Formula Size Lower Bounds and Quantum States

Scott Aaronson

September 27, 2004

Everyone agrees that building a quantum computer is hard. But some people think it must be impossible for a fundamental physical reason. Two computer scientists who think this are Leonid Levin and Oded Goldreich.

Levin, for example, says that the exponentially small amplitudes that arise in Shor's factoring algorithm are absurd—that they take us far beyond quantum mechanics' regime of demonstrated validity, into a realm where nothing sensible can be said. (Although of course, whatever theory describes that realm must conform to the original Extended Church-Turing Thesis, which we know *a priori* to be true.) Similarly, Goldreich says that Shor states have exponential "non-degeneracy" and therefore should take exponential time to prepare.

About a year ago, I wrote a paper called "Multilinear Formulas and Skepticism of Quantum Computing." The first thing I did there is point out that Levin's and Goldreich's arguments are extremely weak, without specifics about *where* our current understanding of quantum mechanics breaks down. Take exponentially small amplitudes. I can produce probabilities of order 2^{-10000} by flipping a coin 10000 times. Does that mean that the laws of probability theory are absurd, or should break down? More concretely, we *can* prepare states with exponentially small amplitudes:

$$\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes n}.$$

A skeptic would say, yeah, but that's not entangled. But there are recent experiments (e.g. Zeilinger's buckyball experiments) that prepare quite large Schrödinger cat states:

$$\frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}.$$

Just by viewing these states in a different basis, we get a combination of exponentially small amplitudes and entanglement across hundreds of particles.

So what exactly is meant by "non-degeneracy"? What criterion separates the states that have already been demonstrated in the lab, from the states arising in Shor's algorithm? In the paper, I called such a criterion a *Sure/Shor separator*, and I challenged quantum computing skeptics to find one.

But in case the skeptics weren't so eager to take up my challenge, I started doing their work for them. My proposal was that we study the complexity theory, not of languages computable with the help of quantum states, but of *quantum states themselves.* In other words, which pure states of n qubits have succinct classical descriptions of various kinds? If we knew that, then we could have a dialogue between theory and experiment: theorists would say, "I conjecture that all physically possible quantum states belong to this class," experimenters would then test that conjecture by trying to prepare states outside of the class, etc.

In the paper, most of my attention focused on "tree states," which are those states representable by a polynomial-size *tree* of linear combination and tensor product gates. Here's an example:You can take a normalized linear combination of any two states on the same set of qubits, or the tensor product of any two states on disjoint sets of qubits. Every leaf vertex must be labeled with either $|0\rangle$ or $|1\rangle$; then the *size* of a tree is the number of leaf vertices. In particular, tensor product states, Schrödinger cat states, and many other simple quantum states on n qubits can all be shown to have tree size bounded by a polynomial in n.

My main result was the following: let $C \subseteq \{0,1\}^n$ be the set of codewords of a uniform random linear code with (say) $n/2$ generators. Then with high probability over C, the state

$$\frac{1}{\sqrt{|C|}} \sum_{x \in C} |x\rangle$$

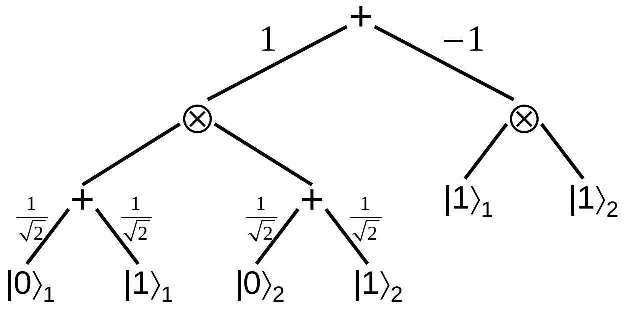

Figure 1: Expressing $(|00\rangle + |01\rangle + |10\rangle - |11\rangle)/2$ by a tree of linear combination and tensor product gates, with scalar multiplication along edges. Subscripts denote the identity of a qubit.

has tree size $n^{\Omega(\log n)}$, which of course is superpolynomial. This result used a very recent technique of Ran Raz for lower-bounding the *multilinear formula size* of functions. Indeed, it's not hard to show that proving a lower bound on the tree size of the state $|\psi\rangle = \sum_x \alpha_x |x\rangle$ is equivalent to proving a lower bound on the multilinear formula size of the function $f(x) = \alpha_x$. I also showed that such states require tree size $n^{\Omega(\log n)}$ even to *approximate* well in L_2 norm. The upshot is that, if we ever prepared such states in the lab, then this would provide evidence against the hypothesis that all states in Nature are tree states. (I conjectured that the states arising in Shor's algorithm also have tree size $n^{\Omega(\log n)}$, but was unable to prove that except by assuming a number theory conjecture.)

After I wrote the paper, some people raised conceptual objections to it—for example, if you *did* prepare a state that provably has superpolynomial tree size, how would you convince a skeptic that you succeeded? More generally, what sort of evidence is needed to establish a state's existence? (My answer was basically that this is like any other empirical question.) Also: how can we go from an asymptotic statement to a statement about physical reality? (My answer: don't let n run off to infinity. Stop it when it reaches 10000.)

But they also raised more specific, technical objections.

First, they complained that my result applies only to a random linear code. What about *explicit* linear codes?

Second, they complained that the tree size notion doesn't seem connected to anything in physics. Are there physical properties of a quantum state that are connected to its having large tree size?

Third, they complained that my lower bound was only $n^{\Omega(\log n)}$ as opposed to exponential. (About this third objection, I might answer: don't blame me, blame Ran Raz! Or rather, admire him for being able to prove even an $n^{\Omega(\log n)}$ lower bound, since formula and circuit lower bounds are among the hardest problems in all of mathematics!)

In the remainder of this talk, I want to report the progress that I made just a few months ago toward answering these technical objections. I think some of the progress might even be of independent interest for *classical* complexity theory.

The first thing I'll show you is that, if you take an equal superposition over the codewords of *any* sufficiently good erasure code, then that will have tree size $n^{\Omega(\log n)}$. I don't really know anything about coding theory. But another Berkeley grad student, Andrej Bogdanov, who does know something, showed me how to get the erasure codes I want by concatenating the Reed-Solomon and Hadamard codes. That yields a derandomization of my original tree size lower bound, which answers the first objection.

Second, I'll show you how Ran Raz's lower bound method is related to a concept in physics called *persistence of entanglement*. Basically, a quantum state is persistently entangled if it still contains a lot of entanglement after a limited amount of interaction with its environment. On the other hand, the connection between large tree size and persistence of entanglement is not a perfect one.

Third, if there's time, I'll say something about a weaker notion than tree size, called "manifestly orthogonal tree size." For this notion, I can prove exponential lower bounds, rather than just $n^{\Omega(\log n)}$. Indeed, I can *exactly* characterize the manifestly orthogonal tree size of linear codeword states.

1 Derandomized Lower Bound

After his main paper on multilinear formula size, Raz wrote a second paper, in which he gave functions that have superpolynomial multilinear formula size but polynomial multilinear *circuit* size. More relevant for us is that he gave a cleaner statement of his original theorem.

Theorem 1 (Raz) *Let $f : \{0,1\}^n \rightarrow \mathbb{C}$. Let $l = n^{\delta}$ for some constant $\delta \in (0,1]$; then let $\mathcal{R}_l$ be the following distribution over restrictions R of f: choose $2l$ inputs uniformly at random, and name them $y = (y_1, \ldots, y_l)$ and $z = (z_1, \ldots, z_l)$. Set each of the remaining $n - 2l$ inputs to 0 or 1 uniformly and independently. This yields a restricted function $f_R : \{0,1\}^l \times \{0,1\}^l \rightarrow \mathbb{C}$. Let $M_{f|R}$ be a $2^l \times 2^l$ matrix whose (y,z) entry is $f_R(y,z)$. Suppose that*

$$\Pr_{R \in \mathcal{R}_l}\left[\operatorname{rank}\left(M_{f|R}\right) \geq 2^{l - l^{1/8}/2}\right] = \Omega(1)$$

Then f has multilinear formula size $n^{\Omega(\log n)}$.

We can use this theorem to prove a tree size lower bound for an explicit linear code. Let

$$V = \begin{pmatrix} 1^0 & 1^1 & \cdots & 1^{k-1} \\ 2^0 & 2^1 & \cdots & 2^{k-1} \\ \vdots & \vdots & & \vdots \\ n^0 & n^1 & \cdots & n^{k-1} \end{pmatrix}$$

be the $n \times k$ Vandermonde matrix, where $1, \ldots, n$ are labels of elements in $\mathbb{F}_{2^d}$ for $d \geq \log_2 n$. Any $k \times k$ submatrix of V has full rank, because the Reed-Solomon (RS) code that V represents is a perfect erasure code (in other words, because a degree-$(k-1)$ polynomial is determined by its values at any k points).

But we want a *binary* linear erasure code with parameters almost as good as this code's. To get one, we concatenate the RS and Hadamard codes. More explicitly, interpret $\mathbb{F}_{2^d}$ as the field of polynomials over $\mathbb{F}_2$, modulo some irreducible of degree d. Then let $m(a)$ be the $d \times d$ Boolean matrix that maps $q \in \mathbb{F}_{2^d}$ to $aq \in \mathbb{F}_{2^d}$, where q and aq are encoded by their $d \times 1$ vectors of coefficients. Let H map a length-d vector to its length-2^d Hadamard encoding. Then $Hm(a)$ is a $2^d \times d$ Boolean matrix that maps $q \in \mathbb{F}_{2^d}$ to the Hadamard encoding of aq. We can now define an $n2^d \times kd$ "binary Vandermonde matrix" as follows:

$$V_b = \begin{pmatrix} Hm\left(1^0\right) & Hm\left(1^1\right) & \cdots & Hm\left(1^{k-1}\right) \\ Hm\left(2^0\right) & Hm\left(2^1\right) & \cdots & Hm\left(2^{k-1}\right) \\ \vdots & \vdots & & \vdots \\ Hm\left(n^0\right) & Hm\left(n^1\right) & \cdots & Hm\left(n^{k-1}\right) \end{pmatrix}.$$

Fix $k = n^{\delta}$ for some $\delta < 1/2$ and $d = O(\log n)$.

Lemma 2 *A $(kd + c) \times kd$ submatrix of V_{bin} chosen uniformly at random has rank kd (that is, full rank) with probability at least $2/3$, for c a sufficiently large constant.*

Proof. We claim that $|V_b u| \geq (n-k)\,2^{d-1}$ for all nonzero vectors $u \in \mathbb{F}_2^{kd}$, where $|\ \ |$ represents the number of '1' bits. To see this, observe that for all nonzero u, the "codeword vector" $Vu \in \mathbb{F}_{2^d}^n$ must have at least $n - k$ nonzero entries by the Fundamental Theorem of Algebra, where here u is interpreted as an element of $\mathbb{F}_{2^d}^k$. Furthermore, the Hadamard code maps any nonzero entry in Vu to 2^{d-1} nonzero bits in $V_b u \in \mathbb{F}_2^{n2^d}$.

Now let W be a uniformly random $(kd + c) \times kd$ submatrix of V_b. By the above claim, for any fixed nonzero vector $u \in \mathbb{F}_2^{kd}$,

$$\Pr_W[Wu = 0] \leq \left(1 - \frac{(n-k)\,2^{d-1}}{n2^d}\right)^{kd+c} = \left(\frac{1}{2} + \frac{k}{2n}\right)^{kd+c}.$$

So by the union bound, Wu is nonzero for all nonzero u (and hence W is full rank) with probability at least

$$1 - 2^{kd}\left(\frac{1}{2} + \frac{k}{2n}\right)^{kd+c} = 1 - \left(1 + \frac{k}{n}\right)^{kd}\left(\frac{1}{2} + \frac{k}{2n}\right)^{c}.$$

Since $k = n^{1/2 - \Omega(1)}$ and $d = O(\log n)$, the above quantity is at least $2/3$ for sufficiently large c. ■

Given an $n2^d \times 1$ Boolean vector x, let $f(x) = 1$ if $V_b^T x = 0$ and $f(x) = 0$ otherwise. Then:

Theorem 3 *f has multilinear formula size $n^{\Omega(\log n)}$.*

Proof. Let V_y and V_z be two disjoint $kd \times (kd+c)$ submatrices of V_b^T chosen uniformly at random. Then by the above lemma together with the union bound, V_y and V_z both have full rank with probability at least 1/3. Letting $l = kd + c$, it follows that

$$\Pr_{R \in \mathcal{R}_l}\left[\operatorname{rank}\left(M_{f|R}\right) \geq 2^{l-c}\right] \geq \frac{1}{3}.$$

The theorem now follows from Raz's lower bound theorem. ■

Above, all I used about the binary Vandermonde matrix is that it corresponds in some sense to a good erasure code. Thus, one can certainly generalize to an equal superposition over codewords of any such code. But I still need to work out the details of how to do that, and what parameters one gets. (There's a mistake in how I do it in the newest version of my paper.)

2 Persistence of Entanglement

We can gain further insight by asking what *physical* properties a codeword state has to have. One important property is "persistence of entanglement," studied by Briegel, Raussendorf, etc. This is the property of remaining highly entangled even after a limited amount of interaction with the environment. For example, the Schrödinger cat state $\left(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}\right)/\sqrt{2}$ is in some sense highly entangled, but it is *not* persistently entangled, since measuring a single qubit in the standard basis destroys all entanglement.

Persistence of entanglement seems related to how one shows tree size lower bounds using Raz's technique. For to apply that technique, one basically "measures" most of a state's qubits, then partitions the unmeasured qubits into two subsystems of equal size, and argues that with high probability those two subsystems are still almost maximally entangled.

The connection is not perfect, though. For one thing, setting most of the qubits to 0 or 1 uniformly at random is not the same as measuring them. Also, that the system is almost maximally entangled *under almost any partition into two subsystems*, seems at least as important as the system remaining entangled after a subset of qubits is measured. Indeed, I can show that there exist states—the *1D cluster states*—that are persistently entangled in some sense yet that have polynomial tree size. On the other hand, I conjecture that 2D cluster states have tree size $n^{\Omega(\log n)}$, but have been unable to prove this.

3 Manifestly Orthogonal Tree Size

Given a state $|\psi\rangle$, the *manifestly orthogonal tree size* of $|\psi\rangle$ is the minimum size of a tree representing $|\psi\rangle$, in which all linear combinations are of two states $|\psi_1\rangle, |\psi_2\rangle$ with "disjoint supports" in the computational basis—that is, either $\langle\psi_1|x\rangle = 0$ or $\langle\psi_2|x\rangle = 0$ for every basis state $|x\rangle$. We can assume without loss of generality that every $+$ or $\otimes$ vertex has at least one child, and that every child of a $+$ vertex is a $\otimes$ vertex and vice versa. Also, given a set $S \subseteq \{0,1\}^n$, let

$$|S\rangle = \frac{1}{\sqrt{|S|}} \sum_{x \in S} |x\rangle$$

be a uniform superposition over the elements of S, and let $M(S)$ be a shorthand for the manifestly orthogonal tree size of $|S\rangle$. I admit this isn't a very physical notion, but it's interesting from a complexity theory perspective.

If $C = \{x : Ax \equiv b \pmod{2}\}$ is the set of codewords of a linear code, then I can give an *exact* characterization of $M(C)$. Intuitively, the characterization says that there's nothing better than the trivial approach of decomposing C into a union of tensor products of codeword states in two subsystems I and J:

$$C = \bigcup_h C_I^{(h)} \otimes C_J^{(h)}.$$

We can then recursively decompose each $C_I^{(h)}$ and $C_J^{(h)}$ in similar manner. Minimizing the resulting tree size over all such decompositions yields the optimal result. The formal statement is as follows:

Theorem 4 *Let $M(A) = M(C)$ (the vector b is irrelevant, so long as $Ax \equiv b$ is solvable). Then*

$$M(A) = \min\left(2^{\operatorname{rank}(A_I)+\operatorname{rank}(A_J)-\operatorname{rank}(A)}\left(M(A_I)+M(A_J)\right)\right)$$

where the minimum is over all nontrivial partitions (A_I, A_J) of the columns of A. As a base case, if A has only one column, then $M(A) = 2$ if $A = 0$ and $M(A) = 1$ otherwise.

This exact characterization has several corollaries. First, it implies that there exists a deterministic $O(n3^n)$-time dynamic programming algorithm that computes $M(A)$, given A as input. Second, it implies that the language $\{A : M(A) \leq s\}$ is in NP. I do not know if this language is NP-complete but conjecture that it is. Third, it implies the following *exponential* lower bound.

Corollary 5 *Suppose the entries of $A \in \mathbb{Z}_2^{k\times n}$ are drawn uniformly and independently at random, where $k \in \left[4\log_2 n, \frac{1}{2}\sqrt{n \ln 2}\right]$. Then $M(A) = \left(n/k^2\right)^{\Omega(k)}$ with probability $\Omega(1)$ over A.*

And fourth, it implies the existence of quantum states with polynomial orthogonal tree size, but *manifestly* orthogonal tree size $n^{\Omega(\log n)}$. For orthogonal tree size, we require that all linear combinations be of two orthogonal states, but do not require the states to have disjoint supports in the computational basis.

Note that in the decomposition

$$C = \bigcup_h C_I^{(h)} \otimes C_J^{(h)},$$

the $C_I^{(h)}$'s and $C_J^{(h)}$'s are unique up to ordering. Furthermore, the quantities $\left|C_I^{(h)}\right|$, $\left|C_J^{(h)}\right|$, $M\left(C_I^{(h)}\right)$, and $M\left(C_J^{(h)}\right)$ remain unchanged as we range over $h \in [H]$. For this reason we suppress the dependence on h when mentioning them.

Let's now prove the characterization theorem. For various sets S, the strategy will be to analyze $M(S)/|S|$, the ratio of tree size to cardinality. We can think of this ratio as the "price per pound" of S: the number of vertices that we have to pay per basis state that we cover with nonzero amplitude. The following lemma says that, under that cost measure, a coset is "as good a deal" as any of its subsets.

Lemma 6 *For all cosets C,*

$$\frac{M(C)}{|C|} = \min\left(\frac{M(S)}{|S|}\right)$$

where the minimum is over nonempty $S \subseteq C$.

Proof. By induction on n. The base case $n = 1$ is obvious, so assume the lemma true for $n-1$. Choose $S^* \subseteq C$ to minimize $M(S^*)/|S^*|$. Let T be a manifestly orthogonal tree for $|S^*\rangle$ of minimum size, and let v be the root of T. We can assume without loss of generality that v is a $\otimes$ vertex, since otherwise v has some $\otimes$ child representing a set $R \subset S^*$ such that $M(R)/|R| \leq M(S^*)/|S^*|$. Therefore for some nontrivial partition (I, J) of $[n]$, and some $S_I^* \subseteq \{0,1\}^{|I|}$ and $S_J^* \subseteq \{0,1\}^{|J|}$, we have

$$\begin{aligned} |S^*\rangle &= |S_I^*\rangle \otimes |S_J^*\rangle, \\ |S^*| &= |S_I^*|\,|S_J^*|, \\ M(S^*) &= M(S_I^*) + M(S_J^*), \end{aligned}$$

where the last equality holds because if $M(S^*) < M(S_I^*) + M(S_J^*)$, then T was not a minimal tree for $|S^*\rangle$. Then

$$\frac{M(S^*)}{|S^*|} = \frac{M(S_I^*) + M(S_J^*)}{|S_I^*|\,|S_J^*|} = \min\left(\frac{M(S_I) + M(S_J)}{|S_I|\,|S_J|}\right)$$

where the minimum is over nonempty $S_I \subseteq \{0,1\}^{|I|}$ and $S_J \subseteq \{0,1\}^{|J|}$ such that $S_I \otimes S_J \subseteq C$. Now there must be an h such that $S_I^* \subseteq C_I^{(h)}$ and $S_J^* \subseteq C_J^{(h)}$, since otherwise some $x \notin C$ would be assigned nonzero amplitude. By the induction hypothesis,

$$\frac{M(C_I)}{|C_I|} = \min\left(\frac{M(S_I)}{|S_I|}\right), \qquad \frac{M(C_J)}{|C_J|} = \min\left(\frac{M(S_J)}{|S_J|}\right),$$

where the minima are over nonempty $S_I \subseteq C_I^{(h)}$ and $S_J \subseteq C_J^{(h)}$ respectively. Define $\beta = |S_I| \cdot |S_J| / M(S_J)$ and $\gamma = |S_J| \cdot |S_I| / M(S_I)$. Then since setting $S_I := C_I^{(h)}$ and $S_J := C_J^{(h)}$ maximizes the four quantities $|S_I|$, $|S_J|$, $|S_I| / M(S_I)$, and $|S_J| / M(S_J)$ simultaneously, this choice also maximizes β and γ simultaneously. Therefore it maximizes their harmonic mean,

$$\frac{\beta\gamma}{\beta+\gamma} = \frac{|S_I|\,|S_J|}{M(S_I)+M(S_J)} = \frac{|S|}{M(S)}.$$

We have proved that setting $S := C_I^{(h)} \otimes C_J^{(h)}$ maximizes $|S| / M(S)$, or equivalently minimizes $M(S) / |S|$. The one remaining observation is that taking the disjoint sum of $C_I^{(h)} \otimes C_J^{(h)}$ over all h leaves the ratio $M(S) / |S|$ unchanged. So setting $S := C$ also minimizes $M(S) / |S|$, and we are done. ■

We are now ready to give a recursive characterization of $M(C)$.

Theorem 7 *If $n \geq 2$, then*

$$M(C) = |C| \min\left(\frac{M(C_I)+M(C_J)}{|C_I|\,|C_J|}\right)$$

where the minimum is over nontrivial partitions (I, J) of $[n]$.

Proof. The upper bound is obvious; we prove the lower bound. Let T be a manifestly orthogonal tree for $|C\rangle$ of minimum size, and let $v^{(1)}, \ldots, v^{(L)}$ be the topmost $\otimes$ vertices in T. Then there exists a partition $\left(S^{(1)}, \ldots, S^{(L)}\right)$ of C such that the subtree rooted at $v^{(i)}$ represents $\left|S^{(i)}\right\rangle$. We have

$$|T| = M\left(S^{(1)}\right) + \cdots + M\left(S^{(L)}\right) = \left|S^{(1)}\right| \frac{M\left(S^{(1)}\right)}{\left|S^{(1)}\right|} + \cdots + \left|S^{(L)}\right| \frac{M\left(S^{(L)}\right)}{\left|S^{(L)}\right|}.$$

Now let $\eta = \min_i \left(M\left(S^{(i)}\right) / \left|S^{(i)}\right|\right)$. We will construct a partition $\left(R^{(1)}, \ldots, R^{(H)}\right)$ of C such that $M\left(R^{(h)}\right) / \left|R^{(h)}\right| = \eta$ for all h, which will imply a new tree T' with $|T'| \leq |T|$. Choose $j \in [L]$ such that $M\left(S^{(j)}\right) / \left|S^{(j)}\right| = \eta$, and suppose vertex $v^{(j)}$ of T expresses $\left|S^{(j)}\right\rangle$ as $|S_I\rangle \otimes |S_J\rangle$ for some nontrivial partition (I, J). Then

$$\eta = \frac{M\left(S^{(j)}\right)}{\left|S^{(j)}\right|} = \frac{M(S_I)+M(S_J)}{|S_I|\,|S_J|}$$

where $M\left(S^{(j)}\right) = M(S_I) + M(S_J)$ follows from the minimality of T. As in the lemma, there must be an h such that $S_I \subseteq C_I^{(h)}$ and $S_J \subseteq C_J^{(h)}$. But the lemma then implies that $M(C_I) / |C_I| \leq M(S_I) / |S_I|$ and that $M(C_J) / |C_J| \leq M(S_J) / |S_J|$. Combining these bounds with $|C_I| \geq |S_I|$ and $|C_J| \geq |S_J|$, we obtain by a harmonic mean inequality that

$$\frac{M(C_I \otimes C_J)}{|C_I \otimes C_J|} \leq \frac{M(C_I)+M(C_J)}{|C_I|\,|C_J|} \leq \frac{M(S_I^*)+M(S_J^*)}{|S_I^*|\,|S_J^*|} = \eta.$$

So setting $R^{(h)} := C_I^{(h)} \otimes C_J^{(h)}$ for all h yields a new tree T' no larger than T. Hence by the minimality of T,

$$M(C) = |T| = |T'| = H \cdot M(C_I \otimes C_J) = \frac{|C|}{|C_I|\,|C_J|} \cdot \left(M(C_I) + M(C_J)\right).$$

■