Quantum Refrigerator

Michael Ben Or   Daniel Gottesman   Avinatan Hassidim

October 5, 2012

The threshold theorem [AB97, KLZ98, Kit97, AGP05] is the central result of the theory of fault-tolerant quantum computation. It states that, provided the error rate per gate or time step is below some constant threshold value, then arbitrarily long quantum computations are possible with only polylogarithmic overhead. The threshold theorem tells us that large quantum computers are possible in principle, provided experimentalists can achieve an error rate below the threshold value.

A number of assumptions are needed to prove the threshold theorem. Some of them can be relaxed and some of them cannot be. One assumption that has been traditionally classified as necessary is the need for fresh ancilla qubits in the course of the computation. In the course of quantum error correction, ancilla qubits are introduced and used to record the error syndrome. This can be viewed as a refrigeration process, where entropy which has been introduced into the data qubits by the noise gets pumped out into the ancilla qubits, cooling down the data qubits. In order for this to work, the ancilla qubits used must be cold themselves, or they cannot absorb the extra entropy from the data. Since ancilla qubits created at the beginning of the computation are themselves subject to the noise process, we can expect them to heat up over time, eventually making them worthless for refrigeration. Based on this intuition, we expect to need a continual stream of new, freshly cooled ancilla qubits to keep the error correction running.

At a rigorous mathematical level, this intuition is supported by the result of [ABIN96], which showed that for noise in the form of a depolarizing channel, it is impossible to compute for longer than $O(\log n)$ time. We study more general channels and reach a different conclusion. In particular, we prove the following main theorem:

Theorem 1 (Main Theorem)**.** *Let C be any non-unitary channel close to the identity, and consider quantum computations which suffer from noise C on each qubit at each time step. Up to unitary equivalence, the limit of repeatedly applying C can be either a point or a diameter in the Bloch sphere.*

1. *If the limit is the center of the Bloch sphere, it is possible to compute for a logarithmic time, and no longer. An example of this class is the depolarizing channel.*
2. *If the limit is a diameter, than it is possible to compute for only a polynomial length of time. An example of this class is the dephasing channel.*
3. *If the limit is a point which is not the center of the Bloch sphere, it is possible to compute for an exponential number of time steps. An example of this class is the amplitude damping channel.*

For the third case, we expect that it is not possible to compute for longer than exponential time, but we have not managed to prove this. We have shown upper bounds for the other cases.

The three classes of channels can also be characterized in terms of how they treat the entropy. The depolarizing class causes entropy to increase for all states until it reaches the maximum for the completely mixed state. For the dephasing class of channels, entropy is merely non-decreasing: For some states it increases, and for others it stays the same. For the amplitude damping class of channels, entropy can decrease under the channel.

The result with the greatest practical significance is 3, as it enables quantum computation without fresh qubits. This means that the usual claim that a quantum computation needs fresh ancilla qubits supplied from the outside is not completely correct. Rather, we have shown that when the channel is in the amplitude damping class, then we can co-opt the noise to create fresh ancilla qubits, allowing us to recycle qubits and perform computation even in an otherwise closed system. This doesn't really violate the intuition discussed above, because noise of this form provides some cooling, taking a maximally mixed state to a less mixed state. For instance, if the system is in thermal contact with a cold bath, if a qubit is not affected by the computation for a long time, it will tend to equilibrate to the temperature of the bath. Viewed this way, of course the system is not truly closed, but it is open in only a single limited way.

Moreover, our protocol works for *any* fixed point (other than the maximally mixed state). Therefore, even a very hot bath, provided it is at finite temperature, can provide enough cooling to leverage into a full-blown fault-tolerant computation. Of course, the overhead is very high when the fixed point is close to maximally mixed, but it is perhaps surprising that it can still be done at all. One might instead have expected a threshold temperature above which the system is too hot to compute. There is a threshold *error strength*, but that is a constraint on the coupling to the bath rather than the temperature of the bath.

We now provide some intuition to the proofs of the various parts of the main theorem.

We work in a model in which gates are perfect, but after each time step, each qubit undergoes a single-qubit channel C. Thus, we alternate layers consisting of gates with layers consisting of noise. Every qubit undergoes the same noise at each time step, and there is no memory or correlation between different qubits or different time steps.

The lower bounds (showing that quantum computation is possible for at least a certain amount of time) are proven by allowing arbitrary single-qubit and two-qubit gates in parallel at each time step, as is usual for fault-tolerant constructions. The upper bounds (showing that computation is impossible for more than a certain amount of time) are more generous, and allow arbitrary many-qubit unitaries between the applications of noise at each time step.

- **Limit(C) is the center of the Bloch sphere:** The proof for this case is just a slight generalization of [ABIN96], and uses the same techniques.

- **Limit(C) is a diameter:** By [KR01], the noise is equivalent to the phase damping channel. We can perform a polynomial-length computation by running the computation on m qubits, leaving $n-m$ qubits in the state $|0\rangle$, a state left unchanged by the noise. At each stage of the computation it is possible to take fresh qubits which were unaffected by the noise. Since at most m qubits are needed every step, the computation can go for time $(n-m)/m$.

  Proving the other direction is a little tricky, as it is not clear how fast the entropy rises. We show that it is impossible to store half of an EPR pair for more than polynomial time. To do this, we take advantage of the fact that the entropy cannot decrease. Therefore, if the state is stored for a long time, there must be some time step during which the entropy rises very little. For the dephasing channel, this can only happen if the qubits are essentially unentangled and diagonal in the standard basis. But a completely dephased state has no entanglement with the outside world at all, and therefore we have failed to maintain the EPR pair.

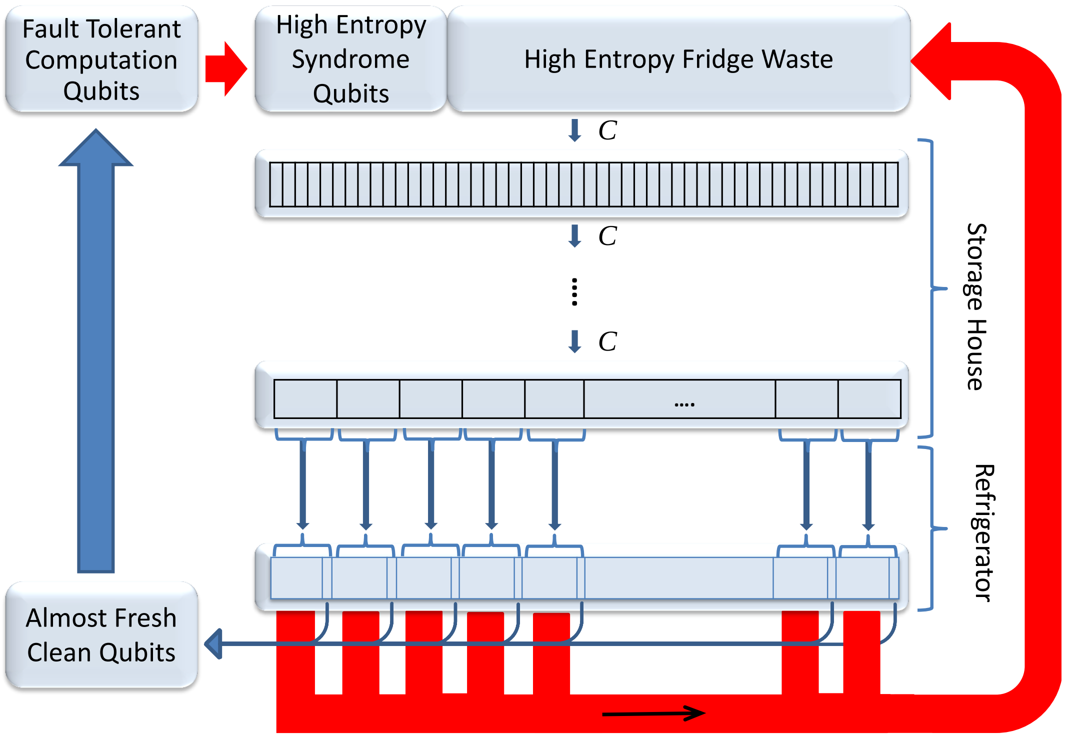

Figure 1: For noise in the amplitude damping class, the computer consists of three components: A computation component, a storage house, and a refrigerator.

- **Limit(C) is a point which is not the center of the sphere:** In this model we divide the system into three parts. The first part is just a standard fault-tolerant computation using n' qubits. At each step we allow this part to use up to n' qubits, and receive the same number of fresh qubits from the other parts.

  The second part of the computation is a storage house. This part holds $n'T$ chunks of data. The size of each chunk is a constant R which depends on the channel, and at each stage of the computation n' chunks are passed to the refrigerator. Qubits stay in the storage house for a time T, which is chosen sufficiently long to get close to the fixed point of C. T depends both on the limit of C and the strength of the noise.

  The last part is the refrigerator, which is actually built of $O(n')$ refrigeration units which work in parallel. Each such unit takes R qubits (where each qubit in the sequence is very similar to the limit of C), and condenses the entropy to $R-1$ qubits and to a fresh qubit. Then the refrigerator passes n' qubits to the computation, which uses as many as it needs. The dirty qubits left over from the refrigerator, the qubits used up by the computation, and any unused fresh qubits are recycled by being sent back to the storehouse. The constant R is a function of the fixed point of C, and is chosen such that R qubits close to the fixed point have total entropy less than $R-1$. The upper bound on the strength of the interaction with the channel (the threshold under which computation is possible) is determined by the usual threshold with fresh ancillas and the time to condense one fresh qubit.

References

[AB97] D. Aharonov and M. Ben-Or, “Fault-Tolerant Quantum Computation With Constant Error,” In *Proceedings of the 29th Annual IEEE Symposyum on the Theory of Computer Science* (STOC’97), ACM, pp. 176-188, 1997.

[ABIN96] D. Aharonov, M. Ben-Or, R. Impagliazzo and N. Nisan “ Limitations of Noisy Reversible Computation,” quant-ph/9611028, 1996.

[AGP05] P. Aliferis, D. Gottesman, and J. Preskill, “Quantum accuracy threshold for concatenated distance-3 codes,” *Quant. Information and Computation* Vol. 6, No. 2, pp. 97–165, quant-ph/0504218, 2006.

[KR01] C. King and M. Beth Ruskai, “Minimal entropy of states emerging from noisy quantum channels,” *IEEE Transactions on Information Theory* Vol. 47, No. 1, pp. 192–209,2001.

[Kit97] A. Yu. Kitaev, “Quantum computations: algorithms and error correction,” *Russian Math. Surveys* Vol. 52, pp. 1191–1249, 1997.

[KLZ98] E Knill, R. Laflamme, W. H. Zurek, “Resilient quantum computation: error models and thresholds,” *Proc. Roy. Soc. London, Ser. A* Vol.454, p. 365, quant-ph/9702058, 1998.