Multilinear Formulas and Skepticism of Quantum Computing

Scott Aaronson*
University of California, Berkeley

Abstract

Several researchers, including Leonid Levin, Gerard 't Hooft, and Stephen Wolfram, have argued that quantum mechanics will break down before the factoring of large numbers becomes possible. If this is true, then there should be a natural "Sure/Shor separator"—that is, a set of quantum states that can account for all experiments performed to date, but *not* for Shor's factoring algorithm. We propose as a candidate the set of states expressible by a polynomial number of additions and tensor products. Using a recent lower bound on multilinear formula size due to Raz, we then show that states arising in quantum error-correction require $n^{\Omega(\log n)}$ additions and tensor products even to approximate, which incidentally yields the first superpolynomial gap between general and multilinear formula size of functions. More broadly, we introduce a complexity classification of pure quantum states, and prove many basic facts about this classification. Our goal is to refine vague ideas about a breakdown of quantum mechanics into specific hypotheses that might be experimentally testable in the near future.

1 Introduction

> QC of the sort that factors long numbers seems firmly rooted in science fiction ... The present attitude would be analogous to, say, Maxwell selling the Daemon of his famous thought experiment as a path to cheaper electricity from heat. —Leonid Levin [33]

Quantum computing presents a dilemma: is it reasonable to study a type of computer that has never been built, and might never be built in one's lifetime? Some researchers strongly believe the answer is 'no.' Their objections generally fall into four categories:

(A) There is a fundamental physical reason why large quantum computers can never be built.

(B) Even if (A) fails, large quantum computers will never be built in practice.

(C) Even if (A) and (B) fail, the speedup offered by quantum computers is of limited theoretical interest.

(D) Even if (A), (B), and (C) fail, the speedup is of limited practical value.[1]

*Email: aaronson@cs.berkeley.edu. Supported by an NSF Graduate Fellowship, by NSF ITR Grant CCR-0121555, and by the Defense Advanced Research Projects Agency (DARPA).

[1] Because of the 'even if' clauses, the objections seem to us logically independent, so that there are 16 possible positions regarding them (or 15 if one is against quantum computing). We ignore the possibility that no speedup exists, in other words that $\mathsf{BPP} = \mathsf{BQP}$. By 'large quantum computer' we mean any computer much faster than its best classical simulation, as a result of asymptotic complexity rather than the speed of elementary operations. Such a computer need not be universal; it might be specialized for (say) factoring.

The objections can be classified along two axes:

	Theoretical	**Practical**
Physical	(A)	(B)
Algorithmic	(C)	(D)

This paper focuses on objection (A). Its goal is not to win a debate about this objection, but to lay the groundwork for a rigorous discussion, and thus hopefully lead to new science. Section 2 provides the philosophical motivation for our paper, by examining the arguments of several quantum computing skeptics, including Leonid Levin, Gerard 't Hooft, and Stephen Wolfram. It concludes that a key weakness of their arguments is their failure to answer the following question: *Exactly what property separates the quantum states we are sure we can create, from those that suffice for Shor's factoring algorithm?* We call such a property a *Sure/Shor separator.* Section 3 develops a complexity theory of pure quantum states, that studies possible Sure/Shor separators. In particular, it introduces *tree states*, which informally are those states $|\psi\rangle \in \mathcal{H}_2^{\otimes n}$ expressible by a polynomial-size 'tree' of addition and tensor product gates. For example, $\alpha|0\rangle^{\otimes n} + \beta|1\rangle^{\otimes n}$ and $(\alpha|0\rangle + \beta|1\rangle)^{\otimes n}$ are both tree states for all α, β. Indeed, we advance the thesis that *all* states created to date are best seen as tree states.

Section 4 proves basic results about tree states; in particular:

(1) The minimum size of a tree representing the state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$ is equal, up to a constant factor, to the minimum size a multilinear formula representing the function $f(x) = \alpha_x$.

(2) Any tree state is representable by a tree of polynomial size and logarithmic depth.

(3) Any *orthogonal* tree state (one in which all additions are of orthogonal states) can be prepared by a polynomial-size quantum circuit.

(4) If the state $|\psi\rangle$ is chosen uniformly at random, then with high probability $|\psi\rangle$ is not even approximated well by any tree of size $2^{o(n)}$.

(5) There exist states that can be approximated by tree states, but not arbitrarily well approximated.

Our main results, proved in Section 5, are lower bounds on tree size for several families of quantum states. Specifically, we show in Section 5.1 that if C is a coset in $\mathbb{Z}_2^n$, then a uniform superposition over the elements of C cannot represented by a tree of size $n^{o(\log n)}$, with high probability if C is chosen at random. Indeed, with high probability such states are not even *approximated* by trees of size $n^{o(\log n)}$. These 'coset states' are exactly what arise in stabilizer codes, a type of quantum error-correcting code. So the practical upshot of our result is this. If error-correcting code states of the type we describe were implemented with a large number of qubits, it would yield evidence against the hypothesis that all states in Nature are tree states. From our perspective, then, the effort to create these states will do more than test the feasibility of quantum error-correction—it will provide an important new test of quantum mechanics itself.

Section 5.1 also gives two corollaries of our error-correction lower bound, which we believe are of independent complexity-theoretic interest. We obtain the first superpolynomial gap between formula size and multilinear formula size of functions $f : \{0,1\}^n \to \mathbb{R}$, as well as a separation between tree states and manifestly orthogonal tree states. Originally, we had hoped to show a tree-size lower bound for states that arise in Shor's algorithm—for example, a uniform superposition over all multiples of a fixed positive integer p, written in binary. However, we were only able to show such a bound assuming a number-theoretic conjecture. Section 5.2 states the conjecture and shows that it implies an $n^{\Omega(\log n)}$ tree size lower bound for states arising in Shor's algorithm.

Our lower bounds use a sophisticated recent technique of Raz [39], which was introduced to show that the permanent and determinant of a matrix require superpolynomial-size multilinear formulas. Currently, Raz's technique is only able to show lower bounds of the form $n^{\Omega(\log n)}$, but we conjecture that $2^{\Omega(n)}$ lower bounds hold in all of the cases discussed above.

Section 6 addresses the following question. If the state of a quantum computer at every time step is a tree state, then can the computer be simulated classically (that is, in BPP)? Although we leave this question open, we do show that $\mathsf{TreeBQP} \subseteq \Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$, where $\mathsf{TreeBQP}$ is the subclass of BQP corresponding to tree states, and $\Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$ is the third level of the polynomial hierarchy PH.[2] By contrast, it is conjectured (see [2]) that $\mathsf{BQP} \not\subset \mathsf{PH}$. Section 7 proposes an experiment to test the idea that all states in Nature are tree states. We give evidence that our experiment is feasible with current liquid NMR technology, and provide detailed numerical tree size calculations to aid in interpreting the experiment. We conclude in Section 8 with some open problems.

2 How Quantum Mechanics Could Fail

This section discusses objection (A), that quantum computing is impossible for a fundamental physical reason. Although this objection has been raised by several physicists, including Gerard 't Hooft [26] and Stephen Wolfram [46], we will begin with the arguments of Leonid Levin [33, 34], since those are the best known to computer scientists. The following is a sample of points made by Levin that we were able to understand. We should mention that Levin does not consider our sample to be an accurate summary of his views; thus, readers are encouraged to consult [33, 34] where Levin makes further points, for example about a distinction between topological and metric approximation.

- Levin draws analogy between quantum computing and the unit-cost arithmetic model, suggesting that if we reject the latter as extravagant, then we should also reject the former. "[Shamir] proved ... that factoring (on infeasibility of which RSA depends) can be done in polynomial number of arithmetic operations. This result uses a so-called 'unit-cost model,' which charges one unit for each arithmetic operation, however long the operands ... The closed-minded cryptographers, however, were not convinced and this result brought a dismissal of the unit-cost model, not RSA" [33]. Levin then says about quantum computing: "Another, not dissimilar, attack is raging this very moment."

- In a newsgroup discussion [34] involving Levin, Daniel Gottesman, and others, Gottesman began a defense of quantum error-correction as follows: "We know linearity and all other laws of quantum mechanics are at least approximately true. Let us fix, for the sake of convenience, some degree of accuracy to which this approximation is correct—say, 20 digits." Levin interjected: "To this accuracy all these amplitudes are 0." Later Levin again said: "Rounded to 10^{-4} (if not to 10^{-10^4} :-), all amplitudes in your algorithm would be 0." To us, the most natural interpretation of these remarks is that Levin wishes to subject amplitudes to *additive* rather than multiplicative error. That is, he imagines an error process that corrupts the amplitude α_x of each basis state $|x\rangle$ to $\alpha_x \pm \varepsilon$, rather than to $\alpha_x (1 \pm \varepsilon)$ as is assumed in results on quantum fault-tolerance due to Aharonov and Ben-Or [4] among others.[3] In the additive case, clearly only classical computation is possible, since an adversary

[2] See http://www.cs.berkeley.edu/~aaronson/zoo.html for information about the complexity classes mentioned in this paper.

[3] In personal correspondence, Levin denied this interpretation, claiming that it makes no sense to discuss *any* equations governing a quantum computer—whether subject to additive, multiplicative, or any other kind of error.

could corrupt all but $O(1/\varepsilon)$ amplitudes to 0.

- Related to the previous point, Levin sees no reason even to hypothesize that quantum mechanics remains valid to the accuracy needed for quantum computing. "We have never seen a physical law valid to over a dozen decimals. Typically, every few new decimal places require major rethinking of most basic concepts. Are quantum amplitudes still complex numbers to such accuracies or do they become quaternions, colored graphs, or sick-humored gremlins?" [33]

- Levin rejects the idea that quantum computing research "wins either way"—either by building quantum computers, or by discovering that our current understanding of quantum mechanics is incomplete. In his words [33]: "[Consider] this scenario. With few q-bits, QC is eventually made to work. The progress stops, though, long before QC factoring starts competing with pencils. The QC people then demand some noble [sic] prize for the correction to the Quantum Mechanics. But the committee wants more specifics than simply a nonworking machine, so something like observing the state of the QC is needed. Then they find the Universe too small for observing individual states of the needed dimensions and accuracy. (Raising sufficient funds to compete with pencil factoring may justify a Nobel Prize in Economics.)"

  Levin points out that, by a simple counting argument, a 'generic' state $|\psi\rangle \in \mathcal{H}_2^{\otimes n}$ is indistinguishable from the set of states $|\varphi\rangle$ such that $|\langle\psi|\varphi\rangle| \leq \varepsilon$ by quantum circuits of subexponential size. "So, what thought experiments can probe the QC to be in the state described with the accuracy needed? I would allow to use the resources of the entire Universe, but *not more*!"

A few responses to Levin's arguments can be offered immediately. First, even classically, one can flip a coin a thousand times to produce probabilities of order 2^{-1000}. Should we dismiss such probabilities as unphysical, or subject them to additive rather than multiplicative noise? At the very least, it is not obvious that amplitudes should behave differently than probabilities with respect to error—since both evolve linearly, and neither is directly observable.

Second, if Levin believes that quantum mechanics will fail, but is agnostic about what will replace it, then his argument can be turned around. How do we know that the successor to quantum mechanics will limit us to BPP, rather than letting us solve (say) PSPACE-complete problems? This is more than a logical point. Abrams and Lloyd [3] show that a wide class of nonlinear variants of the Schrödinger equation would allow NP-complete and even $\#\mathsf{P}$-complete problems to be solved in polynomial time. And Penrose [38], who proposed a model for 'objective collapse' of the wavefunction, believes that his proposal takes us outside the Kleene hierarchy!

Third, to falsify quantum mechanics, it would suffice to show by repeatable experiments that a quantum computer evolved to *some* state far from the state that quantum mechanics predicts. Measuring the exact state is unnecessary. Nobel prizes have been awarded in the past 'merely' for falsifying a previously held theory, rather than replacing it by a new one. For example, Fitch [17] and Cronin [16] won the physics Nobel in 1980 for discovering that CP symmetry is violated.

Perhaps the key to understanding Levin's unease about quantum computing lies in his remark that "we have never seen a physical law valid to over a dozen decimals." Here he touches on a serious epistemological question: *How far should we extrapolate from today's experiments to where quantum mechanics has never been tested?* We will try to address this question by reviewing the evidence for quantum mechanics. For our purposes it will not suffice to declare that "the predictions of quantum mechanics have been verified to one part in a trillion," because we need to distinguish at least three different *types* of prediction: *interference*, *entanglement*, and *Schrödinger cats.* Let us consider these in turn.

(1) **Interference.** If the different paths that an electron could take in its orbit around a nucleus did not interfere destructively, canceling each other out, then electrons would not have quantized energy levels. So being accelerating electric charges, they would lose energy and spiral into their respective nuclei, and all matter in the universe would disintegrate. That this is not observed to happen—together with the results of (for example) single-photon double-slit experiments—is compelling evidence for the reality of quantum interference.

(2) **Entanglement.** One might accept that (say) a single particle's position is described by a wave in three-dimensional phase space, but deny that two particles are described by a wave in *six*-dimensional phase space. However, the Bell inequality experiments of Aspect et al. [8] and successors have convinced all but a few physicists that quantum entanglement exists, can be maintained over large distances, and cannot be explained by local hidden-variable theories.

(3) **Schrödinger Cats.** Accepting two- and three-particle entanglement is not the same as accepting that whole molecules, cats, humans, and galaxies can be in coherent superposition states. However, recently Arndt et al. [7] have performed the double-slit interference experiment using C_{60} molecules (buckyballs) instead of photons; while Friedman et al. [18] have found evidence that a superconducting current, consisting of billions of electrons, can enter a coherent superposition of flowing clockwise around a coil and flowing counterclockwise. Though short of cats, these experiments at least allow us to say the following: *if we could build a general-purpose quantum computer with as many components as have already been placed into coherent superposition, then on certain problems, that computer would outperform any supercomputer in the world today.*

Having reviewed some of the evidence for quantum mechanics, we must now ask what alternatives have been proposed that might also explain the evidence. The simplest alternatives are those in which quantum states "spontaneously collapse" with some probability, as in the GRW (Ghirardi-Rimini-Weber) theory [21]. (Penrose [38] has proposed another such theory, but as mentioned earlier, his suggests that the quantum computing model is *too* restrictive.) The drawbacks of the GRW theory include violations of energy conservation, and parameters that must be fine-tuned to avoid conflicting with experiments. More relevant for us, though, is that even if the GRW theory were true, fairly large quantum computers could still be built.

A second class of alternatives includes those of 't Hooft [26] and Wolfram [46], in which something like a deterministic cellular automaton underlies quantum mechanics. On the basis of his theory, 't Hooft predicts that "[i]t will never be possible to construct a 'quantum computer' that can factor a large number faster, and within a smaller region of space, than a classical machine would do, if the latter could be built out of parts at least as large and as slow as the Planckian dimensions" [26]. Similarly, Wolfram states that

> [i]ndeed within the usual formalism [of quantum mechanics] one can construct quantum computers that may be able to solve at least a few specific problems exponentially faster than ordinary Turing machines. But particularly after my discoveries ... I strongly suspect that even if this is formally the case, it will still not turn out to be a true representation of ultimate physical reality, but will instead just be found to reflect various idealizations made in the models used so far [46, p.771].

The obvious question then is how these theories account for Bell inequality violations. We confess to being unable to understand 't Hooft's answer to this question, except that he believes that the usual notions of causality and locality might no longer apply in quantum gravity. As for Wolfram's theory, which involves "long-range threads" to account for Bell inequality violations, we argued in [1] that it fails Wolfram's own desiderata of causal and relativistic invariance.

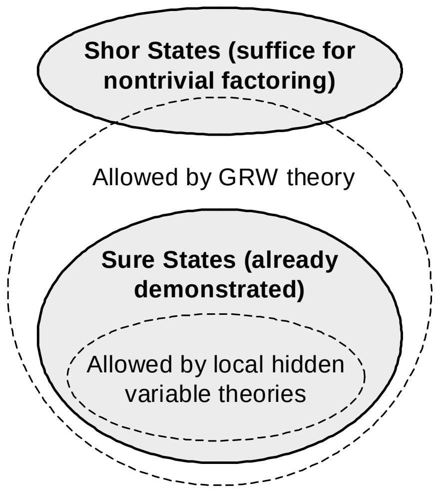

Figure 1: A Sure/Shor separator must contain all Sure states but no Shor states. That is why neither local hidden variables nor the GRW theory yields a Sure/Shor separator.

So the challenge for quantum computing skeptics is clear. Ideally, come up with an alternative to quantum mechanics—even an idealized toy theory—that can account for all present-day experiments (including those of Arndt et al. [7] and Aspect et al. [8]), yet would not allow large-scale quantum computation. Failing that, *at least say what you take quantum mechanics' domain of validity to be.* More concretely, propose a natural set S of quantum states that you believe corresponds to possible physical states of affairs.[4] The set S must contain all "Sure states" (informally, the states that have already been demonstrated in the lab), but no "Shor states" (again informally, the states that can be shown to suffice for factoring, say, 500-digit numbers). If S satisfies both of these constraints, then we call S a *Sure/Shor separator* (see Figure 1).

Of course, an alternative theory need not involve a sharp cutoff between possible and impossible states. So it is perfectly acceptable for a skeptic to define a "complexity measure" $C(|\psi\rangle)$ for quantum states, and then say something like the following:

> If $|\psi_n\rangle$ is a state of n spins, and $C(|\psi_n\rangle)$ is at most, say, n^2, then I predict that $|\psi_n\rangle$ can be prepared using only "polynomial effort." Also, once prepared, $|\psi_n\rangle$ will be governed by standard quantum mechanics to extremely high precision. All states created to date have had small values of $C(|\psi_n\rangle)$. However, if $C(|\psi_n\rangle)$ grows as, say, 2^n, then I predict that $|\psi_n\rangle$ requires "exponential effort" to prepare, or else is not even approximately governed by quantum mechanics. The states that arise in Shor's factoring algorithm have exponential values of $C(|\psi_n\rangle)$. So as my Sure/Shor separator, I propose the set of all infinite families of states $\{|\psi_n\rangle\}_{n\geq 1}$, where $|\psi_n\rangle \in \mathcal{H}_2^{\otimes n}$ has n qubits, such that $C(|\psi_n\rangle) \leq p(n)$ for some polynomial upper bound p.

To understand the importance of Sure/Shor separators, it is helpful to think through some examples. A major theme of Levin's arguments was that exponentially small amplitudes are somehow unphysical.

[4] A skeptic might also specify what happens if a state $|\psi\rangle \in S$ is acted on by a unitary U such that $U|\psi\rangle \notin S$, but this will not be insisted upon.

However, clearly we cannot reject *all* states with tiny amplitudes—for would anyone dispute that the state $2^{-5000}(|0\rangle+|1\rangle)^{\otimes 10000}$ is formed whenever 10,000 photons are each polarized at 45°? Indeed, once we accept $|\psi\rangle$ and $|\varphi\rangle$ as Sure states, we are almost *forced* to accept $|\psi\rangle \otimes |\varphi\rangle$ as well—since we can imagine, if we like, that $|\psi\rangle$ and $|\varphi\rangle$ are prepared in two separate laboratories. So considering a state that arises in Shor's algorithm, such as

$$|\Phi\rangle = \frac{1}{2^{n/2}} \sum_{r=0}^{2^n-1} |r\rangle \left|x^r \bmod N\right\rangle,$$

what property of this state could quantum computing skeptics latch onto as being physically extravagant? They might complain that $|\Phi\rangle$ involves entanglement across hundreds or thousands of particles; but as mentioned earlier, there are other states with that same property, namely the "Schrödinger cats" $\left(|0\rangle^{\otimes n}+|1\rangle^{\otimes n}\right)/\sqrt{2}$, that should be regarded as Sure states. Alternatively, the skeptics might object to the *combination* of exponentially small amplitudes with entanglement across hundreds of particles. However, simply viewing a Schrödinger cat in a different basis produces an equally valid state,

$$\frac{1}{\sqrt{2}}\left[\left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)^{\otimes n}+\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right)^{\otimes n}\right] = \frac{1}{2^{(n-1)/2}} \sum_{x\in\{0,1\}^n\,:\,|x|\equiv 0 (\mathrm{mod}\ 2)} |x\rangle,$$

which has both properties. We seem to be on a slippery slope leading to all of quantum mechanics! Is there any defensible place to draw a line?

The dilemma above is what leads us to propose *tree states* as a candidate Sure/Shor separator. The idea, which might seem more natural to logicians than to physicists, is the following. Once we accept the linear combination and tensor product rules of quantum mechanics—allowing $\alpha|\psi\rangle+\beta|\varphi\rangle$ and $|\psi\rangle\otimes|\varphi\rangle$ into our set S of possible states whenever $|\psi\rangle, |\varphi\rangle \in S$—one of our few remaining hopes for keeping S a proper subset of the set of *all* states is to impose some restriction on how those two rules can be iteratively applied. In particular, we could let S be the closure of $\{|0\rangle, |1\rangle\}$ under a *polynomial number* of linear combinations and tensor products. That is, S is the set of all infinite families of states $\{|\psi_n\rangle\}_{n\geq 1}$ with $|\psi_n\rangle \in \mathcal{H}_2^{\otimes n}$, such that $|\psi_n\rangle$ can be expressed as a "tree" involving at most $p(n)$ addition, tensor product, $|0\rangle$, and $|1\rangle$ gates for some polynomial p (see Figure 2).

One can check that S so defined is rich enough to include Schrödinger cat states, collections of Bell pairs, and many other simple examples of Sure states. Indeed, we advance the thesis that *all* quantum states created to date are best seen as tree states. This thesis has three important caveats. First, for simplicity we are dealing only with pure states, not mixed states. Second, strictly speaking tree states are infinite families of states, one for each number of qubits—so when we say that a state $\{|\psi_n\rangle\}_{n\geq 1}$ was "created," what we really mean is that (say) $|\psi_{50}\rangle$ or $|\psi_{100}\rangle$ was created, and that we have no reason to suppose that creating $|\psi_{1000}\rangle$ or $|\psi_{1000000}\rangle$ would be fundamentally different. Third, states in Nature might be difficult to express in terms of qubits. In particular, it is not obvious how to extend our framework to particle positions and momenta, for which the issues of boson and fermion statistics come into play.

Partly because of these caveats, we would not defend the idea that "all states in Nature are tree states" as a serious physical hypothesis. Our point is simply that to debate objection (A), we need a *foil*—a way the world could be such that (i) large-scale quantum computing is impossible, but (ii) no experiment has yet detected any deviation from quantum mechanics. Several of the obvious ideas for such a foil are nonstarters. Limiting the class of quantum states to those with a certain kind of polynomial-size representation is the simplest example of a foil we could come up with. Our goal in this paper is to investigate where that idea leads.

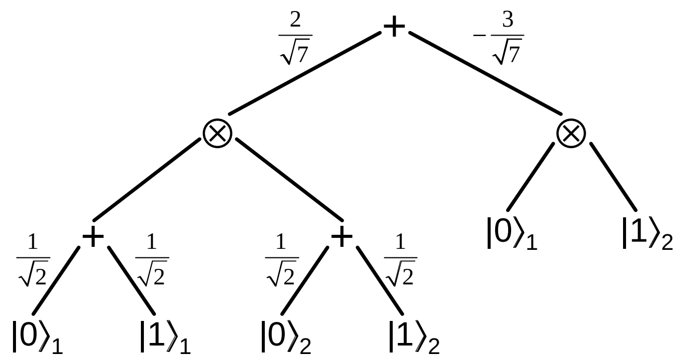

Figure 2: Expressing the state $(|00\rangle - 2|01\rangle + |10\rangle + |11\rangle)/\sqrt{7}$ by a tree of 11 addition, tensor product, $|0\rangle$, and $|1\rangle$ gates, with scalar multiplication along edges. Subscripts denote the identity of a qubit.

3 Classifying Quantum States

In both quantum and classical complexity theory, the objects studied are usually sets of languages or Boolean functions. However, a generic n-qubit quantum state requires exponentially many classical bits to describe, and this suggests looking at *the complexity of quantum states themselves.* That is, which states have polynomial-size classical descriptions of various kinds? This question has been studied from several angles by Aharonov and Ta-Shma [5]; Janzing, Wocjan, and Beth [27]; and Vidal [44]. Here we propose a unified framework for the question. For simplicity, we limit ourselves to pure states $|\psi_n\rangle \in \mathcal{H}_2^{\otimes n}$ with the fixed orthogonal basis $\{|x\rangle : x \in \{0,1\}^n\}$. Also, by 'states' we mean infinite families of states $\{|\psi_n\rangle\}_{n\geq 1}$.

Like complexity classes, pure quantum states can be organized into a hierarchy (see Figure 3). At the bottom are the classical basis states, which have the form $|x\rangle$ for some $x \in \{0,1\}^n$. We can generalize classical states in two directions: to the class $\otimes_1$ of separable states, which have the form $(\alpha_1|0\rangle + \beta_1|1\rangle) \otimes \cdots \otimes (\alpha_n|0\rangle + \beta_n|1\rangle)$; and to the class Σ_1, which consists of all states $|\psi_n\rangle$ that are superpositions of at most $p(n)$ classical states, where p is a polynomial. At the next level, $\otimes_2$ contains the states that can be written as a tensor product of Σ_1 states, with qubits permuted arbitrarily. Likewise, Σ_2 contains the states that can be written as a linear combination of a polynomial number of $\otimes_1$ states. We can continue indefinitely to Σ_3, $\otimes_3$, etc. Containing the whole 'tensor-sum hierarchy' $\cup_{\mathsf{k}}\Sigma_{\mathsf{k}} = \cup_{\mathsf{k}}\otimes_{\mathsf{k}}$ is the class Tree, of all states expressible by a polynomial-size tree of additions and tensor products nested arbitrarily. (Formally, Tree consists of all states $|\psi_n\rangle$ such that $\mathrm{TS}(|\psi_n\rangle) \leq p(n)$ for some polynomial p, where the *tree size* $\mathrm{TS}(|\psi_n\rangle)$ will be defined shortly.) Four other classes deserve mention:

- $\mathsf{Circuit}$, a circuit analog of tree, contains the states $|\psi_n\rangle = \sum_x \alpha_x |x\rangle$ such that for all n, there exists a multilinear algebraic circuit of size $p(n)$ over the complex numbers that outputs α_x given x as input, for some polynomial p.

- AmpP contains the states $|\psi_n\rangle = \sum_x \alpha_x |x\rangle$ such that for all n, b, there exists a classical circuit of size $p(n+b)$ that outputs α_x to b bits of precision given x as input, for some polynomial p.

- Vidal contains the states that are 'polynomially entangled' in the sense of Vidal [44]. Given a par-

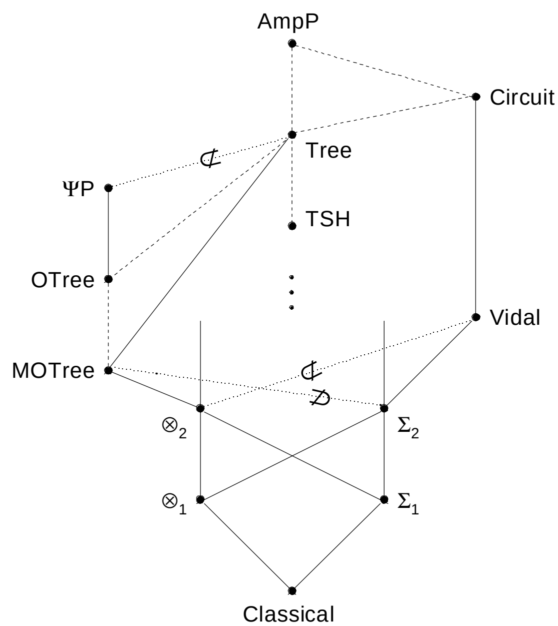

Figure 3: Inclusion diagram of quantum state classes. Solid lines indicate strict containments; dashed lines indicate containments not known to be strict; dashed lines with $\not\subset$ indicate non-containments.

tition of $\{1,\ldots,n\}$ into A and B, let $\chi_A(|\psi_n\rangle)$ be the minimum k for which $|\psi_n\rangle$ can be written as $\sum_{i=1}^{k}\alpha_i|\varphi_i^A\rangle\otimes|\varphi_i^B\rangle$, where $|\varphi_i^A\rangle$ and $|\varphi_i^B\rangle$ are states of qubits in A and B respectively. ($\chi_A(|\psi_n\rangle)$ is known as the *Schmidt rank*.) Let $\chi(|\psi_n\rangle)=\max_A\chi_A(|\psi_n\rangle)$. Then $|\psi_n\rangle\in\mathsf{Vidal}$ if and only if $\chi(|\psi_n\rangle)\leq p(n)$ for some polynomial p.

- $\Psi\mathsf{P}$ contains the states $|\psi_n\rangle$ such that for all n and $\varepsilon>0$, there exists a quantum circuit of size $p(n+\log(1/\varepsilon))$ that maps the initial state $|0\rangle^{\otimes p(n)}$ to a state $|\varphi\rangle$ such that $\left|\langle\varphi|\left(|\psi_n\rangle\otimes|0\rangle^{\otimes(p(n)-n)}\right)\right|\geq 1-\varepsilon$, for some polynomial p. Because of the Solovay-Kitaev Theorem [28, 37], the definition of $\Psi\mathsf{P}$ is invariant under the choice of universal gate set.

Proposition 1 *(i)* $\mathsf{Tree}\cup\mathsf{Vidal}\subseteq\mathsf{Circuit}\subseteq\mathsf{AmpP}$.

(ii) All states in Vidal *have tree size* $n^{O(\log n)}$.

(iii) $\Sigma_2\subseteq\mathsf{Vidal}$ *but* $\otimes_2\not\subset\mathsf{Vidal}$.

(iv) $\otimes_2\subsetneq\mathsf{MOTree}$.

(v) Σ_1, Σ_2, Σ_3, $\otimes_1$, $\otimes_2$, *and* $\otimes_3$ *are all distinct. Also,* $\otimes_3\neq\Sigma_4\cap\otimes_4$.

Like most proofs in this paper, the proof of Proposition 1 can be found in Appendix 9.

We now formalize the notion of *tree size* of a quantum state, which will be used throughout this paper.

Definition 2 *A quantum state tree over* $\mathcal{H}_2^{\otimes n}$ *is a rooted tree where each leaf vertex is labeled with either* $|0\rangle$ *or* $|1\rangle$*, and each non-leaf vertex (called a gate) is labeled with either* $+$ *or* $\otimes$. *Each vertex* v *is also labeled with a set* $S(v)\subseteq\{1,\ldots,n\}$*, such that*

(i) If v *is a leaf then* $|S(v)|=1$,

(ii) If v *is the root then* $S(v)=\{1,\ldots,n\}$,

(iii) If v *is a* $+$ *gate and* w *is a child of* v*, then* $S(w)=S(v)$,

(iv) If v *is a* $\otimes$ *gate and* $w_1,\ldots,w_k$ *are the children of* v*, then* $S(w_1),\ldots,S(w_k)$ *are pairwise disjoint and form a partition of* S_v.

Finally, if v *is a* $+$ *gate, then the outgoing edges* $e_1,\ldots,e_k$ *of* v *are labeled with complex numbers* $\beta_1,\ldots,\beta_k$. *For each vertex* v*, the subtree rooted at* v *represents a quantum state of the qubits in* $S(v)$ *in the obvious way. We require this state to be normalized for each* v.[5]

We say a tree is *orthogonal* if it satisfies the further condition that if v is a $+$ gate, then any two children w_1,w_2 of v represent $|\psi_1\rangle,|\psi_2\rangle$ with $\langle\psi_1|\psi_2\rangle=0$. If the condition $\langle\psi_1|\psi_2\rangle=0$ can be replaced by the stronger condition that for all basis states $|x\rangle$, either $\langle\psi_1|x\rangle=0$ or $\langle\psi_2|x\rangle=0$, then we say the tree is *manifestly orthogonal*.

The *size* $|T|$ of a tree T is the number of vertices, including leaf vertices, and the *depth* of T is the maximum number of edges in any path from the root to a leaf. Then given a state $|\psi\rangle\in\mathcal{H}_2^{\otimes n}$, the *tree size* $\mathrm{TS}(|\psi\rangle)$ is the minimum size of a tree that represents $|\psi\rangle$. The *orthogonal tree size* $\mathrm{OTS}(|\psi\rangle)$ and *manifestly orthogonal tree size* $\mathrm{MOTS}(|\psi\rangle)$ are defined similarly. Then we can let OTree be the class of $|\psi_n\rangle$ such that $\mathrm{OTS}(|\psi_n\rangle)\leq p(n)$ for some polynomial p, and MOTree be the class such that $\mathrm{MOTS}(|\psi_n\rangle)\leq p(n)$ for some p. Also:

[5] Requiring only the *whole* tree to represent a normalized state clearly yields no further generality.

Proposition 3

(i) For every $|\psi\rangle$,

$$n+1 \leq \operatorname{TS}(|\psi\rangle) \leq \operatorname{OTS}(|\psi\rangle) \leq \operatorname{MOTS}(|\psi\rangle) \leq (n+1)\,2^n + 1.$$

(ii) The set of $|\psi\rangle$ *such that* $\operatorname{TS}(|\psi\rangle) < 2^n$ *has measure* 0 *in* $\mathcal{H}_2^{\otimes n}$.

We can also define the *ε-approximate tree size* $\operatorname{TS}_\varepsilon(|\psi\rangle)$ to be the minimum size of a tree representing a state $|\varphi\rangle$ such that $|\langle\psi|\varphi\rangle|^2 \geq 1-\varepsilon$, and define $\operatorname{OTS}_\varepsilon(|\psi\rangle)$ and $\operatorname{MOTS}_\varepsilon(|\psi\rangle)$ similarly.

Definition 4 *An arithmetic formula (over the ring* $\mathbb{C}$ *and* n *variables) is a rooted binary tree where each leaf vertex is labeled with either a complex number or a variable in* $\{x_1,\ldots,x_n\}$, *and each non-leaf vertex is labeled with either* $+$ *or* $\times$. *Such a tree represents a polynomial* $p(x_1,\ldots,x_n)$ *in the obvious way. We call a polynomial multilinear if no variable appears raised to a higher power than* 1, *and an arithmetic formula multilinear if the polynomials computed by each of its subtrees are multilinear.*

The *size* $|\Phi|$ of a multilinear formula Φ is the number of vertices. Given a multilinear polynomial p, the multilinear formula size $\operatorname{MFS}(p)$ is the minimum size of a multilinear formula that represents p. Then given a function $f:\{0,1\}^n \rightarrow \mathbb{C}$, we define

$$\operatorname{MFS}(f) = \min_{p\,:\,p(x)=f(x)\ \forall x\in\{0,1\}^n} \operatorname{MFS}(p).$$

(Actually p turns out to be unique.) We can also define the ε-approximate multilinear formula size of f, $\operatorname{MFS}_\varepsilon(f) = \min_{p\,:\,\|p-f\|_2^2<\varepsilon} \operatorname{MFS}(p)$ where $\|p-f\|_2^2 = \sum_{x\in\{0,1\}^n} |p(x)-f(x)|^2$. Now given a state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$ in $\mathcal{H}_2^{\otimes n}$, let f_ψ be the function from $\{0,1\}^n$ to $\mathbb{C}$ defined by $f_\psi(x) = \alpha_x$.

Theorem 5 *For all* $|\psi\rangle$,

(i) $\operatorname{MFS}(f_\psi) \leq 6\operatorname{TS}(|\psi\rangle)$.

(ii) $\operatorname{TS}(|\psi\rangle) = O(\operatorname{MFS}(f_\psi)+n)$.

(iii) $\operatorname{MFS}_\delta(f_\psi) \leq 6\operatorname{TS}_\varepsilon(|\psi\rangle)$ *where* $\delta = 2-2\sqrt{1-\varepsilon}$.

(iv) $\operatorname{TS}_{2\varepsilon}(|\psi\rangle) = O(\operatorname{MFS}_\varepsilon(f_\psi)+n)$.

4 Basic Results

Before studying the tree size of specific quantum states, we would like to know in general how tree size behaves as a complexity measure. In this section we show that tree size has three rather nice properties. First, any tree of size S can be made to have depth $O(\log S)$ with only a polynomial blowup in size (Theorem 6). Second, any orthogonal tree state can be prepared by a polynomial-size quantum circuit (Theorem 7). And third, most quantum states are *not* approximated by any tree of subexponential size (Theorem 8).

Theorem 6 *For all* $\varepsilon > 0$, *there exists a tree representing* $|\psi\rangle$ *of size* $O\left(\operatorname{TS}(|\psi\rangle)^{1+\varepsilon}\right)$ *depth* $O(\log \operatorname{TS}(|\psi\rangle))$, *and a manifestly orthogonal tree of size* $O\left(\operatorname{MOTS}(|\psi\rangle)^{1+\varepsilon}\right)$ *and depth* $O(\log \operatorname{MOTS}(|\psi\rangle))$.

Theorem 7 *The state $|\psi\rangle$ can be prepared by a quantum circuit of size polynomial in* $\mathrm{OTS}(|\psi\rangle)$*. Therefore* $\mathsf{OTree} \subseteq \Psi\mathsf{P}$.

Theorem 8 *If $|\psi\rangle \in \mathcal{H}_2^{\otimes n}$ is chosen uniformly at random (with respect to the Haar measure), then* $\mathrm{TS}_{1/16}(|\psi\rangle) = 2^{\Omega(n)}$ *with probability $1 - o(1)$.*

A corollary of Theorem 8 is the following 'nonamplification' property: there exist states that can be approximated to within, say, 1% by trees of polynomial size, but that require exponentially large trees to approximate to within a smaller margin (say 0.01%).

Corollary 9 *For all $\delta \in (0,1]$, there exists a state $|\psi\rangle$ such that* $\mathrm{MOTS}_{\delta}(|\psi\rangle) = n+1$ *but* $\mathrm{TS}_{\varepsilon}(|\psi\rangle) = 2^{\Omega(n)}$ *where $\varepsilon = \delta/32 - \delta^2/4096$.*

5 Lower Bounds

We want to show that certain quantum states of interest to us are not represented by trees of polynomial size. At first this seems like a hopeless task. Proving superpolynomial formula-size lower bounds for 'explicit' functions is a notorious open problem, as it would imply complexity class separations such as $\mathsf{NC}^1 \neq \mathsf{P}$. Indeed, Naor and Reingold [36] gave pseudorandom functions as hard as factoring in NC^1, so the arguments of Razborov and Rudich [40] imply that any 'natural' proof of a superpolynomial lower bound on formula size would yield a subexponential-time classical factoring algorithm.

Here, though, we are only concerned with *multilinear* formulas. Could this make it easier to prove a lower bound? The answer is not obvious, but very recently, for reasons unrelated to quantum computing, Raz [39] showed the first superpolynomial lower bounds on multilinear formula size. In particular, he showed that multilinear formulas computing the permanent or determinant of an $n \times n$ matrix over any field have size $n^{\Omega(\log n)}$. We do not know of subexponential-size multilinear formulas for either problem, but Raz's method is currently unable to prove exponential lower bounds; $n^{\Omega(\log n)}$ is the best it can show.

Raz's method is a beautiful combination of the Furst-Saxe-Sipser technique of random restrictions [19], with matrix rank arguments as used in communication complexity. We now outline the method. Given a function $f : \{0,1\}^n \rightarrow \mathbb{C}$, let a *$k$-restriction* R (for $0 \leq k \leq n/2$) set $n - 2k$ of the variables of f to either 0 or 1, and partition the remaining $2k$ variables into two collections $y = (y_1, \ldots, y_k)$ and $z = (z_1, \ldots, z_k)$. This yields a restricted function $f_{|R}(y,z) : \{0,1\}^k \times \{0,1\}^k \rightarrow \mathbb{C}$. Then let $M_{f|R}$ be a $2^k \times 2^k$ matrix whose rows are labeled by assignments $y \in \{0,1\}^k$, and whose columns are labeled by assignments $z \in \{0,1\}^k$. The (y,z) entry of $M_{f|R}$ equals $f_{|R}(y,z)$. Let $\mathrm{rank}(M_{f|R})$ be the rank of $M_{f|R}$ over the complex numbers. The following is a special case[6] of Raz's main theorem [39]; recall that $\mathrm{MFS}(f)$ is the minimum size of a multilinear formula for f.

Theorem 10 (Raz) *Let $\mathcal{D}_k$ be the uniform distribution over k-restrictions of f, meaning that $y_1, \ldots, y_k$ and $z_1, \ldots, z_k$ are chosen uniformly at random, and each of the remaining $n - 2k$ variables is set to 1 with independent probability $1/2$ and to 0 otherwise. Set $k = n^{\delta}$, and suppose that for some constants $\delta \in (0, 1/3]$ and $c > 0$,*

$$\Pr_{R \in \mathcal{D}_k}\left[\mathrm{rank}\left(M_{f|R}\right) \geq c2^k\right] = \Omega(1).$$

Then $\mathrm{MFS}(f) = n^{\Omega(\log n)}$.

[6] Raz uses a distribution over restrictions that is more tailored to the permanent and determinant functions, but examining his proof, it is easy to see that our distribution works equally well.

The intuition behind Theorem 10 is the following. Multilinear formulas can compute functions whose associated matrices have large rank. For example, the inner product function

$$\mathrm{IP}(x, w) = x_1 w_1 + \cdots + x_{n/2} w_{n/2} \pmod{2}$$

has a multilinear formula of linear size, even though M_{IP}, the $2^{n/2} \times 2^{n/2}$ matrix whose (x, w) entry is $\mathrm{IP}(x, w)$, has rank $2^{n/2}$. But now suppose that we apply a random k-restriction to IP to obtain a function $\mathrm{IP}_{|R}$. Then with high probability, the $2^k \times 2^k$ matrix $M_{\mathrm{IP}\,|R}$ will have small (in fact constant) rank—since for almost all products $x_i w_i$, either x_i or w_i has been replaced by a constant. Raz shows that similar behavior applies not only to the inner product, but to *any* function f that has a small multilinear formula Φ. For a random restriction makes many vertices v of Φ *unbalanced*, in the sense that one child of v depends on substantially more variables in $\{y_1, \ldots, y_k, z_1, \ldots, z_k\}$ than the other child. In particular, say a path from the root of Φ to a leaf is *central* if every vertex in the path depends on at least as many variables as its sibling. Then Raz shows that any central path contains at least one unbalanced vertex with probability $1 - n^{-\Omega(\log n)}$ over R. This is because a central path contains at least $\log n$ vertices, each vertex is balanced with probability about $n^{-\varepsilon}$ for some $\varepsilon > 0$, and the probabilities do not depend too strongly on each other. If the number of central paths is $n^{o(\log n)}$, then it follows by the union bound that *every* central path contains at least one unbalanced vertex with probability $1 - o(1)$. In that situation, Raz shows that $\operatorname{rank}(M_{f|R}) = o(2^k)$. So proving

$$\Pr_{R \in \mathcal{D}_k}\left[\operatorname{rank}\left(M_{f|R}\right) \geq c2^k\right] = \Omega(1)$$

implies that the number of central paths (and hence the number of vertices) is $n^{\Omega(\log n)}$.

We will apply Raz's theorem to obtain $n^{\Omega(\log n)}$ tree size lower bounds for two classes of quantum states: states arising in quantum error-correction in Section 5.1, and (assuming a number-theoretic conjecture) states arising in Shor's factoring algorithm in Section 5.2. In the remainder of this section we give two simple extensions of Raz's result, which will be used in Section 5.1. The first extension yields lower bounds on approximate tree size. Given an $N \times N$ matrix $M = (m_{ij})$, let $\operatorname{rank}_\varepsilon(M) = \min_{L\,:\,\|L-M\|_2^2 < \varepsilon} \operatorname{rank}(L)$ where $\|L - M\|_2^2 = \sum_{i,j=1}^{N} |l_{ij} - m_{ij}|^2$.

Corollary 11 *Letting $\mathcal{D}_k$ be as before, suppose that for some κ,*

$$\Pr_{R \in \mathcal{D}_k}\left[\operatorname{rank}_\delta\left(M_{f|R}\right) \geq c2^k\right] = \frac{1}{\kappa} + \Omega(1)$$

where $\|f\|_2^2 = 1$ and $\delta = \kappa\varepsilon 2^{2k}/2^n$. Then $\mathrm{MFS}_\varepsilon(f) = n^{\Omega(\log n)}$.

The second extension of Raz's result yields lower bounds on manifestly orthogonal tree size. Let $\operatorname{morank}(M_{f|R})$ be the minimum number of rank-1 matrices $A_1, \ldots, A_r$ such that $A_1 + \cdots + A_r = M_{f|R}$ and for all y, z, at most one A_i has $A_i(y, z) \neq 0$.

Theorem 12 *Given $|\psi\rangle$, let $\mathcal{D}_k$ be the uniform distribution over k-restrictions of f_ψ, and suppose that for some constant c,*

$$\Pr_{R \in \mathcal{D}_k}\left[\operatorname{morank}\left(M_{f|R}\right) \geq c2^k\right] = \Omega(1).$$

Then $\mathrm{MOTS}(|\psi\rangle) = n^{\Omega(\log n)}$.

Proof. Raz's proof goes through without change, if we assume the formula Φ to be manifestly orthogonal, and replace every occurrence of rank by morank. ■

5.1 Coset States

Let the elements of $\mathbb{Z}_2^n$ be labeled by n-bit strings. Given a coset C in $\mathbb{Z}_2^n$, we define the *coset state* $|C\rangle$ as follows:

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

Coset states arise as codewords in the class of quantum error-correcting codes known as stabilizer codes [15, 23, 42]. Our interest in these states, however, arises from their large tree size rather than their error-correcting properties.

For an integer $k \geq 0$, let $\mathcal{E}_{k,n}$ be the following distribution over cosets C. Choose a $k \times n$ matrix A and $k \times 1$ vector v by setting each entry to 0 or 1 uniformly and independently. Then let $C = \{x \mid Ax \equiv v\}$ (here all congruences are mod 2). By Theorem 5, it suffices to consider the multilinear formula size of the function $f_C(x)$, which is 1 if $x \in C$ and 0 otherwise. Throughout this subsection we set $k = n^{1/3}$.

Theorem 13 *If C is drawn from $\mathcal{E}_{k,n}$, then* $\operatorname{MFS}(f_C) = n^{\Omega(\log n)}$ *(and hence* $\operatorname{TS}(|C\rangle) = n^{\Omega(\log n)}$*), with probability $\Omega(1)$ over C.*

Proof. Let R be a random k-restriction of f_C: that is, it renames $2k$ randomly chosen inputs $y_1, \ldots, y_k$, $z_1, \ldots, z_k$, and restricts the remaining $n - 2k$ inputs to 0 or 1 each with independent probability $1/2$. Then there is a coset $\widehat{C}$ in $\mathbb{Z}_2^k \times \mathbb{Z}_2^k$, such that $f_{C|R}(y,z) = 1$ if $(y,z) \in \widehat{C}$ and $f_{C|R}(y,z) = 0$ otherwise, for all $y, z \in \mathbb{Z}_2^k$. The key point is that we can consider $\widehat{C}$ as a uniform random coset of $\mathbb{Z}_2^{2k}$; that is, as drawn from $\mathcal{E}_{k,2k}$. Let $M_{C|R}$ be the matrix whose (y,z) entry is $f_{C|R}(y,z)$. Then there is a fixed $k \times k$ matrix A, such that $\operatorname{rank}(M_{C|R})$ equals the number of distinct $k \times 1$ vectors v for which

(i) there is at least one y such that $(y,z) \in \widehat{C}$ if and only if $Az \equiv v$, and

(ii) there is at least one z satisfying $Az \equiv v$.

Furthermore, there is a fixed $k \times k$ matrix B and vector w such that (i) holds if and only if v can be written as $Bz + w$ for some z. It is easily checked that A, B, and w are all uniformly random and independent as we range over C and R. Now, the probability that a random $k \times k$ matrix over $\mathbb{Z}_2$ is invertible is

$$\frac{1}{2} \cdot \frac{3}{4} \cdot \cdots \cdot \frac{2^k - 1}{2^k} > 0.288.$$

So the probability that A and B are both invertible is at least 0.288^2. In that case $\operatorname{rank}(M_{C|R}) = 2^k$. By Markov's inequality, it follows that for at least an 0.04 fraction of C's, $\operatorname{rank}(M_{C|R}) = 2^k$ for at least an 0.04 fraction of R's. Theorem 10 then yields the desired result. ■

For a fixed subgroup $S \leq \mathbb{Z}_2^n$, suppose we had a multilinear formula of size t to compute $f_S(x)$, which is 1 if $x \in S$ and 0 otherwise. Then with a multilinear formula of size $O(t)$, we could compute $f_C(x)$ for any coset C of S, by simply negating the appropriate bits of x. So a corollary of Theorem 13 is that for a randomly chosen S, multilinear formulas for $f_S(x)$ are of size $n^{\Omega(\log n)}$ with probability $\Omega(1)$ over S. Since coset states are easily prepared by polynomial-size quantum circuits, a second corollary is that $\Psi\mathsf{P} \not\subset \mathsf{Tree}$. A third corollary is the following.

Corollary 14 *There exists a family of functions $g_n : \{0,1\}^n \to \mathbb{R}$ that has polynomial-size (nonuniform) arithmetic formulas, but no polynomial-size multilinear formulas.*

Proof. Clearly f_C has an arithmetic formula of size $O(nk) = O\left(n^{4/3}\right)$, since there is a $k \times n$ matrix A and vector b such that $x \in C$ if and only if $Ax \equiv b \pmod 2$. ■

The reason Corollary 14 does not follow from Raz's results is that polynomial-size formulas for the permanent and determinant are not known; the smallest known formulas for the determinant have size $n^{O(\log n)}$ (see [14]).

We have shown that not all coset states are tree states, but it is still conceivable that all coset states are extremely well *approximated* by tree states. Let us now rule out the latter possibility. We first need a lemma about matrix rank.

Lemma 15 *Let M be an $N \times N$ complex matrix, and let I_N be the $N \times N$ identity matrix. Then $\|M - I_N\|_2^2 \geq N - \operatorname{rank}(M)$.*

Let $\widehat{f}_C(x) = f_C(x) / \sqrt{|C|}$ be f_C normalized to have $\left\|\widehat{f}_C\right\|_2^2 = 1$.

Theorem 16 *For $\varepsilon < 0.02$, if C is drawn from $\mathcal{E}_{k,n}$, then $\mathrm{MFS}_\varepsilon\left(\widehat{f}_C\right) = n^{\Omega(\log n)}$ with probability $\Omega(1)$ over C.*

Proof. As in Theorem 13, we look at the matrix $M_{C|R}$ induced by a random k-restriction R of $\widehat{f}_C$. We have already seen that for at least an 0.04 fraction of C's, $M_{C|R}$ is a permutation of $I_{2^k}/\sqrt{|C|}$ for at least an 0.04 fraction of R's, where I_{2^k} is the identity. In this case $\operatorname{rank}_\delta\left(M_{C|R}\right) \geq 2^k - \delta|C|$ by Lemma 15. Furthermore, since for these C's there exists an R that makes the matrices A and B from Theorem 13 invertible, it follows that the k equations that define C are linearly independent and solvable. Therefore $|C| = 2^{n-k}$. So taking $\delta = \kappa\varepsilon 2^{2k}/2^n$ with $\kappa = 1/(2\varepsilon)$, we have

$$\Pr_{R \in \mathcal{D}_k}\left[\operatorname{rank}_\delta\left(M_{C|R}\right) \geq 2^{k-1}\right] \geq 0.04 > 2\varepsilon = \frac{1}{\kappa},$$

and Corollary 11 yields the desired result. ■

A corollary of Theorem 16 and of Theorem 5, part (iii), is that $\mathrm{TS}_\varepsilon(|C\rangle) = n^{\Omega(\log n)}$ with probability $\Omega(1)$ over C, for $\varepsilon < 0.0199$.

Finally, we show a separation between tree size and manifestly orthogonal tree size.

Theorem 17 *There exists a family of states $|\psi_n\rangle \in \mathcal{H}_2^{\otimes n}$ such that $\mathrm{TS}(|\psi_n\rangle) = O\left(n^{4/3}\right)$ while $\mathrm{MOTS}(|\psi_n\rangle) = n^{\Omega(\log n)}$. Therefore $\mathsf{MOTree} \neq \mathsf{Tree}$ (in fact we get $\Sigma_2 \not\subset \mathsf{MOTree}$).*

5.2 Shor States

Since the motivation for our theory was to study possible Sure/Shor separators, an obvious question is, *do tree states constitute a Sure/Shor separator?* Unfortunately, we are only able to answer this question assuming a number-theoretic conjecture. To formalize the question, let p be a prime and a an integer with $0 \leq a < p < 2^n$. Then letting $w = \lfloor (2^n - a - 1)/p \rfloor$, define the 'Shor state' $|a + p\mathbb{Z}\rangle = w^{-1/2} \sum_{i=0}^{w} |a + pi\rangle$, where each integer is written as an n-bit string. This is a possible state of the first register in Shor's factoring algorithm, after the second register is measured but before the Fourier transform is applied.[7] So a lower bound on $\mathrm{TS}(|a + p\mathbb{Z}\rangle)$ would imply an equivalent lower bound for the joint state of the two registers. By Theorem 5, it suffices to consider the multilinear formula size of the function $f_{n,p,a}$, that given an n-bit integer x outputs 1 if $x \equiv a \pmod p$ and 0 otherwise. We have the following.

[7] In general there is no reason for the period p to be prime, but this seems like a convenient assumption for lower bound purposes.

Proposition 18

(i) *For all* a, b, $\mathrm{MFS}(f_{n,p,a}) \le \mathrm{MFS}(f_{n+\log p,p,b})$. *Thus, it suffices to consider Shor states of the form* $|p\mathbb{Z}\rangle$, *corresponding to functions* $f_{n,p}(x) = f_{n,p,0}(x)$ *that test whether* x *is divisible by* p.

(iii) $\mathrm{MFS}(f_{n,p}) = O(n2^n/p)$.

(ii) $\mathrm{MFS}(f_{n,p}) = O(np)$.

We now state our number-theoretic conjecture.

Conjecture 19 *Let* S *be a set of nonnegative integers with* $|S| = 32t$ *and* $x \le 2^{(\log t)^c}$ *for all* $x \in S$ *and some constant* $c > 0$. *Let* $S(\mathrm{mod}\, p) = \{x \bmod p : x \in S\}$. *For sufficiently large* t, *if we choose a prime* p *uniformly at random from* $[t, 5t/4]$, *then* $\Pr_p[|S(\mathrm{mod}\, p)| \ge 3t/4] \ge 3/4$.

Proposition 20 *Conjecture 19 implies that, if we choose a prime* p *uniformly at random from the range* $\left[2^{n^{1/c}}, 1.25 \cdot 2^{n^{1/c}}\right]$, *then with* $\Omega(1)$ *probability,* $\mathrm{TS}(|p\mathbb{Z}\rangle) = n^{\Omega(\log n)}$ *and* $\mathrm{TS}_{\varepsilon}(|p\mathbb{Z}\rangle) = n^{\Omega(\log n)}$ *for some fixed* $\varepsilon > 0$.

6 Computing With Tree States

Suppose a quantum computer is restricted to being in a tree state at all times. (We can imagine that if the tree size ever exceeds some polynomial bound, the quantum computer explodes, destroying our laboratory.) Does the computer then have an efficient classical simulation? In other words, letting TreeBQP be the class of languages accepted by such a machine, does TreeBQP = BPP? A positive answer would make tree states more attractive as a Sure/Shor separator. For once we admit any states incompatible with the polynomial-time Church-Turing thesis, it seems like we might as well go all the way, and admit *all* states preparable by polynomial-size quantum circuits! The TreeBQP versus BPP problem is closely related to the problem of finding an efficient (classical) algorithm to *learn* multilinear formulas. In light of Raz's lower bound, and of the connection between lower bounds and learning noticed by Linial, Mansour, and Nisan [35], the latter problem might be less hopeless than it looks. In this section we show a weaker result: that TreeBQP is contained in $\Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$, the third level of the polynomial hierarchy. Since BQP is not known to lie in PH, this result could be taken as weak evidence that TreeBQP $\neq$ BQP. (On the other hand, we do not yet have oracle evidence even for BQP $\not\subset$ AM, though not for lack of trying [2].)

Definition 21 TreeBQP *is the class of languages accepted by a* BQP *machine subject to the constraint that at every time step* t, *the machine's state* $\left|\psi^{(t)}\right\rangle$ *is exponentially close to a tree state. More formally, the initial state is* $\left|\psi^{(0)}\right\rangle = |0\rangle^{\otimes(p(n)-n)} \otimes |x\rangle$ *(for an input* $x \in \{0,1\}^n$ *and polynomial bound* p*), and a uniform classical polynomial-time algorithm generates a sequence of gates* $g^{(1)}, \ldots, g^{(p(n))}$. *Each* $g^{(t)}$ *can be either be selected from some finite universal basis of unitary gates (the choice will turn out not to matter), or can be a* 1*-qubit measurement. When we perform a measurement, the state evolves to one of two possible pure states, with the usual probabilities, rather than to a mixed state. We require that the final gate* $g^{(p(n))}$ *is a measurement of the first qubit. If at least one intermediate state* $\left|\psi^{(t)}\right\rangle$ *had* $\mathrm{TS}_{1/2^{\Omega(n)}}\left(\left|\psi^{(t)}\right\rangle\right) > p(n)$, *then the outcome of the final measurement is chosen adversarially; otherwise it is given by the usual Born probabilities. The measurement must return* 1 *with probability at least* 2/3 *if the input is in the language, and with probability at most* 1/3 *otherwise.*

Some comments on the definition: we allow $|\psi^{(t)}\rangle$ to deviate from a tree state by an exponentially small amount, in order to make the model independent of the choice of gate set. We allow intermediate measurements because otherwise it is unclear even how to simulate BPP.[8] The rule for measurements follows the "Copenhagen interpretation," in the sense that if a qubit is measured to be 1, then subsequent computation is not affected by what would have happened were the qubit measured to be 0. In particular, if measuring 0 would have led to states of tree size greater than $p(n)$, that does not invalidate the results of the path where 1 is measured.

The following theorem shows that $\mathsf{TreeBQP}$ has many of the properties we would want it to have.

Theorem 22

(i) If $|\psi^{(t+1)}\rangle$ *is obtained from* $|\psi^{(t)}\rangle$ *by applying a* k*-qubit unitary* $g^{(t+1)}$*, then*

$$\mathrm{TS}\left(\left|\psi^{(t+1)}\right\rangle\right) \leq 2^k\,\mathrm{TS}\left(\left|\psi^{(t)}\right\rangle\right) + 1.$$

(ii) The definition of $\mathsf{TreeBQP}$ *is invariant under the choice of gate set.*

(iii) The probabilities $(1/3, 2/3)$ *can be replaced by any* $(p, 1-p)$ *with* $2^{-2^{\sqrt{\log n}}} < p < 1/2$.

(iv) $\mathsf{BPP} \subseteq \mathsf{TreeBQP} \subseteq \mathsf{BQP}$.

Theorem 23 $\mathsf{TreeBQP} \subseteq \Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$.

In the proof of Theorem 23, the only fact about tree states we use is that $\mathsf{Tree} \subseteq \mathsf{AmpP}$; that is, there is a polynomial-time classical algorithm that computes the amplitude α_x of any basis state $|x\rangle$. So if we define AmpP-BQP analogously to $\mathsf{TreeBQP}$ except that any states in AmpP are allowed, then AmpP-$\mathsf{BQP} \subseteq \Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$ as well.

7 Experimental Proposal

We believe our results underscore the importance of experimental efforts to implement quantum computing. For whether or not those efforts lead eventually to a practical quantum computer, they are testing *quantum mechanics itself* in a way qualitatively different from all previous experiments we know of. In particular, neither the Bell inequality experiments of Aspect et al. [8] and their successors, nor "macroscopic" interference experiments such as those of Arndt et al. [7], sought to prepare quantum states with large tree size. At the same time, our results suggest new experiments that would test candidate Sure/Shor separators more directly. To illustrate, we pose the following challenge to experimentalists.

> *Prepare a 'generic' Clifford group state [24]—that is, a state obtainable by applying Hadamard,* CNOT*, and* $\pi/4$ *phase rotation gates, starting from* $|0\rangle^{\otimes n}$*—on as many qubits as possible. For concreteness, a 9-qubit state stabilized by the following 7 Pauli operators could be taken as a benchmark:* $XYXIIYIZI$, $ZIZXIXXXY$, $IYZYYZIII$, $XXIIIYXIZ$, $XIXIIYZXY$, $ZIIIZIYYZ$, $YXXZXYZIY$. *(See Nielsen and Chuang [37] for an explanation of stabilizer notation.)*

[8] If we try to simulate BPP in the standard way, we might produce complicated entanglement between the computation register and the register containing the random bits, and no longer have a tree state.

The point of this experiment is to extend the tree sizes for which quantum mechanics has been explored by roughly an order of magnitude. Were the experiment successful, we would consider it evidence against the hypothesis that all states in Nature are tree states. Since tree states are an asymptotic notion, obviously that hypothesis is formally unfalsifiable. However, the larger the n for which generic n-qubit Clifford group states were demonstrated, the more it would look like special pleading to claim tree states (or a subset thereof) as a Sure/Shor separator.

As evidence for our experiment's feasibility, we mention two experiments that have already been performed in liquid-state NMR. First, Knill, Laflamme, Martinez, and Negrevergne [30] demonstrated encoding, error detection, and error correction for a code that maps 1 qubit to 5 qubits. This involved preparing nuclear spin states such as the following:

$$|\psi\rangle=\frac{1}{4}\left(\begin{array}{c}|00000\rangle+|10010\rangle+|01001\rangle+|10100\rangle+|01010\rangle-|11011\rangle-|00110\rangle-|11000\rangle\\ -|11101\rangle-|00011\rangle-|11110\rangle-|01111\rangle-|10001\rangle-|01100\rangle-|10111\rangle+|00101\rangle\end{array}\right).$$

The best upper bound we know of on $\mathrm{TS}(|\psi\rangle)$ is 69, which follows from the decomposition

$$|\psi\rangle=\frac{1}{4}\left(\begin{array}{c}(|01\rangle+|10\rangle)\otimes(|010\rangle-|111\rangle)+(|01\rangle-|10\rangle)\otimes(|001\rangle-|100\rangle)\\ -(|00\rangle+|11\rangle)\otimes(|011\rangle+|110\rangle)+(|00\rangle-|11\rangle)\otimes(|000\rangle+|101\rangle)\end{array}\right).$$

Second, Knill, Laflamme, Martinez, and Tseng [31] created a 7-qubit cat state, $(|0000000\rangle+|1111111\rangle)/\sqrt{2}$. This has recently been improved to 12 qubits (R. Laflamme, personal communication).

Our proposed experiment differs in two respects from experiments designed to test quantum error-correction. First, we do not ask that error-correction mechanisms be demonstrated, only preparation of a codeword state. This might make it easier to work with more qubits. And second, it is important that the codeword state not have special structure that entails a small tree size.

But what about the Clifford group states mentioned in the challenge? Do *those* have small tree size? Observe that the coset states studied in Section 5.1 are obtainable from $|0\rangle^{\otimes n}$ by applying Hadamard and CNOT gates, and therefore form a subclass of Clifford group states (in which all amplitudes have the same phase). Also, using the techniques of that section, it is easy to show that a uniform random Clifford group state (suitably defined) has superpolynomial tree size with high probability.

This is not very convincing, however. The lower bounds in Section 5.1 are only asymptotic; worse, they have the form $n^{\varepsilon\log n}$ where ε in Raz's proof [39] is about 10^{-6}. So it would be good to compute tight lower bounds explicitly for small n. Since we have not found a feasible way to do that, instead we will give explicit *upper* bounds that we conjecture are close to tight. Given integers $0\leq k\leq n$, let $S(n,k)$ be the maximum, over all coset states $|C\rangle\in\mathcal{H}_2^{\otimes n}$ with $|C|=2^k$, of $\mathrm{TS}(|C\rangle)$. Also, let $S'(n,k)$ be the maximum, over all Clifford group states $|C\rangle\in\mathcal{H}_2^{\otimes n}$ with 2^k nonzero amplitudes, of $\mathrm{TS}(|G\rangle)$. Then we have the following inequalities (the second uses the Fourier transform as in Proposition 1, part (v), while the third and fifth use divide-and-conquer):

$$\begin{aligned}
&S(n,k)\leq 1+2^k(n+1)\\
&S(n,k)\leq 1+2^{n-k}(3n+1)\\
&S(n,k)\leq 1+2^{n-k}\min_{0<m<n}\left(1+S(m,k+m-n)+S(n-m,k-m)\right)\\
&S'(n,k)\leq 1+2^k(n+1)\\
&S'(n,k)\leq 1+2^{n-k+2}\min_{0<m<n,\,a\in\{1,2\}}\left(1+S'(m,k+m-n-a)+S'(n-m,k-m+a-2)\right).
\end{aligned}$$

The following tables list, for $1 \leq n \leq 12$ and $0 \leq k \leq n$, the best upper bounds on $S(n,k)$ and $S'(n,k)$ that we could obtain by applying the above inequalities as well as a few further optimizations (for example, collapsing multiple gates of the same type). Boldface indicates that divide-and-conquer yielded a smaller tree than either the naïve or Fourier strategies.

n / k	0	1	2	3	4	5	6	7	8	9	10	11	12	Max
1	1	3												3
2	3	7	7											7
3	4	9	17	10										17
4	5	11	21	27	13									27
5	6	13	25	49	33	16								49
6	7	15	29	57	**77**	39	19							77
7	8	17	33	65	**121**	89	45	22						121
8	9	19	37	73	145	**185**	101	51	25					185
9	10	21	41	81	161	**305**	225	113	57	28				305
10	11	23	45	89	177	353	**433**	249	125	63	31			433
11	12	25	49	97	193	385	**705**	545	273	137	69	34		705
12	13	27	53	105	209	417	833	**993**	593	297	149	75	37	993

Table 1: Upper bounds on $S(n,k)$, the tree size of an n-qubit coset state with 2^k nonzero amplitudes

n / k	0	1	2	3	4	5	6	7	8	9	10	11	12	Max
1	1	3												3
2	3	7	**11**											11
3	4	9	17	**25**										25
4	5	11	21	41	**53**									53
5	6	13	25	49	**89**	**85**								89
6	7	15	29	57	113	**153**	**133**							153
7	8	17	33	65	129	**225**	**233**	**189**						233
8	9	19	37	73	145	289	**369**	**345**	**301**					369
9	10	21	41	81	161	321	**545**	**561**	**537**	**413**				561
10	11	23	45	89	177	353	705	**865**	**817**	**793**	**541**			865
11	12	25	49	97	193	385	769	**1281**	**1313**	**1265**	**1177**	**733**		1313
12	13	27	53	105	209	417	833	1665	**1985**	**1889**	**1841**	**1689**	**957**	1985

Table 2: Upper bounds on $S'(n,k)$, the tree size of an n-qubit Clifford group state with 2^k nonzero amplitudes

Two patterns stand out in the above tables. First, for a fixed n, Clifford group states have about twice the maximum tree size of coset states—the reason being that trees for the former need to handle phases ($+1$, -1, $+i$, and $-i$). That is why our challenge specifies Clifford group states instead of the less general coset states. Second, when k is small, the minimum-size tree is obtained by simply summing all 2^k basis states with nonzero amplitude. But once k exceeds $n/2$, for coset states it becomes better to 'switch the Fourier basis' (and better still to use divide-and-conquer at an intermediate stage). Finally, when $k = n$, the only coset state is the uniform superposition, which has tree size $3n + 1$. For Clifford group states, by contrast, switching to the Fourier basis may not decrease the tree size even when k is large, so divide-and-conquer pays off for a larger range of k.

A caveat concerning our experiment is in order. When we say a pure state $|\psi\rangle \in \mathcal{H}_2^{\otimes n}$ is created in liquid NMR, what is actually created is the 'pseudo-pure' state $\rho = \varepsilon |\psi\rangle \langle\psi| + (1 - \varepsilon) 2^{-n} I_{2^n}$, where I_{2^n} is the identity (so $I_{2^n}/2^n$ is the maximally mixed state), and ε scales roughly as $n/2^n$ (in current experiments

$\varepsilon \approx 10^{-5}$). Braunstein et al. [11] have shown that, if the number n of qubits is less than about 14, then such states cannot be entangled. That is, there exists some representation of ρ as a mixture of pure states, each of which is a product $(\alpha_1 |0\rangle + \beta_1 |1\rangle) \otimes \cdots \otimes (\alpha_n |0\rangle + \beta_n |1\rangle)$ and which therefore has tree size $O(n)$. This is a well-known limitation of liquid NMR, in which molecules are kept at room temperature. To prove *conclusively* that states with large tree size exist, one would need to repeat our experiment in (say) solid-state NMR, in which nuclei are confined in a lattice and cooled to sub-milliKelvin temperatures, and consequently $\varepsilon \approx 1$.

A final remark: it seems conceivable that states with large tree size could have been created in condensed matter systems, for instance those exhibiting the quantum Hall effect. We do not know how to represent the states of such systems in terms of qubits; the issue deserves further investigation.

8 Conclusion and Open Problems

A crucial step in quantum computing was to separate the question of whether quantum computers can be built from the question of what one could do with them. This separation allowed computer scientists to make great advances on the latter question, despite knowing nothing about the former. We have argued, however, that the tools of computational complexity theory are relevant to both questions. The claim that quantum computing is possible in principle is really a claim that certain *states* can exist—that quantum mechanics will not break down if we try to prepare those states. Moreover, what distinguishes these states from states we have seen must be more than precision in amplitudes, or the number of qubits maintained coherently. The distinguishing property, the Sure/Shor separator, must instead be some sort of *complexity*. That is, Sure states must have succinct representations of a type that Shor states do not.

We have tried to show that, by adopting this viewpoint, we make the debate about whether quantum computing is possible less ideological and more scientific. By studying particular examples of Sure/Shor separators, quantum computing skeptics would strengthen their case—for they would then have a plausible research program aimed at identifying what, exactly, the barriers to quantum computation are. We hope, however, that the 'complexity theory of quantum states' initiated in this paper will be taken up by quantum computing proponents as well. This theory offers a new perspective on the transition from quantum to classical computing, and a new connection between quantum computing and the powerful circuit lower bound techniques of classical complexity theory.

We end with some open problems.

(1) Can Raz's technique be improved to show exponential rather than quasipolynomial tree size lower bounds? What if we restrict to orthogonal or manifestly orthogonal trees?

(2) Can we prove Conjecture 19 about primes, implying an $n^{\Omega(\log n)}$ tree size lower bound for Shor states? A perhaps related question is, can we prove Theorem 13 for particular examples of codes, such as Reed-Solomon codes, rather than a uniform random code?

(3) Corollary 14 shows a superpolynomial separation between formula size and multilinear formula size for a function $f : \{0,1\}^n \rightarrow \mathbb{R}$. Can we achieve such a separation for a multilinear polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$? (We thank R. Raz for this problem.)

(4) Let $|\varphi\rangle$ be a uniform superposition over all n-bit strings of Hamming weight $n/2$. It is easy to show by divide-and-conquer that $\mathrm{TS}(|\varphi\rangle) = n^{O(\log n)}$. Is this upper bound tight? More generally, can we show a superpolynomial tree size lower bound for any state with permutation symmetry? Such states can arise, for example, in ion traps with pairwise-interaction Hamiltonians.

(5) Is $\mathsf{Tree} = \mathsf{OTree}$? That is, are there tree states that are not orthogonal tree states?

(6) Is the tensor-sum hierarchy of Section 3 infinite? That is, do we have $\Sigma_\mathsf{k} \neq \Sigma_{\mathsf{k+1}}$ for all k?

(7) Is $\mathsf{TreeBQP} = \mathsf{BPP}$? That is, can a quantum computer that is always in a tree state be simulated classically? The key question seems to be whether the concept class of multilinear formulas is efficiently learnable. Intriguingly, Klivans and Shpilka [29] gave an efficient algorithm to learn depth-3 multilinear formulas. However, the quantum simulation algorithm that their result implies turns out to be subsumed by the simulation algorithm of Vidal [44].

(8) Can we obtain explicit lower bounds on tree size for, say, 9-qubit Clifford group states? If our proposed experiment succeeded, such lower bounds would provide more confidence in interpreting the results.

(9) What is the best way to extend our theory to mixed states? Possibilities include letting $\mathrm{TS}(\rho)$ be the minimum of $\mathrm{TS}(|\psi\rangle)$ over all purifications $|\psi\rangle$ of ρ, or letting $\mathrm{TS}(\rho)$ be the minimum, over all $|\psi_1\rangle, \ldots, |\psi_m\rangle$ such that $\rho = \sum_{i=1}^{m} \lambda_i |\psi_i\rangle \langle\psi_i|$, of $\max_i \mathrm{TS}(|\psi_i\rangle)$.

Acknowledgments

I thank Ran Raz for fruitful correspondence and for sharing an early version of his paper; Leonid Levin for clarifying his views; and Mike Mosca, Ashwin Nayak, John Preskill, Umesh Vazirani, Guifre Vidal, and Avi Wigderson for helpful discussions.

References

[1] S. Aaronson. Book review on A New Kind of Science, *Quantum Information and Computation (QIC)* 2(5), 2002. quant-ph/0206089.

[2] S. Aaronson. Quantum lower bound for recursive Fourier sampling, *Quantum Information and Computation (QIC)* 3(2), 2003. quant-ph/0209060.

[3] D. S. Abrams and S. Lloyd. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and #P problems, *Phys. Rev. Lett.* 81:3992–3995, 1998. quant-ph/9801041.

[4] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error, in *Proc. ACM STOC'1997*, pp. 176–188, 1997. quant-ph/9906129.

[5] D. Aharonov and A. Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge, in *Proc. of ACM STOC'03*, pp. 20–29, 2003. quant-ph/0301023.

[6] A. Ambainis, L. Schulman, A. Ta-Shma, U. Vazirani, and A. Wigderson. The quantum communication complexity of sampling, to appear in *SIAM J. Comp.*, 2003. Earlier version in *Proc. IEEE FOCS'98*, pp. 342–351, 1998.

[7] M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, and A. Zeilinger. Wave-particle duality of C_{60} molecules, *Nature* 401:680–682, 1999.

[8] A. Aspect, P. Grangier, and G. Roger. Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: a new violation of Bell's inequalities, *Phys. Rev. Lett.* 49:91–94, 1982.

[9] E. Bernstein and U. Vazirani. Quantum complexity theory, *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[10] M. L. Bonet and S. R. Buss. Size-depth tradeoff for Boolean formulae, *Inf. Proc. Lett.* 11:151–155, 1994.

[11] S. L. Braunstein, C. M. Caves, N. Linden, S. Popescu, and R. Schack. Separability of very noisy mixed states and implications for NMR quantum computing, *Phys. Rev. Lett.* 83:1054–1057, 1999. quant-ph/9811018.

[12] R. P. Brent. The parallel evaluation of general arithmetic expressions, *J. ACM* 21:201–206, 1974.

[13] N. H. Bshouty, R. Cleve, and W. Eberly. Size-depth tradeoffs for algebraic formulae, in *Proc. IEEE FOCS'91*, pp. 334–341, 1991.

[14] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic Complexity Theory*, Springer-Verlag, 1997.

[15] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist, *Phys. Rev. A* 54:1098–1105, 1996.

[16] J. Cronin. CP symmetry violation—the search for its origin, Nobel Lecture, December 8, 1980. http://www.nobel.se/physics/laureates/1980/cronin-lecture.pdf.

[17] V. Fitch. The discovery of charge-conjugation parity asymmetry, Nobel Lecture, December 8, 1980. http://www.nobel.se/physics/laureates/1980/fitch-lecture.pdf.

[18] J. R. Friedman, V. Patel, W. Chen, S. K. Tolpygo, and J. E. Lukens. Quantum superposition of distinct macroscopic states, *Nature* 406:43–46, 2000.

[19] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial time hierarchy, *Math. Systems Theory* 17:13–27, 1984.

[20] S. B. Gashkov. The complexity of the realization of Boolean functions by networks of functional elements and by formulas in bases whose elements realize continuous functions, *Prob. Kibernetiki* 37:52–118, 1980.

[21] G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems, *Phys. Rev. D*, 34:470, 1986.

[22] S. Goldwasser and M. Sipser. Private coins versus public coins in interactive proof systems, in *Randomness and Computation*, Vol. 5 of *Advances in Computing Research*, JAI Press, 1989.

[23] D. Gottesman. Class of quantum error-correcting codes saturating the quantum Hamming bound, *Phys. Rev. A.* 54:1862–1868, 1996. quant-ph/9604038.

[24] D. Gottesman. The Heisenberg representation of quantum computers, *Int. Conf. on Group Theoretic Methods in Physics*, 1998. quant-ph/9807006.

[25] A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix, *Duke J. Math.* 20:37–39, 1953.

[26] G. 't Hooft. Quantum gravity as a dissipative deterministic system, *Classical and Quantum Gravity* 16:3263–3279, 1999.

[27] D. Janzing, P. Wocjan, and T. Beth. Cooling and low energy state preparation for 3-local Hamiltonians are FQMA-complete, manuscript, 2003. quant-ph/0303186.

[28] A. Yu. Kitaev. Quantum computation: algorithms and error correction, *Russian Math. Surveys* 52(6):1191-1249, 1997.

[29] A. Klivans and A. Shpilka. Learning arithmetic circuits via partial derivatives, in *Proc. 16th Annual Conf. on Computational Learning Theory (COLT)*, pp. 463–476, 2003.

[30] E. Knill, R. Laflamme, R. Martinez, and C. Negrevergne. Implementation of the five qubit error correction benchmark, *Phys. Rev. Lett.* 86:5811, 2001. quant-ph/0101034.

[31] E. Knill, R. Laflamme, R. Martinez, and C.-H. Tseng. An algorithmic benchmark for quantum information processing, *Nature* 404:368–370, 2000. quant-ph/9908051.

[32] E. Knill, R. Laflamme, and W. Zurek. Resilient quantum computation, *Science* 279:342, 1998. quant-ph/9702058.

[33] L. Levin. Polynomial time and extravagant models, in The tale of one-way functions, *Problems of Information Transmission* 39(1), 2003. http://www.cs.bu.edu/fac/lnd/misc/qc.html.

[34] L. Levin, D. Gottesman, P. Shor, et al. Discussion on sci.physics.research newsgroup, February-March 2000. http://www.google.com/groups?threadm=FrsHL0.ILr%40research.att.com.

[35] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability, *J. ACM* 40(3):607–620, 1993.

[36] M. Naor and O. Reingold. Number-theoretic constructions of efficient pseudo-random functions, in *Proc. IEEE FOCS'97*, pp. 458–467, 1997.

[37] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*, Cambridge Univ. Press, 2000.

[38] R. Penrose. *The Emperor's New Mind*, Oxford Univ. Press, 1989.

[39] R. Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size, submitted, 2003. ECCC TR03-067.

[40] A. A. Razborov and S. Rudich. Natural proofs, *J. Comp. Sys. Sci.* 55:24–35, 1997.

[41] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. Comput.* 26(5):1484–1509, 1997. quant-ph/9508027.

[42] A. Steane. Multiple particle interference and quantum error correction, *Proc. Roy. Soc. Lond. A*, 452:2551–2577, 1996. quant-ph/9601029.

[43] G. Turán and F. Vatan. On the computation of Boolean functions by analog circuits of bounded fan-in (extended abstract), in *Proc. IEEE FOCS'94*, pp. 553–564, 1994.

[44] G. Vidal. Efficient classical simulation of slightly entangled quantum computations, *Phys. Rev. Lett.* 91:147902, 2003. quant-ph/0301063.

[45] H. E. Warren. Lower bounds for approximation by non-linear manifolds, *Trans. Amer. Math. Soc.* 133:167–178, 1968.

[46] S. Wolfram. *A New Kind of Science*, Wolfram Media, 2002.

9 Appendix: Proofs

Proof of Proposition 1.

(i) Obvious ($\mathsf{Vidal} \subseteq \mathsf{Circuit}$ follows from results of Vidal [44]).

(ii) Using results in [44], a Vidal state $|\psi_n\rangle$ can be recursively decomposed as $\sum_{i=1}^{\chi(|\psi_n\rangle)} \alpha_i \left|\varphi_i^A\right\rangle \otimes \left|\varphi_i^B\right\rangle$, where $A = \{1, \ldots, \lfloor n/2 \rfloor\}$ and for each i, $\chi\left(\left|\varphi_i^A\right\rangle\right) \leq \chi\left(|\psi_n\rangle\right)$ and $\chi\left(\left|\varphi_i^B\right\rangle\right) \leq \chi\left(|\psi_n\rangle\right)$.

(iii) $\Sigma_2 \subseteq \mathsf{Vidal}$ follows since a sum of t separable states has $\chi \leq t$, while $\otimes_2 \not\subset \mathsf{Vidal}$ follows from the example of $n/2$ Bell pairs: $2^{-n/4}\left(|00\rangle + |11\rangle\right)^{\otimes n/2}$.

(iv) $\otimes_2 \subseteq \mathsf{MOTree}$ is obvious, while $\mathsf{MOTree} \not\subset \otimes_2$ follows from the example of $|P_n\rangle$, an equal superposition over all n-bit strings of even parity. Using divide-and-conquer, $\text{MOTS}\left(|P_n\rangle\right) = O\left(n^2\right)$, while $|P_n\rangle \notin \otimes_2$ follows from $|P_n\rangle \notin \Sigma_1$ and the fact that $|P_n\rangle$ has no nontrivial tensor product decomposition.

(v) $\otimes_1 \not\subset \Sigma_1$ and $\Sigma_1 \not\subset \otimes_1$ are obvious. $\otimes_2 \not\subset \Sigma_2$ (and hence $\otimes_1 \neq \otimes_2$) follows from part (iii). $\Sigma_2 \not\subset \otimes_2$ (and hence $\Sigma_1 \neq \Sigma_2$) follows from part (iv), together with the fact that $|P_n\rangle$ has a Σ_2 formula based on the Fourier transform:

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

$\Sigma_2 \neq \Sigma_3$ follows from $\otimes_2 \not\subset \Sigma_2$ and $\otimes_2 \subseteq \Sigma_3$. Also, $\Sigma_3 \not\subset \otimes_3$ follows from $\Sigma_2 \neq \Sigma_3$, together with the fact that we can easily construct states in $\Sigma_3 \setminus \Sigma_2$ that have no nontrivial tensor product decomposition. $\otimes_2 \neq \otimes_3$ follows from $\Sigma_2 \not\subset \otimes_2$ and $\Sigma_2 \subseteq \otimes_3$. Finally, $\otimes_3 \neq \Sigma_4 \cap \otimes_4$ follows from $\Sigma_3 \not\subset \otimes_3$ and $\Sigma_3 \subseteq \Sigma_4 \cap \otimes_4$.

■

Proof of Theorem 5.

(i) Given a tree representing $|\psi\rangle$, replace every unbounded fan-in gate by a collection of binary gates, every $\otimes$ by $\times$, every $|1\rangle_i$ vertex by x_i, and every $|0\rangle_i$ vertex by a formula for $1 - x_i$. Push all multiplications by constants at the edges down to $\times$ gates at the leaves.

(ii) Given a multilinear formula Φ for f_ψ, let $p(v)$ be the polynomial computed at vertex v of Φ, and let $S(v)$ be the set of variables that appears in $p(v)$. First, call Φ *syntactic* if at every $\times$ gate with children v and w, $S(v) \cap S(w) = \varnothing$. A lemma of Raz [39] states that we can always make Φ syntactic without increasing its size.

Second, at every $+$ gate u with children v and w, enlarge both $S(v)$ and $S(w)$ to $S(v) \cup S(w)$, by multiplying $p(v)$ by $x_i + (1 - x_i)$ for every $x_i \in S(w) \setminus S(v)$, and multiplying $p(w)$ by $x_i + (1 - x_i)$ for every $x_i \in S(v) \setminus S(w)$. Doing this does not invalidate any $\times$ gate that is an ancestor of u, since by the assumption that Φ is syntactic, $p(u)$ is never multiplied by any polynomial containing variables in $S(v) \cup S(w)$. Similarly, enlarge $S(r)$ to $\{x_1, \ldots, x_n\}$ where r is the root of Φ.

Third, call v *max-linear* if $|S(v)| = 1$ but $|S(w)| > 1$ where w is the parent of v. If v is max-linear and $p(v) = a + bx_i$, then replace the tree rooted at v by a tree computing $a|0\rangle_i + (a+b)|1\rangle_i$. Also, replace all multiplications by constants higher in Φ by multiplications at the edges. (Because of the second step, there are no additions by constants higher in Φ.) Replacing every $\times$ by $\otimes$ then gives a tree representing $|\psi\rangle$, whose size is easily seen to be $O(|\Phi| + n)$.

(iii) Apply the reduction from part (i). Let the resulting multilinear formula compute polynomial p; then

$$\sum_{x \in \{0,1\}^n} |p(x) - f_\psi(x)|^2 = 2 - 2 \sum_{x \in \{0,1\}^n} p(x) \overline{f_\psi(x)} \leq 2 - 2\sqrt{1-\varepsilon} = \delta.$$

(iv) Apply the reduction from part (ii). Let $(\beta_x)_{x \in \{0,1\}^n}$ be the resulting amplitude vector; since this vector might not be normalized, divide each β_x by $\sum_x |\beta_x|^2$ to produce β'_x. Then

$$\begin{aligned} \left| \sum_{x \in \{0,1\}^n} \beta'_x \overline{\alpha_x} \right|^2 &= 1 - \frac{1}{2} \sum_{x \in \{0,1\}^n} |\beta'_x - \alpha_x|^2 \\ &\geq 1 - \frac{1}{2} \left(\sqrt{\sum_{x \in \{0,1\}^n} |\beta'_x - \beta_x|^2} + \sqrt{\sum_{x \in \{0,1\}^n} |\beta_x - \alpha_x|^2} \right)^2 \\ &\geq 1 - \frac{1}{2} \left(2\sqrt{\varepsilon}\right)^2 = 1 - 2\varepsilon. \end{aligned}$$

■

Proof of Theorem 6. A classical theorem of Brent [12] says that given an arithmetic formula Φ, there exists an equivalent formula of depth $O(\log|\Phi|)$ and size $O(|\Phi|^c)$, where c is a constant. Bshouty, Cleve, and Eberly [13] (see also Bonet and Buss [10]) improved Brent's theorem to show that c can be taken to be $1 + \varepsilon$ for any $\varepsilon > 0$. So it suffices to show that, for 'division-free' formulas, these theorems preserve multilinearity (and in the MOTS case, preserve manifest orthogonality).

Brent's theorem is proven by induction on $|\Phi|$. Here is a sketch: choose a subformula I of Φ size between $|\Phi|/3$ and $2|\Phi|/3$ (which one can show always exists). Then identifying a subformula with the polynomial computed at its root, $\Phi(x)$ can be written as $G(x) + H(x)I(x)$ for some formulas G and H. Furthermore, G and H are both obtainable from Φ by removing I and then applying further restrictions. So $|G|$ and $|H|$ are both at most $|\Phi| - |I| + O(1)$. Let $\widehat{\Phi}$ be a formula equivalent to Φ that evaluates G, H, and I separately, and then returns $G(x) + H(x)I(x)$. Then $\left|\widehat{\Phi}\right|$ is larger than $|\Phi|$ by at most a constant factor, while by the induction hypothesis, we can assume the formulas for G, H, and I have logarithmic depth. Since the number of induction steps is $O(\log|\Phi|)$, the total depth is logarithmic and the total blowup in formula size is polynomial in $|\Phi|$. Bshouty, Cleve, and Eberly's improvement uses a more careful decomposition of Φ, but the basic idea is the same.

Now, if Φ is syntactic multilinear, then clearly G, H, and I are also syntactic multilinear. Furthermore, H cannot share variables with I, since otherwise a subformula of Φ containing I would have been multiplied by a subformula containing variables from I. Thus multilinearity is preserved. To see that manifest orthogonality is preserved, suppose we are evaluating G and H 'bottom up,' and let G_v and H_v be the polynomials computed at vertex v of Φ. Let $v_0 = \operatorname{root}(I)$, let v_1 be the parent of v_0, let v_2 be the parent of v_1, and so on until $v_k = \operatorname{root}(\Phi)$. It is clear that, for every x, either $G_{v_0}(x) = 0$ or $H_{v_0}(x) = 0$. Furthermore, suppose that property holds for $G_{v_{i-1}}, H_{v_{i-1}}$; then by induction it holds for G_{v_i}, H_{v_i}. If v_i is a $\times$ gate, then this follows from multilinearity (if $|\psi\rangle$ and $|\varphi\rangle$ are manifestly orthogonal, then $|0\rangle \otimes |\psi\rangle$ and $|0\rangle \otimes |\varphi\rangle$ are also manifestly orthogonal). If v_i is a $+$ gate, then letting $\operatorname{supp}(p)$ be the set of x such that $p(x) \neq 0$, any polynomial p added to $G_{v_{i-1}}$ or $H_{v_{i-1}}$ must have

$$\operatorname{supp}(p) \cap \left(\operatorname{supp}\left(G_{v_{i-1}}\right) \cup \operatorname{supp}\left(H_{v_{i-1}}\right)\right) = \emptyset,$$

and manifest orthogonality follows. ■

Proof of Theorem 7. Let $\Gamma(|\psi\rangle)$ be the minimum size of a circuit needed to prepare $|\psi\rangle \in \mathcal{H}_2^{\otimes n}$ starting from $|0\rangle^{\otimes n}$. We prove by induction on $\Gamma(|\psi\rangle)$ that $\Gamma(|\psi\rangle) \leq q(\operatorname{OTS}(|\psi\rangle))$ for some polynomial q. The base case $\operatorname{OTS}(|\psi\rangle) = 1$ is clear. Let T be an orthogonal state tree for $|\psi\rangle$, and assume without loss of generality that every gate has fan-in 2 (this increases $|T|$ by at most a constant factor). Let T_1 and T_2 be the subtrees of $\operatorname{root}(T)$, representing states $|\psi_1\rangle$ and $|\psi_2\rangle$ respectively; note that $|T| = |T_1| + |T_2| + 1$. First suppose $\operatorname{root}(T)$ is a $\otimes$ gate; then clearly $\Gamma(|\psi\rangle) \leq \Gamma(|\psi_1\rangle) + \Gamma(|\psi_2\rangle)$.

Second, suppose $\operatorname{root}(T)$ is a $+$ gate, with $|\psi\rangle = \alpha|\psi_1\rangle + \beta|\psi_2\rangle$ and $\langle\psi_1|\psi_2\rangle = 0$. Let U be a quantum circuit that prepares $|\psi_1\rangle$, and V be a circuit that prepares $|\psi_2\rangle$. Then we can prepare $\alpha|0\rangle|0\rangle^{\otimes n} + \beta|1\rangle U^{-1}V|0\rangle^{\otimes n}$. Observe that $U^{-1}V|0\rangle^{\otimes n}$ is orthogonal to $|0\rangle^{\otimes n}$, since $|\psi_1\rangle = U|0\rangle^{\otimes n}$ is orthogonal to $|\psi_2\rangle = V|0\rangle^{\otimes n}$. So applying a NOT to the first register, conditioned on the OR of the bits in the second register, yields $|0\rangle \otimes \left(\alpha|0\rangle^{\otimes n} + \beta U^{-1}V|0\rangle^{\otimes n}\right)$, from which we obtain $\alpha|\psi_1\rangle + \beta|\psi_2\rangle$ by applying U to the second register. The size of the circuit used is $O(|U| + |V| + n)$, with a possible constant-factor blowup arising from the need to condition on the first register. If we are more careful, however, we can combine the 'conditioning' steps across multiple levels of the recursion, producing a circuit of size $|V| + O(|U| + n)$. By symmetry, we can also reverse the roles of U and V to obtain a circuit of size $|U| + O(|V| + n)$. Therefore

$$\Gamma(|\psi\rangle) \leq \min\left\{\Gamma(|\psi_1\rangle) + c\Gamma(|\psi_2\rangle) + cn,\ c\Gamma(|\psi_2\rangle) + \Gamma(|\psi_1\rangle) + cn\right\}$$

for some constant $c \geq 2$. Solving this recurrence we find that $\Gamma(|\psi\rangle)$ is polynomial in $\operatorname{OTS}(|\psi\rangle)$. ■

Proof of Theorem 8. To generate a uniform random state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$, we can choose $\widehat{\alpha}_x, \widehat{\beta}_x \in \mathbb{R}$ for each x independently from a Gaussian distribution with mean 0 and variance 1, then let $\alpha_x = \left(\widehat{\alpha}_x + i\widehat{\beta}_x\right)/\sqrt{R}$ where $R = \sum_{x\in\{0,1\}^n}\left(\widehat{\alpha}_x^2 + \widehat{\beta}_x^2\right)$. Let

$$\Lambda_\psi = \left\{x : (\operatorname{Re}\alpha_x)^2 < \frac{1}{4\cdot 2^n}\right\},$$

and let $\mathcal{G}$ be the set of $|\psi\rangle$ for which $|\Lambda_\psi| < 2^n/5$. We claim that $\Pr_{|\psi\rangle}[|\psi\rangle \in \mathcal{G}] = 1 - o(1)$. First, $\operatorname{EX}[R] = 2^{n+1}$, so by a standard Hoeffding-type bound, $\Pr[R < 2^n]$ is doubly-exponentially small in n. Second, assuming $R \geq 2^n$, for each x

$$\Pr[x \in \Lambda_\psi] \leq \Pr\left[\widehat{\alpha}_x^2 < \frac{1}{4}\right] = \operatorname{erf}\left(\frac{1}{4\sqrt{2}}\right) < 0.198,$$

and the claim follows by a Chernoff bound.

For $g : \{0,1\}^n \to \mathbb{R}$, let $A_g = \{x : \operatorname{sgn}(g(x)) \neq \operatorname{sgn}(\operatorname{Re}\alpha_x)\}$, where $\operatorname{sgn}(y)$ is 1 if $y \geq 0$ and -1 otherwise. Then if $|\psi\rangle \in \mathcal{G}$, clearly

$$\sum_{x\in\{0,1\}^n} |g(x) - f_\psi(x)|^2 \geq \frac{|A_g| - |\Lambda_\psi|}{4\cdot 2^n}$$

where $f_\psi(x) = \operatorname{Re}\alpha_x$, and thus

$$|A_g| \leq \left(4\|g - f_\psi\|_2^2 + \frac{1}{5}\right)2^n.$$

Therefore to show that $\mathrm{MFS}_{1/15}(f_\psi) = 2^{\Omega(n)}$ with probability $1 - o(1)$, we need only show that for almost all Boolean functions $f : \{0,1\}^n \to \{-1,1\}$, there is no arithmetic formula Φ of size $2^{o(n)}$ such that

$$|\{x : \operatorname{sgn}(\Phi(x)) \neq f(x)\}| \leq 0.49 \cdot 2^n.$$

Here an arithmetic formula is real-valued, and can include addition, subtraction, and multiplication gates of fan-in 2 as well as constants. We do not need to assume multilinearity, and it is easy to see that the assumption of bounded fan-in is without loss of generality. Let W be the set of Boolean functions *sign-represented* by an arithmetic formula Φ of size $2^{o(n)}$, in the sense that $\operatorname{sgn}(\Phi(x)) = f(x)$ for all x. Then it suffices to show that $|W| = 2^{2^{o(n)}}$, since the number of functions sign-represented on an 0.51 fraction of inputs is at most $|W| \cdot 2^{2^n H(0.51)}$. (Here H denotes the binary entropy function.)

Let Φ be an arithmetic formula that takes as input the binary string $x = (x_1, \ldots, x_n)$ as well as constants $c_1, c_2, \ldots$. Let Φ_c denote Φ under a particular assignment c to $c_1, c_2, \ldots$. Then a result of Gashkov [20] (see also Turán and Vatan [43]), which follows from Warren's Theorem [45] in real algebraic geometry, shows that as we range over all c, Φ_c sign-represents at most $\left(2^{n+4}|\Phi|\right)^{|\Phi|}$ distinct Boolean functions, where $|\Phi|$ is the size of Φ. Furthermore, excluding constants, the number of distinct arithmetic formulas of size $|\Phi|$ is at most $\left(3|\Phi|^2\right)^{|\Phi|}$. When $|\Phi| = 2^{o(n)}$, this gives $\left(3|\Phi|^2\right)^{|\Phi|} \cdot \left(2^{n+4}|\Phi|\right)^{|\Phi|} = 2^{2^{o(n)}}$. We have shown that $\mathrm{MFS}_{1/15}(f_\psi) = 2^{\Omega(n)}$; by Theorem 5, part (iii), this implies that $\mathrm{TS}_{1/16}(|\psi\rangle) = 2^{\Omega(n)}$. ∎

Proof of Corollary 9. It is clear from Theorem 8 that there exists a state $|\varphi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$ such that $\mathrm{TS}_{1/16}(|\varphi\rangle) = 2^{\Omega(n)}$ and $\alpha_{0^n} = 0$. Take $|\psi\rangle = \sqrt{1-\delta}\,|0\rangle^{\otimes n} + \sqrt{\delta}\,|\varphi\rangle$. Since $\left|\langle\psi|0\rangle^{\otimes n}\right|^2 = 1 - \delta$, we have $\mathrm{MOTS}_\delta(|\psi\rangle) = n + 1$. On the other hand, suppose some $|\phi\rangle = \sum_{x\in\{0,1\}^n} \beta_x |x\rangle$ with $\mathrm{TS}(|\phi\rangle) = 2^{o(n)}$ satisfies $|\langle\phi|\psi\rangle|^2 \geq 1 - \varepsilon$. Then

$$\sum_{x \neq 0^n} \left(\sqrt{\delta}\alpha_x - \beta_x\right)^2 \leq 2 - 2\sqrt{1-\varepsilon}.$$

Thus, letting $f_\varphi(x) = \alpha_x$, we have $\mathrm{MFS}_c(f_\varphi) = O(\mathrm{TS}(|\phi\rangle) + n)$ where $c = \left(2 - 2\sqrt{1-\varepsilon}\right)/\delta$. By Theorem 5, part (iv), this implies that $\mathrm{TS}_{2c}(|\varphi\rangle) = O(\mathrm{TS}(|\phi\rangle) + n)$. But $2c = 1/16$ when $\varepsilon = \delta/32 - \delta^2/4096$, contradiction. ∎

Proof of Corollary 11. Suppose there were a function $g : \{0,1\}^n \to \mathbb{C}$ with $\mathrm{MFS}(g) = n^{o(\log n)}$, such that $\|g - f\|_2^2 \leq \varepsilon$. Then by Theorem 10 we would have that for all c,

$$\Pr_{R\in\mathcal{D}_k}\left[\operatorname{rank}\left(M_{g|R}\right) \geq c2^k\right] = o(1).$$

Furthermore, by Markov's inequality

$$\Pr_{R\in\mathcal{D}_k}\left[\left\|M_{g|R} - M_{f|R}\right\|_2^2 > \delta\right] < \frac{1}{\kappa},$$

the distribution over f values being uniform. Hence

$$\Pr_{R\in\mathcal{D}_k}\left[\operatorname{rank}_\delta\left(M_{f|R}\right) \geq c2^k\right] \leq \frac{1}{\kappa} + o(1).$$

■

Proof of Lemma 15. The Hoffman-Wielandt inequality [25] (see also [6]) states that for any two $N \times N$ matrices M, P,

$$\sum_{i=1}^{N} \left(\sigma_i\left(M\right) - \sigma_i\left(P\right)\right)^2 \leq \left\|M - P\right\|_2^2,$$

where $\sigma_i\left(M\right)$ is the i^{th} singular value of M (that is, $\sigma_i\left(M\right) = \sqrt{\lambda_i\left(M\right)}$, where $\lambda_1\left(M\right) \geq \cdots \geq \lambda_N\left(M\right) \geq 0$ are the eigenvalues of MM^*, and M^* is the conjugate transpose of M). Clearly $\sigma_i\left(I_N\right) = 1$ for all i. On the other hand, M has only $\operatorname{rank}\left(M\right)$ nonzero singular values, so

$$\sum_{i=1}^{N} \left(\sigma_i\left(M\right) - \sigma_i\left(I_N\right)\right)^2 \geq N - \operatorname{rank}\left(M\right).$$

■

Proof of Theorem 17. Let $C_1, \ldots, C_k$ be size-2^{n-1} cosets of $\mathbb{Z}_2^n$, each drawn independently from $\mathcal{E}_{1,n}$. (The probability that any of the cosets have size 0 or 2^n is exponentially small.) Then let $\left|\varphi\right\rangle = \alpha\left(\left|C_1\right\rangle + \cdots + \left|C_k\right\rangle\right)$, where α is a suitable normalizing constant. Our first claim is that $\operatorname{TS}\left(\left|C_i\right\rangle\right) = O\left(n\right)$ for all i, from which it follows that $\operatorname{TS}\left(\left|\varphi\right\rangle\right) = O\left(n^{4/3}\right)$. Since each C_i is just the set of strings satisfying some linear equation mod 2, it suffices to recall from Proposition 1, part (v), that $\operatorname{TS}\left(\left|P_n\right\rangle\right) = O\left(n\right)$, where $\left|P_n\right\rangle$ is the uniform superposition over all n-bit strings having even parity. We actually obtain a formula of depth 2, implying $\left|\varphi\right\rangle \in \Sigma_2$.

Our second claim is that $\operatorname{MOTS}\left(\left|\varphi\right\rangle\right) = n^{\Omega(\log n)}$ with probability $\Omega\left(1\right)$ over $C_1, \ldots, C_k$. Let $f_\varphi\left(x\right)$ be the amplitude of basis state $\left|x\right\rangle$ in $\left|\varphi\right\rangle$. Let $C = C_1 \cap \cdots \cap C_k$, and let $\beta = f_\varphi\left(x\right)$ for any $x \in C$ (clearly all elements of C have the same amplitude). Then β is the maximum amplitude, in the sense that $f_\varphi\left(x'\right) < \beta$ for any $x' \notin C$. Let R be a random k-restriction of f_φ, and let $M_{\varphi|R}$ be the $2^k \times 2^k$ matrix whose (y,z) entry is $f_{\varphi|R}\left(y,z\right)$. Then by the analysis of Theorem 13, with probability at least 0.288^2 it holds that for every setting of $y = y_1 \ldots y_k$, there exists a unique setting of $z = z_1 \ldots z_k$ such that $f_{\varphi|R}\left(y,z\right) = \beta$. In that case it is clear that $\operatorname{morank}\left(M_{\varphi|R}\right) = 2^k$, since a rank-1 matrix can cover at most one (y,z) such that $f_{\varphi|R}\left(y,z\right) = \beta$. Thus $\operatorname{MOTS}\left(\left|\varphi\right\rangle\right) = n^{\Omega(\log n)}$ by Theorem 12. ■

Proof of Proposition 18.

(i) Take the formula for $f_{n+\log p,p,b}$, and restrict the most significant $\log p$ bits to sum to a number congruent to $b - a \bmod p$ (this is always possible since $\mathbb{Z}_p$ is cyclic).

(ii) Obvious.

(iii) We use the Fourier transform, similarly to Proposition 1, part (v). We have

$$f_{n,p}\left(x\right) = \frac{1}{2^{n/2}\sqrt{p}} \sum_{h=0}^{p-1} \prod_{j=1}^{n} \exp\left(2\pi i h 2^{n-j} x_j / p\right)$$

which immediately yields a sum-of-products formula of size $O\left(np\right)$.

■

Proof of Proposition 20. Let $k = n^{1/c} + 5$, and let R be a uniform random k-restriction of $f_{n,p}$; that is, letting $x = 2^{n-1}x_{n-1} + \cdots + 2^0 x_0$, it restricts $n - 2k$ randomly chosen x_i bits to 0 or 1 both with

$1/2$ probability, and renames the remaining $2k$ bits $y_1, \ldots, y_k, z_1, \ldots, z_k$. Let S_y be the set of 2^k integers whose n-bit binary representation agrees with R and has $z_1 = \cdots = z_k = 0$, so that only $y_1, \ldots, y_k$ are allowed to vary. Likewise, let S_z be the set of 2^k integers whose binary representation agrees with R and has $y_1 = \cdots = y_k = 0$. Then setting $t = 2^{n^{1/c}}$, we have $|S_y| = |S_z| = 32t$. Also, if $x \in S_y \cup S_z$ then $x \leq 2^n = 2^{(\log t)^c}$. Since p is drawn uniformly from $[t, 5t/4]$, Conjecture 19 together with the union bound implies that

$$\Pr_p \left[\left(|S_y \,(\operatorname{mod} p)| \geq \frac{3t}{4} \right) \wedge \left(|S_z \,(\operatorname{mod} p)| \geq \frac{3t}{4} \right) \right] \geq \frac{1}{2}.$$

When this occurs, then by a pigeonhole argument, for every $a \in \mathbb{Z}_p$ there exist at least $t/4$ distinct congruence classes $b \in S_y \,(\operatorname{mod} p)$ such that $b+c \equiv a \,(\operatorname{mod} p)$ for some $c \in S_z \,(\operatorname{mod} p)$. Letting $M_{p|R}$ be the $2^k \times 2^k$ matrix whose (y,z) entry is $f_{n,p|R}(y,z)$, it follows that $\operatorname{rank}(M_{p|R}) \geq t/4 = \Omega(2^k)$. Thus $\operatorname{MFS}(f_{n,p}) = n^{\Omega(\log n)}$ by Theorem 10, from which $\operatorname{TS}(|p\mathbb{Z}\rangle) = n^{\Omega(\log n)}$ follows by Theorem 5. To lower-bound $\operatorname{TS}_\varepsilon(|p\mathbb{Z}\rangle)$, we use Corollary 11 and the Hoffman-Wielandt inequality, exactly as in Theorem 16. ■

Proof of Theorem 22.

(i) Let $\left|\psi^{(t)}\right\rangle = \sum_{y \in \{0,1\}^{p(n)}} \alpha_y^{(t)} |y\rangle$, where $y = y_1 \ldots y_{p(n)}$, and suppose without loss of generality that $g^{(t+1)}$ is applied to the first k qubits. Let T be a tree representing $\left|\psi^{(t)}\right\rangle$. Then we can write $\alpha_y^{(t)}$ as

$$\sum_{S \subseteq \{1,\ldots,k\}} \gamma_S \left(\prod_{i \in S} y_i \right) \left(\prod_{i \in \{1,\ldots,k\} \setminus S} (1 - y_i) \right) T_{y[k]} \left(y_{k+1}, \ldots, y_{p(n)} \right),$$

where each $T_{y[k]}$ is a restriction of T. This has tree size at most $2^k |T| + 1$. Applying $g^{(t+1)}$ then corresponds simply to changing the coefficients γ_S.

(ii) The Solovay-Kitaev Theorem [28, 37] shows that given a universal gate set, we can approximate any k-qubit unitary to accuracy $1/\varepsilon$ using k qubits and a circuit of size $O(\operatorname{polylog}(1/\varepsilon))$. So let $\left|\psi^{(0)}\right\rangle, \ldots, \left|\psi^{(p(n))}\right\rangle \in \mathcal{H}_2^{\otimes p(n)}$ be a sequence of states, with $\left|\psi^{(t)}\right\rangle$ produced from $\left|\psi^{(t-1)}\right\rangle$ by applying a k-qubit unitary $g^{(t)}$ (where $k = O(1)$). Then using a polynomial-size circuit, we can approximate each $\left|\psi^{(t)}\right\rangle$ to accuracy $1/2^{\Omega(n)}$, as in the definition of $\mathsf{TreeBQP}$. Furthermore, since the approximation circuit for $g^{(t)}$ acts only on k qubits, any intermediate state $|\varphi\rangle$ it produces satisfies $\operatorname{TS}_{1/2^{\Omega(n)}}(|\varphi\rangle) \leq 2^k \operatorname{TS}_{1/2^{\Omega(n)}}\left(\left|\psi^{(t-1)}\right\rangle\right) + 1$ by part (i).

(iii) To amplify to a constant probability, run k copies of the computation in tensor product, then output the majority answer. By part (i), outputting the majority can increase the tree size by a factor of at most 2^{k+1}. To amplify to $2^{-2^{\sqrt{\log n}}}$, observe that the Boolean majority function on k bits has a multilinear formula of size $k^{O(\log k)}$. For let $T_k^h(x_1, \ldots, x_k)$ equal 1 if $x_1 + \cdots + x_k \geq h$ and 0 otherwise; then

$$T_k^h(x_1, \ldots, x_k) = 1 - \prod_{i=0}^{h} \left(1 - T_{\lfloor k/2 \rfloor}^i \left(x_1, \ldots, x_{\lfloor k/2 \rfloor} \right) T_{\lceil k/2 \rceil}^{h-i} \left(x_{\lfloor k/2 \rfloor + 1}, \ldots, x_k \right) \right),$$

so $\operatorname{MFS}(T_k^h) \leq 2h \max_i \operatorname{MFS}\left(T_{\lceil k/2 \rceil}^h\right) + O(1)$, and solving this recurrence yields $\operatorname{MFS}\left(T_k^{k/2}\right) = k^{O(\log k)}$. Substituting $k = 2^{\sqrt{\log n}}$ into $k^{O(\log k)}$ yields $n^{O(1)}$, meaning the tree size increases by at most a polynomial factor.

(iv) To simulate BPP, we just perform a classical reversible computation, applying a Hadamard followed by a measurement to some qubit whenever we need a random bit. Since the number of basis states with nonzero amplitude is at most 2, the simulation is clearly in TreeBQP. The other containment is obvious.

■

Proof of Theorem 23. Since TreeBQP is closed under complement, it suffices to show that $\mathsf{TreeBQP} \subseteq \Pi_3^{\mathsf{P}}$. Our proof will combine approximate counting with a predicate to verify the correctness of a TreeBQP computation. Let C be a uniformly-generated quantum circuit, and let $M = \left(m^{(1)}, \ldots, m^{(p(n))}\right)$ be a sequence of binary measurement outcomes. We adopt the convention that after making a measurement, the state vector is *not* rescaled to have norm 1. That way the probabilities across all 'measurement branches' continue to sum to 1. Let $\left|\psi_{M,x}^{(0)}\right\rangle, \ldots, \left|\psi_{M,x}^{(p(n))}\right\rangle$ be the sequence of unnormalized pure states under measurement outcome sequence M and input x, where $\left|\psi_{M,x}^{(t)}\right\rangle = \sum_{y \in \{0,1\}^{p(n)}} \alpha_{y,M,x}^{(t)} |y\rangle$. Also, let $\Lambda(M,x)$ express that $\mathrm{TS}_{1/2^{\Omega(n)}}\left(\left|\psi_{M,x}^{(t)}\right\rangle\right) \leq p(n)$ for every t. Then C accepts if

$$W_x = \sum_{M\,:\,\Lambda(M,x)} \sum_{y \in \{0,1\}^{p(n)-1}} \left|\alpha_{1y,M,x}^{(p(n))}\right|^2 \geq \frac{2}{3},$$

while C rejects if $W_x \leq 1/3$. If we could compute each $\left|\alpha_{1y,M,x}^{(p(n))}\right|$ efficiently (as well as $\Lambda(M,x)$), we would then have a Π_2^{P} predicate expressing that $W_x \geq 2/3$. This follows since we can do approximate counting via hashing in $\mathsf{AM} \subseteq \Pi_2^{\mathsf{P}}$ [22], and thereby verify that an exponentially large sum of nonnegative terms is at least 2/3, rather than at most 1/3. The one further fact we need is that in our Π_2^{P} ($\forall\exists$) predicate, we can take the existential quantifier to range over tuples of 'candidate solutions'—that is, (M,y) pairs together with lower bounds β on $\left|\alpha_{1y,M,x}^{(p(n))}\right|$.

It remains only to show how we verify that $\Lambda(M,x)$ holds and that $\left|\alpha_{1y,M,x}^{(p(n))}\right| = \beta$. First, we extend the existential quantifier so that it guesses not only M and y, but also a sequence of trees $T^{(0)}, \ldots, T^{(p(n))}$, representing $\left|\psi_{M,x}^{(0)}\right\rangle, \ldots, \left|\psi_{M,x}^{(p(n))}\right\rangle$ respectively. Second, using the last universal quantifier to range over $\widehat{y} \in \{0,1\}^{p(n)}$, we verify the following:

(1) $T^{(0)}$ is a fixed tree representing $|0\rangle^{\otimes(p(n)-n)} \otimes |x\rangle$.

(2) $\left|\alpha_{1y,M,x}^{(p(n))}\right|$ equals its claimed value to $\Omega(n)$ bits of precision.

(3) Let $g^{(1)}, \ldots, g^{(p(n))}$ be the gates applied by C. Then for all t and $\widehat{y}$, if $g^{(t)}$ is unitary then $\alpha_{\widehat{y},M,x}^{(t)} = \langle\widehat{y}| \cdot g^{(t)} \left|\psi_{M,x}^{(t-1)}\right\rangle$ to $\Omega(n)$ bits of precision. Here the right-hand side is a sum of 2^k terms (k being the number of qubits acted on by $g^{(t)}$), each term efficiently computable given $T^{(t-1)}$. Similarly, if $g^{(t)}$ is a measurement of the i^{th} qubit, then $\alpha_{\widehat{y},M,x}^{(t)} = \alpha_{\widehat{y},M,x}^{(t-1)}$ if the i^{th} bit of $\widehat{y}$ equals $m^{(t)}$, while $\alpha_{\widehat{y},M,x}^{(t)} = 0$ otherwise.

■