizing noise* in a *local* two-dimensional measurement-based scheme (combination of numerics and analytic argument).
-- Raussendorf, Harrington, Goyal

Teleported: $\varepsilon_0 > 6.7 \times 10^{-4}$ proven for local stochastic noise using concatenated error-detecting codes ($\varepsilon_0 > 1.25 \times 10^{-3}$ for depolarizing noise); simulations indicate $\varepsilon_0 \sim 1 \times 10^{-2}$ for depolarizing noise.
-- Knill; Aliferis, Preskill



Local coherent noise

Terhal, Burkard
Aliferis, Gottesman, Preskill
Aharonov, Kitaev, Preskill

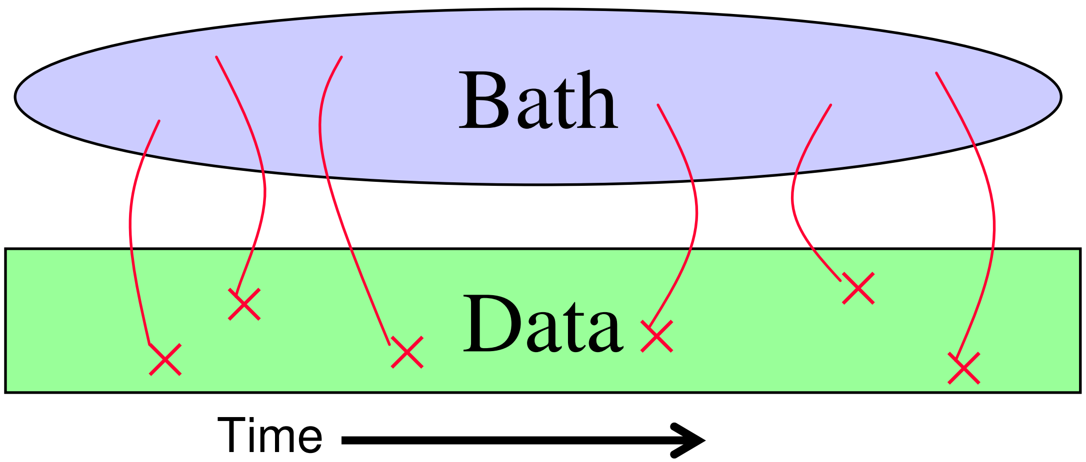
Non-Markovian noise with a *nonlocal* bath.

$$H = H_{System} + H_{Bath} + H_{System-Bath}$$

where

$$H_{System-Bath} = \sum_{\substack{\text{terms } a \text{ acting} \\ \text{locally on the system}}} H^{(a)}_{System-Bath}$$

and



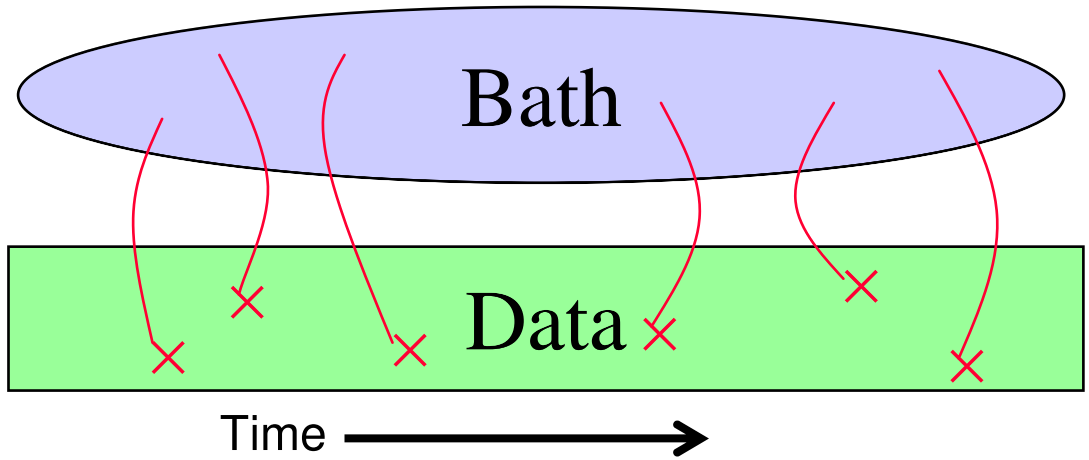
Threshold condition becomes:

$$\boxed{\left\| H^{(a)}_{System-Bath} \right\| t_0 < \varepsilon_0 \approx 10^{-4}}$$

(where t_0 is the time to execute a gate). Really a condition on the error amplitude (probably overly pessimistic).

Quantum error *correction* works as long as the coupling of the system to the bath is *local* (only a few system qubits are jointly coupled to the bath) and *weak* (sum of terms, each with a small norm). Arbitrary (nonlocal) couplings among the bath degrees of freedom are allowed.

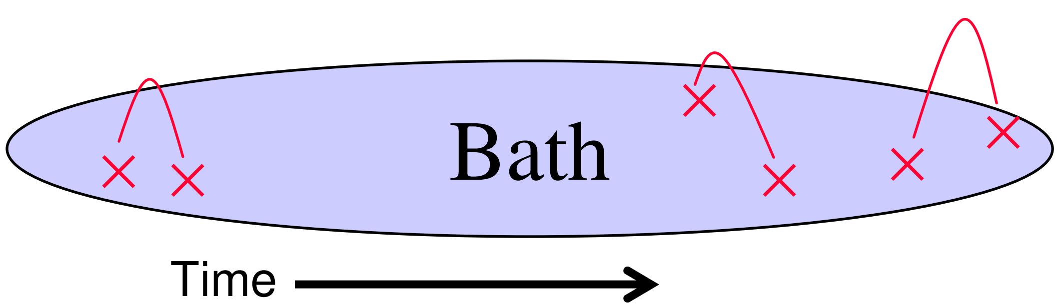
Local non-Markovian noise



$$\left\| H^{(a)}_{System-Bath} \right\| t_0 < \varepsilon_0$$

Expressing the threshold condition in terms of the norm of the system-bath coupling has disadvantages --- The norm is not measured in experiments, and e.g., in the case of a bath of harmonic oscillators, the norm is *infinite*.

It is more natural, and more broadly applicable, to express the threshold condition in terms of the *correlation functions* of the bath.



Such results can apply to an oscillator bath, and take into account the properties of the bath.

Gaussian noise model

Ng, Preskill

In the *Gaussian noise model*, each system qubit couples to a bath of harmonic oscillators:

$$H = H_S + H_B + H_{SB} \qquad H_B = \sum_k \omega_k a_k^\dagger a_k$$ (uncoupled oscillators)

$$H_{SB}(t) = \sum_x \sum_\alpha \phi_\alpha(x,t)\sigma_\alpha(x)$$

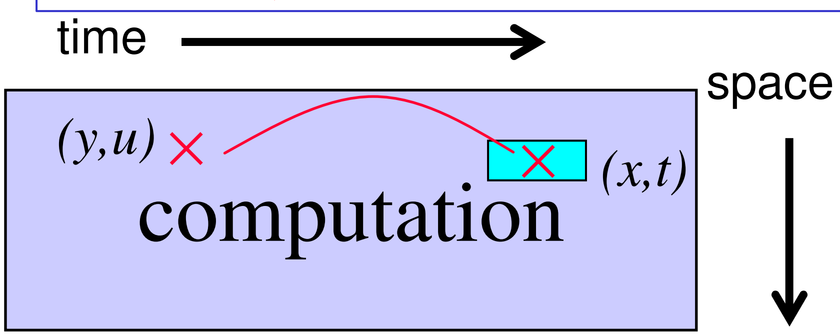
$$\phi_\alpha(x,t) = \sum_k g_{k,\alpha}(x,t)a_k + g_{k,\alpha}(x,t)^* a_k^\dagger$$

(x labels qubit position, ϕ is a Hermitian bath operator, σ is Pauli operator acting on the system qubit, g is a coupling constant.)

If the state of the bath is Gaussian (e.g., thermal), the provable threshold condition can be expressed in terms of the two-point correlation function of the bath variables:

$$\boxed{\varepsilon = \max\left(\int_{\Box} dt \int_{all} du \sum_{y,\alpha,\beta} \left|\langle \phi_\alpha(x,t)\phi_\beta(y,u)\rangle_B\right|\right)^{1/2} < \varepsilon_0}$$

(x,t) is integrated over one circuit "location", while (y,u) is integrated over all of spacetime.



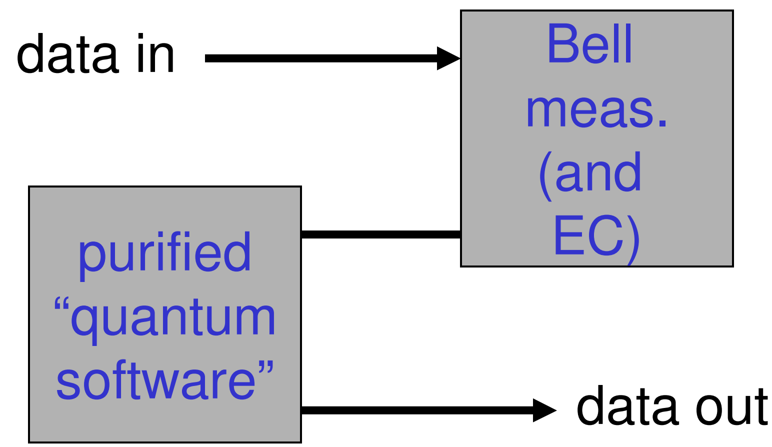
This condition has sensitivity to high-frequency noise that is probably spurious, but it is hard to do better without using specific properties of the system Hamiltonian (i.e., the qubits).

Some useful ideas

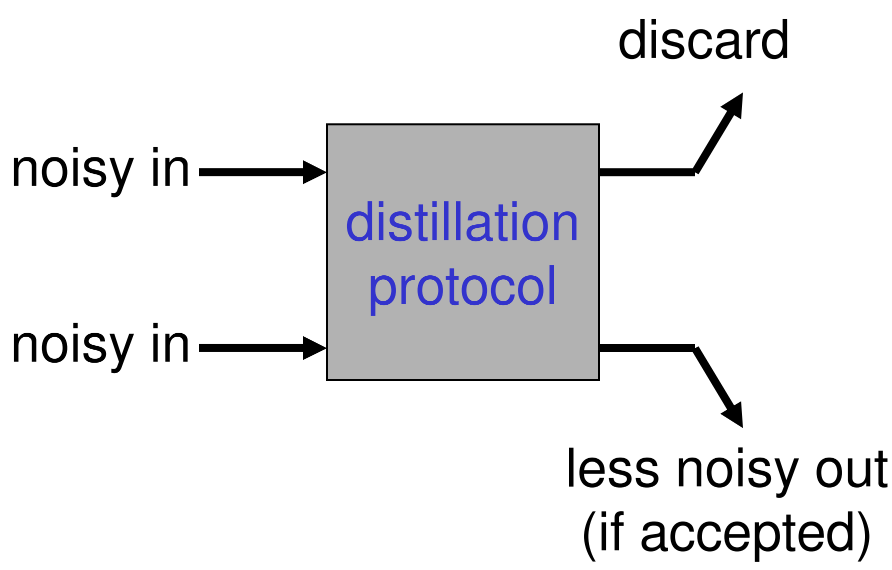
1) Gate teleportation and state distillation
2) Subsystem codes
3) Topological codes
4) Message passing in block decoding
5) Fault-tolerant gates for highly biased noise
6) Protected devices and gates

Gate teleportation and state distillation

In fault-tolerant schemes, a version of quantum teleportation is used to complete a universal set of protected quantum gates. Suitable “quantum software” is prepared and verified offline, then measurements are performed that transform the incoming data to outgoing data, with a “twist” (an encoded operation) determined by the software.



Reliable software is obtained from noisy software via a multi-round state distillation protocol. In each round (which uses CNOT gates and measurements), two noisy copies of the software are compared, and one copy is destroyed. The other copy, if accepted, is less noisy than the input.



Gottesman, Chuang; Knill; Bravyi, Kitaev

Subsystem codes

Hilbert space decomposes as:

$$\mathcal{H} = \bigoplus_{s} \mathcal{H}_{\mathcal{L}}^{(s)} \otimes \mathcal{H}_{\mathcal{T}}^{(s)}$$

syndrome

"logical" subsystem

"gauge" subsystem

A subsystem code is really the same thing as a standard quantum code, but where we don't use some of the k qubits encoded in the code block. These unused qubits are called "gauge qubits" --- we don't care about their quantum state and we don't have to correct their errors.

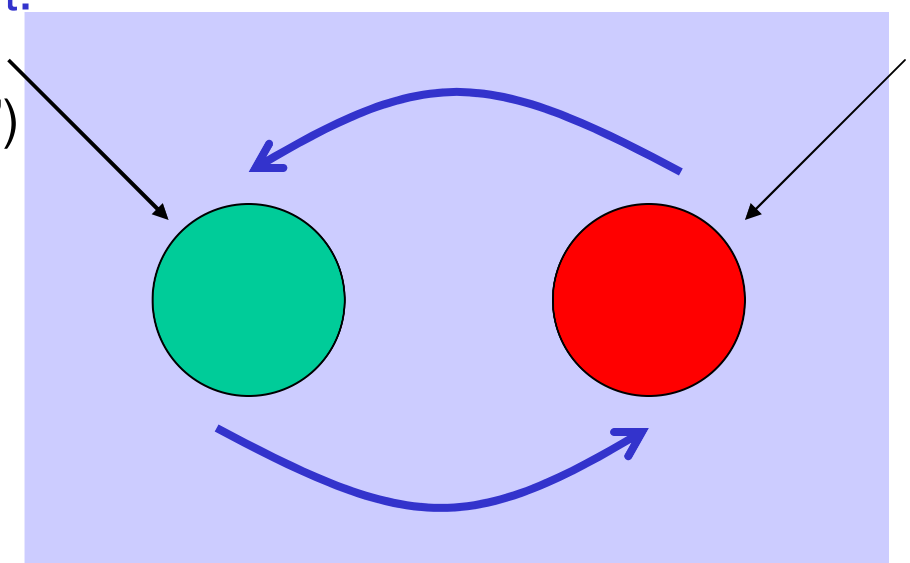
Choosing not to correct the gauge qubits can be surprisingly useful. For one thing, we are free to measure the gauge qubits, and the measurement outcomes can provide useful information about the logical qubits that we really do want to protect.

For example, in the "Bacon-Shor code" it is easier to diagnose the errors by measuring the gauge qubits (measurements of weight-two Pauli operators) than by measuring the standard check operators of the code (measurements of weight-six Pauli operators).This method was used to prove the lower bound on the threshold $\varepsilon_0 > 1.94 \times 10^{-4}$ for a recursive scheme subject to local stochastic noise.

-- Kribs, Laflamme, Poulin; Bacon; Aliferis, Cross

Local fault tolerance with 2D topological codes

Qubits are arranged on a two-dimensional lattice with holes in it. Protected qubits are encoded (in either of two complementary bases) by placing “electric” charges inside “primal” holes or “magnetic” charges inside “dual” holes. The quantum information is well protected if the holes are large and far apart.



A controlled-NOT gate can be executed by “braiding the holes” which is achieved by a sequence of local gates or measurements.

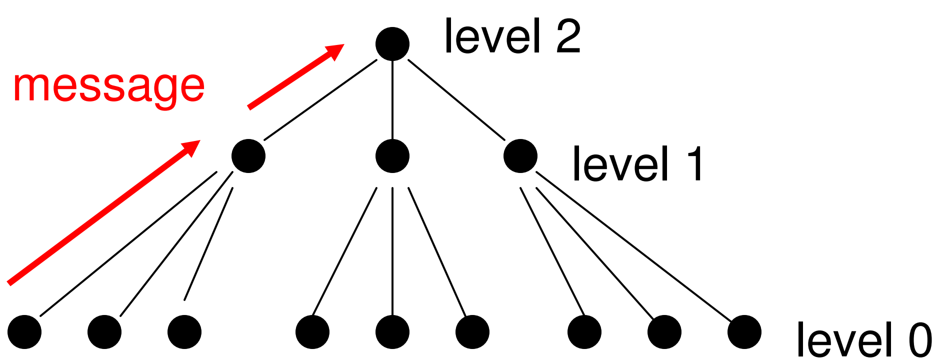
The protected gates and error syndrome extraction can be done with local two-qubit gates or measurements. Numerical studies indicate an upper bound on the threshold for independent depolarizing noise:

$\varepsilon_0 \sim 7.5 \times 10^{-3}$

Raussendorf, Harrington, Goyal
Dennis, Kitaev, Landahl, Preskill

Improved decoding via message passing

The simplest way to decode a concatenated code is “one level at a time”, starting at the lowest level. However, the decoding is more reliable if information about the error syndrome found at lower levels is used to infer the best way to decode at higher levels.



In particular, “message passing” allows a concatenated error-*detecting* code to *correct* errors at higher levels, because an error-detecting code can correct errors that occur at *known* positions in the code block.

This observation is useful because gadgets for error-detecting codes are simpler than gadgets for error-correcting codes, and hence can tolerate stronger noise. These ideas were applied to prove a lower bound on the accuracy threshold for independent depolarizing noise:

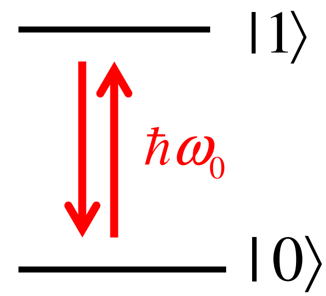
$\varepsilon_0 > 1.25 \times 10^{-3}$

(simulations indicate $\varepsilon_0 \sim 1 \times 10^{-2}$)

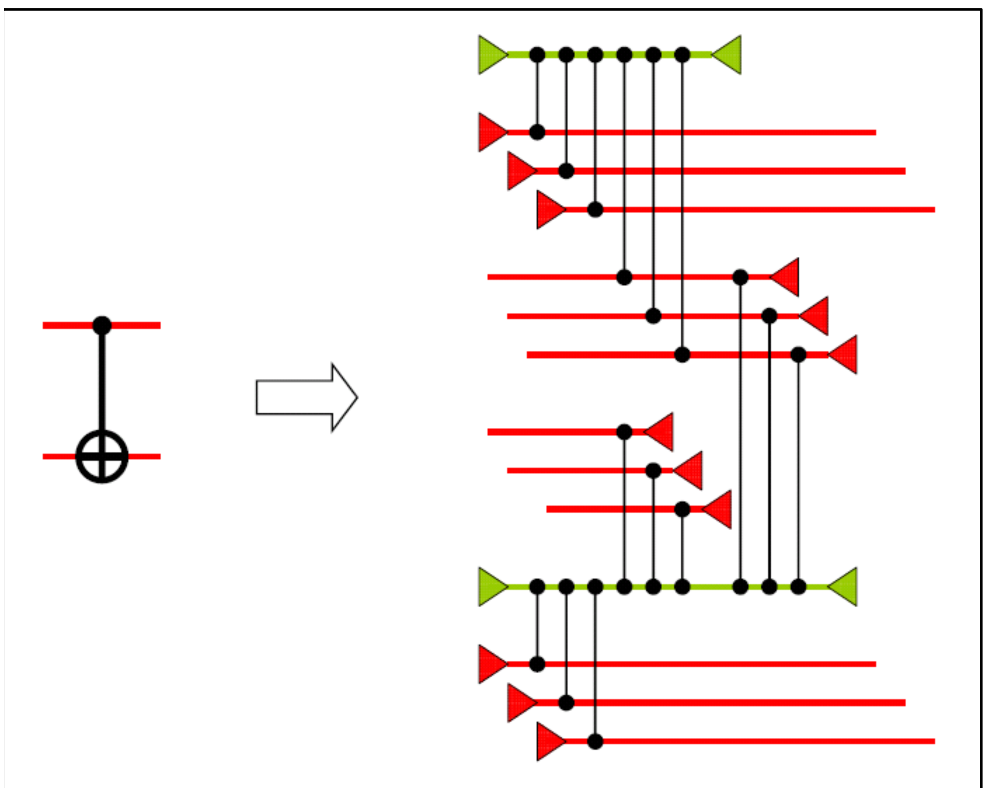
Poulin; Knill; Aliferis, Preskill

Fault-tolerant quantum computing versus biased noise

In many physical implementations of quantum gates, Z noise (dephasing) is stronger than X noise (relaxation). Dephasing arises from low frequency noise, while relaxation arises from noise with frequency $\sim\omega_0$, which is typically much weaker.



Using only controlled-phase two-qubit gates (which are diagonal in the computational basis), plus qubit preparations and measurements, we can do universal fault-tolerant computation with good protection against dephasing. Shown is a circuit that uses these elements to achieve a controlled-NOT gate protected by a phase repetition code. When the noise is highly biased, the threshold improves by about a factor of 5.

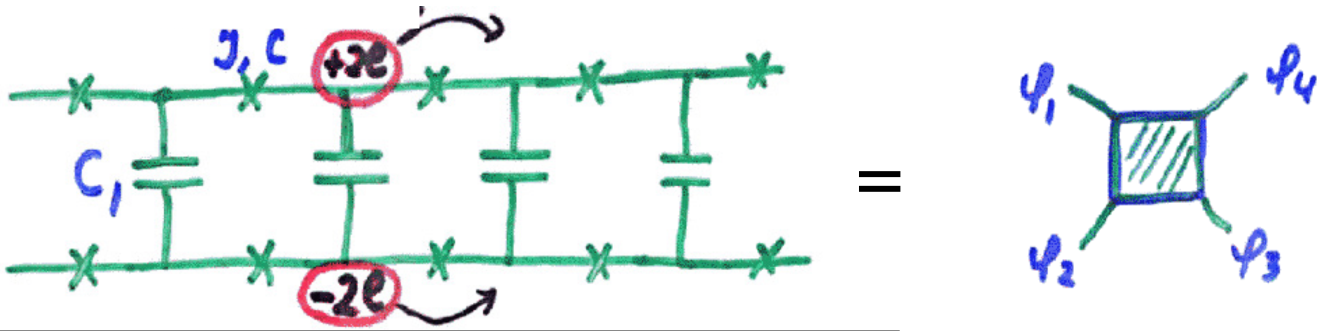


-- Aliferis, Preskill; Brito, Aliferis, et al.

Protected superconducting qubit

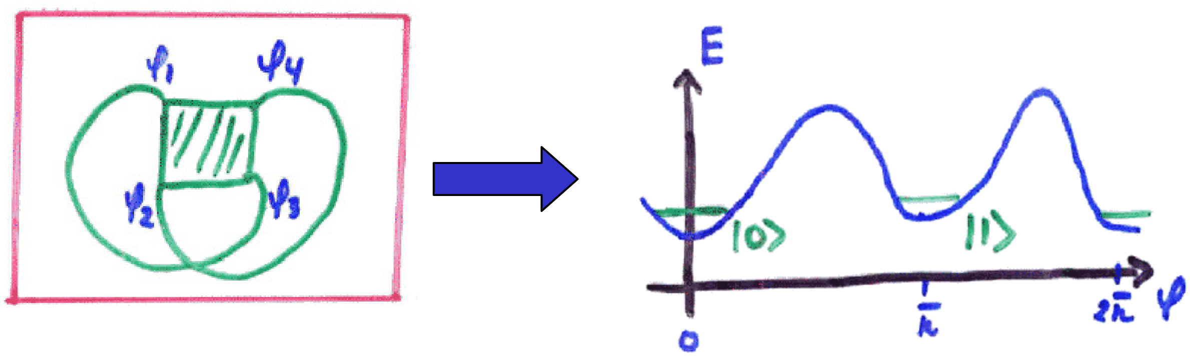
Feigel'man, Ioffe
Kitaev

A superconducting “current mirror” can be realized e.g. using a two-rung ladder of Josephson junctions shunted by capacitors that have much larger capacitance than the intrinsic capacitance of the JJs.



$$\boxed{E = f(\varphi_1 - \varphi_2 + \varphi_3 - \varphi_4) + \text{exp. small}}$$

(The ladder conducts only excitons)



Connecting leads as shown,

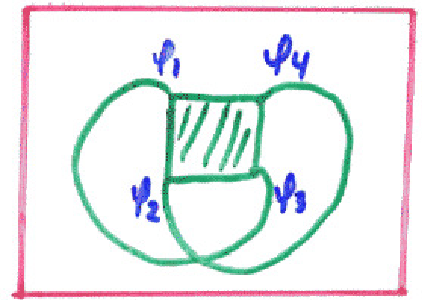
$$\boxed{E \approx f(2(\varphi_1 - \varphi_2))}$$

Two nearly degenerate minima $\longrightarrow$ qubit.

The barrier is high enough to suppress bit flips, and the stable degeneracy suppresses phase errors. Protection arises because the encoding of quantum information is highly nonlocal. Degeneracy is split in an order of perturbation theory linear in length of ladder (compare topological protection).

Protected superconducting qubit

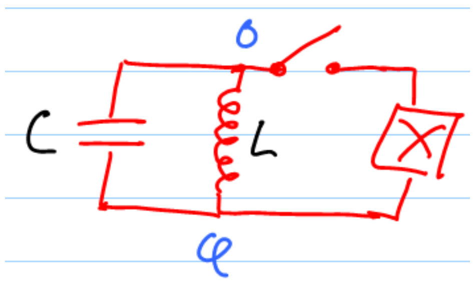
Kitaev



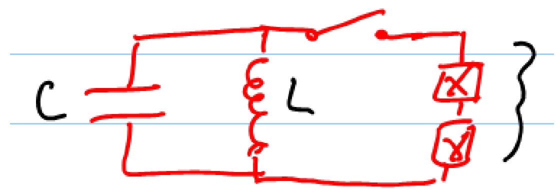
Some gates are also protected: we can execute

$$\boxed{\exp\left(i\frac{\pi}{4}Z\right) \text{ and } \exp\left(i\frac{\pi}{4}Z_1Z_2\right)}$$

with exponential precision. This is achieved by coupling a qubit or a pair of qubits to a "superinductor" with large phase fluctuations:

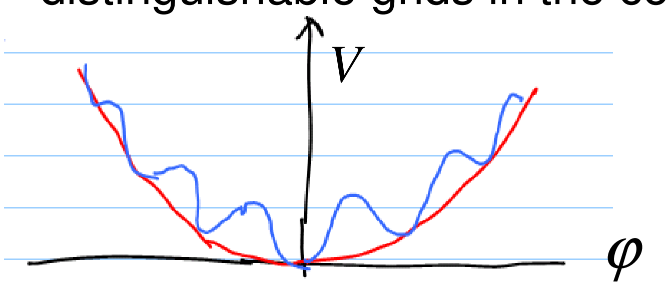


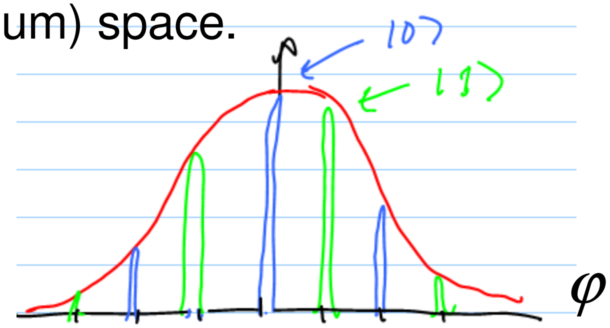
qubit



two qubits

The harmonic wave function of the superinductor evolves adiabatically to a "Gaussian grid state,'' where the $|0\rangle$ and $|1\rangle$ grids differ by a displacement by π. The phase of $|0\rangle$ advances by $2\pi \times (\text{integer})^2$ and the phase of $|1\rangle$ advances by $2\pi \times (\text{integer}+ ½)^2$. The relative phase is robust because states are also distinguishable grids in the conjugate (momentum) space.





Some issues

1) Can we rigorously justify the “error phase randomization hypothesis” (error probability linear in number of gates)?
2) In Hamiltonian noise models, can we further reduce the sensitivity of threshold estimates to high-frequency noise, include non-Gaussian correlations, etc.?
3) Adapting fault-tolerance to properties of noise and specific algorithms.
4) What other schemes are scalable (besides concatenated codes and topological codes)?
5) Optimizing overhead cost.
6) Incorporating spin echo, dynamical decoupling, etc. into analysis of fault-tolerant protocols.
7) Self-correcting systems and devices (“finite-temperature topological order” in fewer than four spatial dimensions).
8) Systems engineering challenges (wires, power, cooling, ..)

Quantum fault tolerance

Operating a large-scale quantum computer will be a grand scientific and engineering achievement.

Judicious application of the principles of fault-tolerant quantum computing will be the key to making it happen.

Fascinating connections with statistical physics, quantum many-body theory, device physics, and decoherence make the study of quantum fault tolerance highly rewarding.