Multilinear Formulas and Skepticism of Quantum Computing

Scott Aaronson*

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 set of quantum states that can account for all quantum computing experiments performed to date, but *not* for Shor's factoring algorithm. We investigate 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 [34]

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]

*University of California, Berkeley. Email: aaronson@cs.berkeley.edu. Part of this work was done at the Perimeter Institute (Waterloo, Canada). Supported by an NSF Graduate Fellowship, by the Defense Advanced Research Projects Agency (DARPA), and by ARO grant DAAD19-03-1-0082.

[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. Section 4 investigates basic properties of this class of states. Among other results, it shows that any tree state is representable by a tree of polynomial size and logarithmic depth; and that most states do not even have large inner product with any tree state.

Our main results, proved in Section 5, are lower bounds on tree size for various natural families of quantum states. In particular, Section 5.1 analyzes "subgroup states," which are uniform superpositions $|S\rangle$ over all elements of a subgroup $S \leq \mathbb{Z}_2^n$. The importance of these states arises from their central role in stabilizer codes, a type of quantum error-correcting code. We first show that if S is chosen uniformly at random, then with high probability $|S\rangle$ cannot be represented by any tree of size $n^{o(\log n)}$. This result has a corollary of independent complexity-theoretic interest: the first superpolynomial gap between the formula size and the multilinear formula size of a function $f : \{0,1\}^n \to \mathbb{R}$. We then present two improvements of our basic lower bound. First, we show that a random subgroup state cannot even be *approximated* well in trace distance by any tree of size $n^{o(\log n)}$. Second, we "derandomize" the lower bound, by using Reed-Solomon codes to construct an *explicit* subgroup state with tree size $n^{\Omega(\log n)}$.

Section 5.2 analyzes the states that arise in Shor's factoring algorithm—for example, a uniform superposition over all multiples of a fixed positive integer p, written in binary. Originally, we had hoped to show a superpolynomial tree size lower bound for these states as well. However, we are only able to show such a bound assuming a number-theoretic conjecture.

Our lower bounds use a sophisticated recent technique of Raz [40, 41], 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.

One might wonder how superpolynomial tree size relates to more *physical* properties of a quantum state. Section 5.3 addresses this question, by pointing out how Raz's lower bound technique is connected to a notion that physicists call "persistence of entanglement" [13, 18]. On the other hand, we also give examples showing that the connection is not exact.

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? In other words, letting $\mathsf{TreeBQP}$ be the class of languages accepted by such a machine, does $\mathsf{TreeBQP} = \mathsf{BPP}$? 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! Although we leave this question open, we do show that $\mathsf{TreeBQP} \subseteq \Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$, where $\Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$ is the third level of the polynomial hierarchy PH. By contrast, it is conjectured that $\mathsf{BQP} \not\subset \mathsf{PH}$, though admittedly not on strong evidence.

Section 7 discusses the implications of our results for experimental physics. It advocates a dialogue between theory and experiment, in which theorists would propose a class of quantum states that encompasses everything seen so far, and then experimenters would try to prepare states not in that class. It also asks whether states with superpolynomial tree size have already been observed in condensed-matter systems; and more broadly, what sort of evidence is needed to establish a state's existence. Other issues addressed in Section 7 include how to deal with mixed states and particle position and momentum states, and the experimental relevance of asymptotic bounds.

Finally, two appendices investigate quantum state complexity measures other than tree size. Appendix 3 shows relationships among tree size, circuit size, bounded-depth tree size, Vidal's χ complexity [46], and several other measures. It also relates questions about quantum state classes to more traditional questions about computational complexity classes. Appendix 10 studies a weakening of tree size called "manifestly orthogonal tree size," and shows that this measure can sometimes be characterized *exactly*, enabling us to prove *exponential* lower bounds. Our techniques in Appendix 10 might be of independent interest to complexity theorists.

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. Among computer scientists, this objection is most closely associated with Leonid Levin [34].[2] The following passage captures much of the flavor of his critique:

> The major problem [with quantum computing] is the requirement that basic quantum equations hold to multi-hundredth if not millionth decimal positions where the significant digits of the relevant quantum amplitudes reside. 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? [34]

Among other things, Levin argues that quantum computing is analogous to the unit-cost arithmetic model, and should be rejected for essentially the same reasons; that claims to the contrary rest on a confusion between metric and topological approximation; that quantum fault-tolerance theorems depend on extravagant assumptions; and that even if a quantum computer failed, we could not measure its state to prove a breakdown of quantum mechanics, and thus would be unlikely to learn anything new.

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 one dismiss such probabilities as unphysical? 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.

[2]Since this paper was written, Oded Goldreich [25] has also put forward an argument against quantum computing. Compared to Levin's arguments, Goldreich's is easily understood: he believes that Shor states have exponential "non-degeneracy" and therefore take exponential time to prepare, and that there is no burden on those who hold this view to suggest a definition of non-degeneracy.

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 [4] argue that a wide class of nonlinear variants of the Schrödinger equation would allow NP-complete and even #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 set of computable functions entirely!

Third, to falsify quantum mechanics, it would suffice to show 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. An example is the physics Nobel awarded to Fitch [19] and Cronin [17] in 1980 for discovering CP symmetry violation.

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 the predictions of quantum mechanics "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 would disintegrate. That this has not happened—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 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. [20] 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 (see Leggett [33] for a survey of such experiments). 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 computer 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 [23].[3] 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 the collapses postulated by the theory are only in the position basis, so that quantum information stored in internal degrees of freedom (such as spin) is unaffected. Furthermore, even if we extended the theory to collapse those internal degrees, large quantum computers could still be built. For the theory predicts roughly one collapse per particle per 10^{15} seconds, with a collapse affecting everything in a 10^{-7}-meter vicinity. So even if we confined ourselves to such a vicinity, we could perform a computation involving (say) 10^{10} particles for 10^5 seconds. Finally, as pointed out to us by Rob Spekkens, standard quantum error-correction techniques might be used to overcome GRW-type decoherence.

A second class of alternatives includes those of 't Hooft [43] and Wolfram [48], 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" [43]. 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" [48, 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.

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, yet would not allow large-scale quantum computation. Failing that, *at least say what you take quantum mechanics' domain of validity to be.* One way to do this would be to propose a 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, or else does not even make sense in the context of an alternative theory. The states that arise in Shor's factoring algorithm have exponential values of $C(|\psi_n\rangle)$.*

[3] Penrose [38] has proposed another such theory, but as mentioned earlier, his theory suggests that the quantum computing model is *too* restrictive.

[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.

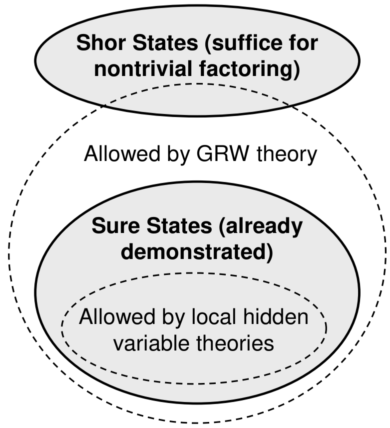

Figure 1: A Sure/Shor separator must contain all Sure states but no Shor states. That is why neither local hidden variables nor the GRW theory yields a Sure/Shor separator.

So as my Sure/Shor separator, I propose the set of all infinite families of states $\{|\psi_n\rangle\}_{n\geq1}$*, where* $|\psi_n\rangle$ *has* n *qubits, such that* $C(|\psi_n\rangle) \leq p(n)$ *for some polynomial* 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. 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.[5] So considering a Shor state such as

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

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 state in the Hadamard basis produces an equal superposition over all strings of even parity, 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 led us to propose *tree states* as a possible Sure/Shor separator. The idea, which might seem more natural to logicians than to physicists, is this. 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

[5] A reviewer comments that in Chern-Simons theory (for example), there is no clear tensor product decomposition. However, the only question that concerns us is whether $|\psi\rangle \otimes |\varphi\rangle$ is a Sure state, *given* that $|\psi\rangle$ and $|\varphi\rangle$ are both Sure states that are well-described in tensor product Hilbert spaces.

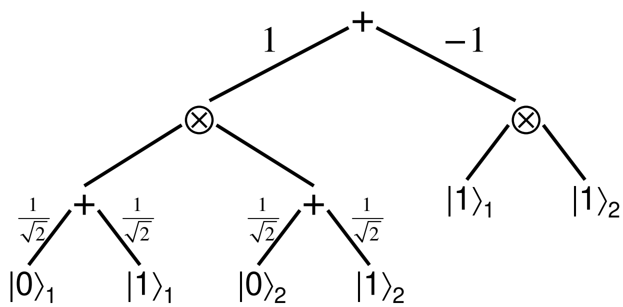

Figure 2: Expressing $\left(|00\rangle + |01\rangle + |10\rangle - |11\rangle\right)/2$ by a tree of linear combination and tensor product gates, with scalar multiplication along edges. Subscripts denote the identity of a qubit.

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).

To be clear, we are *not* advocating that "all states in Nature are tree states" as a serious physical hypothesis. Indeed, even if we believed firmly in a breakdown of quantum mechanics,[6] there are other choices for the set S that seem equally reasonable. For example, define *orthogonal tree states* similarly to tree states, except that we can only form the linear combination $\alpha|\psi\rangle + \beta|\varphi\rangle$ if $\langle\psi|\varphi\rangle = 0$. Rather than choose among tree states, orthogonal tree states, and the other candidate Sure/Shor separators that occurred to us, our approach will be to prove everything we can about all of them. If we devote more space to tree states than to others, that is simply because tree states are the subject of our most interesting results. On the other hand, if we show (for example) that $\{|\psi_n\rangle\}$ is not a tree state, then we have also shown that $\{|\psi_n\rangle\}$ is not an orthogonal tree state. So many candidate separators are related to each other; and indeed, their relationships will be a major theme of the paper.

Let us summarize. To debate whether quantum computing is fundamentally impossible, we need at least one proposal for how it *could* be impossible. Since even skeptics admit that quantum mechanics is valid within some "regime," a key challenge for any such proposal is to separate the regime of acknowledged validity from the quantum computing regime. Though others will disagree, we do not see any choice but to *identify those two regimes with classes of quantum states.* For gates and measurements that suffice for quantum computing have already been demonstrated experimentally. Thus, if we tried to identify the two regimes with classes of gates or measurements, then we could equally well talk about the class of *states* on which all 1- and 2-qubit operations behave as expected. A similar argument would apply if we identified the two regimes with classes of quantum circuits—since any "memory" that a quantum system retains of the previous gates in a circuit, is part of the system's state by definition. So: states, gates, measurements, circuits—what else is there?

We should stress that none of the above depends on the interpretation of quantum mechanics. In particular, it is irrelevant whether we regard quantum states as "really out there" or as representing subjective knowledge—since in either case, the question is whether there can exist systems that we would *describe* by $|\psi\rangle$ based on their observed behavior.

[6] which we don't

Once we agree to seek a Sure/Shor separator, we quickly find that the obvious ideas—based on precision in amplitudes, or entanglement across of hundreds of particles—are nonstarters. The only idea that we have found plausible is to limit the class of allowed quantum states to those with some kind of succinct representation. That still leaves numerous possibilities; and for each one, it might be a difficult problem to decide whether a given $|\psi\rangle$ is succinctly representable or not. Thus, constructing a useful theory of Sure/Shor separators will not be easy. But we should start somewhere.

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 [30]; Vidal [46]; and Green et al. [28]. Here we propose a general 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* $\mathrm{TS}(|\psi_n\rangle)$ is defined as follows.

Definition 1 *A quantum state tree over $\mathcal{H}_2^{\otimes n}$ is a rooted tree where each leaf vertex is labeled with $\alpha|0\rangle + \beta|1\rangle$ for some $\alpha, \beta \in \mathsf{C}$, 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 of v are labeled with complex numbers. For each 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.[7]

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

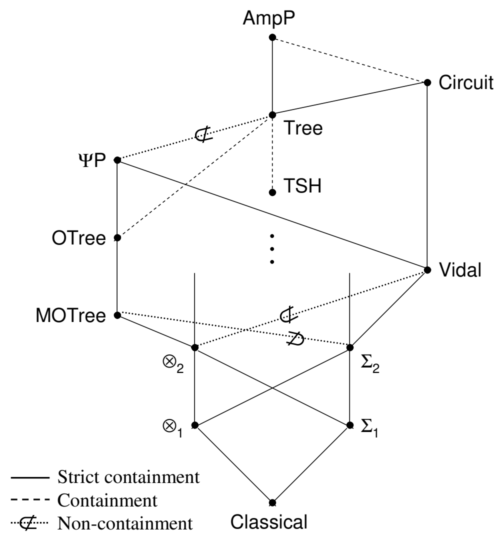

Figure 3: Relations among quantum state classes.

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*. Manifest orthogonality is an extremely unphysical definition; we introduce it only because it is interesting from a lower bounds perspective.

For reasons of convenience, we define the *size* $|T|$ of a tree T to be the number of leaf vertices. 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 OTree is the class of $|\psi_n\rangle$ such that $\mathrm{OTS}(|\psi_n\rangle) \leq p(n)$ for some polynomial p, and MOTree is the class such that $\mathrm{MOTS}(|\psi_n\rangle) \leq p(n)$ for some p.

It is easy to see that

$$n \leq \mathrm{TS}(|\psi\rangle) \leq \mathrm{OTS}(|\psi\rangle) \leq \mathrm{MOTS}(|\psi\rangle) \leq n2^n$$

for every $|\psi\rangle$, and that the set of $|\psi\rangle$ such that $\mathrm{TS}(|\psi\rangle) < 2^n$ has measure 0 in $\mathcal{H}_2^{\otimes n}$. Two other important properties of TS and OTS are as follows:

Proposition 2

(i) TS *and* OTS *are invariant under local*[8] *basis changes, up to a constant factor of* 2.

(ii) If $|\phi\rangle$ *is obtained from* $|\psi\rangle$ *by applying a* k*-qubit unitary, then* $\mathrm{TS}(|\phi\rangle) \leq k4^k\, \mathrm{TS}(|\psi\rangle)$ *and* $\mathrm{OTS}(|\phi\rangle) \leq k4^k\, \mathrm{OTS}(|\psi\rangle)$.

Proof.

(i) Simply replace each occurrence of $|0\rangle$ in the original tree by a tree for $\alpha|0\rangle + \beta|1\rangle$, and each occurrence of $|1\rangle$ by a tree for $\gamma|0\rangle + \delta|1\rangle$, as appropriate.

[8]Several people told us that a reasonable complexity measure must be invariant under *all* basis changes. Alas, this would imply that all pure states have the same complexity!

(ii) Suppose without loss of generality that the gate is applied to the first k qubits. Let T be a tree representing $|\psi\rangle$, and let T_y be the restriction of T obtained by setting the first k qubits to $y \in \{0,1\}^k$. Clearly $|T_y| \leq |T|$. Furthermore, we can express $|\phi\rangle$ in the form $\sum_{y\in\{0,1\}^k} S_y T_y$, where each S_y represents a k-qubit state and hence is expressible by a tree of size $k2^k$.

■

We can also define the *ε-approximate tree size* $\mathrm{TS}_\varepsilon(|\psi\rangle)$ 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 $\mathrm{OTS}_\varepsilon(|\psi\rangle)$ and $\mathrm{MOTS}_\varepsilon(|\psi\rangle)$ similarly.

Definition 3 *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 leaf vertices. Given a multilinear polynomial p, the multilinear formula size $\mathrm{MFS}(p)$ is the minimum size of a multilinear formula that represents p. Then given a function $f : \{0,1\}^n \rightarrow \mathbb{C}$, we define

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

(Actually p turns out to be unique [37].) We can also define the ε-approximate multilinear formula size of f,

$$\mathrm{MFS}_\varepsilon(f) = \min_{p\,:\,\|p-f\|_2^2\leq\varepsilon} \mathrm{MFS}(p)$$

where $\|p-f\|_2^2 = \sum_{x\in\{0,1\}^n} |p(x)-f(x)|^2$. (This metric is closely related to the inner product $\sum_x p(x)^* f(x)$, but is often more convenient to work with.) 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 4 *For all $|\psi\rangle$,*

(i) $\mathrm{MFS}(f_\psi) = O(\mathrm{TS}(|\psi\rangle))$.

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

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

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

Proof.

(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 [40] 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

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

■

Besides Tree, OTree, and MOTree, four other classes of quantum states deserve mention:

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 arithmetic circuit of size $p(n)$ over the complex numbers that outputs α_x given x as input, for some polynomial p. (Multilinear circuits are the same as multilinear trees, except that they allow unbounded fanout—that is, polynomials computed at intermediate points can be reused arbitrarily many times.)

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 [46]. Given a partition 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 $\sum_{i=1}^k \alpha_i \left|\varphi_i^A\right\rangle \otimes \left|\varphi_i^B\right\rangle$, where $\left|\varphi_i^A\right\rangle$ and $\left|\varphi_i^B\right\rangle$ are states of qubits in A and B respectively. ($\chi_A(|\psi_n\rangle)$ is known as the *Schmidt rank*; see [36] for more information.) Let $\chi(|\psi_n\rangle) = \max_A \chi_A(|\psi_n\rangle)$. Then $|\psi_n\rangle \in$ Vidal if and only if $\chi(|\psi_n\rangle) \leq p(n)$ for some polynomial p.

ΨP 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 all-0 state to a state some part of which has trace distance at most $1 - \varepsilon$ from $|\psi_n\rangle$, for some polynomial p. Because of the Solovay-Kitaev Theorem [31, 36], ΨP is invariant under the choice of universal gate set.

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 prove three rather nice properties of tree size.

Theorem 5 *For all $\varepsilon > 0$, there exists a tree representing $|\psi\rangle$ of size $O\left(\mathrm{TS}(|\psi\rangle)^{1+\varepsilon}\right)$ and depth $O(\log \mathrm{TS}(|\psi\rangle))$, as well as a manifestly orthogonal tree of size $O\left(\mathrm{MOTS}(|\psi\rangle)^{1+\varepsilon}\right)$ and depth $O(\log \mathrm{MOTS}(|\psi\rangle))$.*

Proof. 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 [14] (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 $v_0 = \mathrm{root}(I)$, let v_1 be the parent of v_0, let v_2 be the parent of v_1, and so on until $v_k = \mathrm{root}(\Phi)$. 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 $\mathrm{supp}(p)$ 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

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

and manifest orthogonality follows. ■

Theorem 6 *Any $|\psi\rangle$ can be prepared by a quantum circuit of size polynomial in $\mathrm{OTS}(|\psi\rangle)$. Thus $\mathsf{OTree} \subseteq \Psi\mathsf{P}$.*

Proof. 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 $\Gamma(|\psi\rangle) \leq q(\mathrm{OTS}(|\psi\rangle))$ for some polynomial q. The base case $\mathrm{OTS}(|\psi\rangle) = 1$ 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 $|T| = |T_1| + |T_2|$. 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 $\operatorname{OTS}(|\psi\rangle)$.
■

Theorem 7 *If $|\psi\rangle \in \mathcal{H}_2^{\otimes n}$ is chosen uniformly at random under the Haar measure, then* $\operatorname{TS}_{1/16}(|\psi\rangle) = 2^{\Omega(n)}$ *with probability* $1 - o(1)$.

Proof. 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 $\operatorname{MFS}_{1/15}(f_\psi) = 2^{\Omega(n)}$ 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 [22] (see also Turán and Vatan [44]), which follows from Warren's Theorem [47] in real algebraic geometry, shows that as we range over all c, the formula 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 4, part (iii), this implies that $\mathrm{TS}_{1/16}(|\psi\rangle) = 2^{\Omega(n)}$. ∎

A corollary of Theorem 7 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 8 *For all $\delta \in (0,1]$, there exists a state $|\psi\rangle$ such that $\mathrm{TS}_\delta(|\psi\rangle) = n$ but $\mathrm{TS}_\varepsilon(|\psi\rangle) = 2^{\Omega(n)}$ where $\varepsilon = \delta/32 - \delta^2/4096$.*

Proof. It is clear from Theorem 7 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 $\mathrm{MOTS}_\delta(|\psi\rangle) = n$. 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 $\mathrm{MFS}_c(f_\varphi) = O(\mathrm{TS}(|\phi\rangle))$ where $c = \left(2 - 2\sqrt{1-\varepsilon}\right)/\delta$. By Theorem 4, part (iv), this implies that $\mathrm{TS}_{2c}(|\varphi\rangle) = O(\mathrm{TS}(|\phi\rangle))$. But $2c = 1/16$ when $\varepsilon = \delta/32 - \delta^2/4096$, contradiction. ∎

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 notoriously hard open problem, as it would imply complexity class separations such as $\mathsf{NC}^1 \neq \mathsf{P}$.

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 [40, 41] 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 $n^{\Omega(\log n)}$.

Raz's technique is a beautiful combination of the Furst-Saxe-Sipser method of random restrictions [21], 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 P be a partition of the input variables $x_1, \ldots, x_n$ into two collections $y = (y_1, \ldots, y_{n/2})$ and $z = (z_1, \ldots, z_{n/2})$. This yields a function $f_P(y,z) : \{0,1\}^{n/2} \times \{0,1\}^{n/2} \rightarrow \mathbb{C}$. Then let $M_{f|P}$ be a $2^{n/2} \times 2^{n/2}$ matrix whose rows are labeled by assignments $y \in \{0,1\}^{n/2}$, and whose columns are labeled by assignments $z \in \{0,1\}^{n/2}$. The (y,z) entry of $M_{f|P}$ is $f_P(y,z)$. Let $\operatorname{rank}(M_{f|P})$ be the rank of $M_{f|P}$ over the complex numbers. Finally, let $\mathcal{P}$ be the uniform distribution over all partitions P.

The following, Corollary 3.6 in [41], is one statement of Raz's main theorem; recall that $\mathrm{MFS}(f)$ is the minimum size of a multilinear formula for f.

Theorem 9 ([41]) *Suppose that*

$$\Pr_{P\in\mathcal{P}}\left[\operatorname{rank}\left(M_{f|P}\right) \geq 2^{n/2-(n/2)^{1/8}/2}\right] = n^{-o(\log n)}.$$

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

An immediate corollary yields lower bounds on *approximate* multilinear formula size. Given an $N \times N$ matrix $M = (m_{ij})$, let $\operatorname{rank}_\varepsilon(M) = \min_{L \,:\, \|L-M\|_2^2 \leq \varepsilon} \operatorname{rank}(L)$ where $\|L-M\|_2^2 = \sum_{i,j=1}^{N} |l_{ij} - m_{ij}|^2$.

Corollary 10 *Suppose that*

$$\Pr_{P\in\mathcal{P}}\left[\operatorname{rank}_\varepsilon\left(M_{f|P}\right) \geq 2^{n/2-(n/2)^{1/8}/2}\right] = n^{-o(\log n)}.$$

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

Proof. Suppose $\mathrm{MFS}_\varepsilon(f) = n^{o(\log n)}$. Then for all g such that $\|f-g\|_2^2 \leq \varepsilon$, we would have $\mathrm{MFS}(g) = n^{o(\log n)}$, and therefore

$$\Pr_{P\in\mathcal{P}}\left[\operatorname{rank}\left(M_{g|P}\right) \geq 2^{n/2-(n/2)^{1/8}/2}\right] = n^{-\Omega(\log n)}.$$

by Theorem 9. But $\operatorname{rank}_\varepsilon\left(M_{f|P}\right) \leq \operatorname{rank}\left(M_{g|P}\right)$, and hence

$$\Pr_{P\in\mathcal{P}}\left[\operatorname{rank}_\varepsilon\left(M_{f|P}\right) \geq 2^{n/2-(n/2)^{1/8}/2}\right] = n^{-\Omega(\log n)},$$

contradiction. ∎

Another simple corollary gives lower bounds in terms of *restrictions* of f. Let $\mathcal{R}_l$ be the following distribution over restrictions R: choose $2l$ variables of f uniformly at random, and rename them $y = (y_1, \ldots, y_l)$ and $z = (z_1, \ldots, z_l)$. Set each of the remaining $n - 2l$ variables to 0 or 1 uniformly and independently at random. This yields a restricted function $f_R(y,z)$. Let $M_{f|R}$ be a $2^l \times 2^l$ matrix whose (y,z) entry is $f_R(y,z)$.

Corollary 11 *Suppose that*

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

where $l = n^\delta$ for some constant $\delta \in (0,1]$. Then $\mathrm{MFS}(f) = n^{\Omega(\log n)}$.

Proof. Under the hypothesis, clearly there exists a *fixed* restriction $g : \{0,1\}^{2l} \to \mathbb{C}$ of f, which leaves $2l$ variables unrestricted, such that

$$\Pr_{P \in \mathcal{P}}\left[\operatorname{rank}\left(M_{g|P}\right) \geq 2^{l-l^{1/8}/2}\right] = n^{-o(\log n)} = l^{-o(\log l)}.$$

Then by Theorem 9,

$$\mathrm{MFS}(f) \geq \mathrm{MFS}(g) = l^{\Omega(\log l)} = n^{\Omega(\log n)}.$$

■

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.

5.1 Subgroup States

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

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

Coset states arise as codewords in the class of quantum error-correcting codes known as stabilizer codes [16, 27, 42]. Our interest in these states, however, arises from their large tree size rather than their error-correcting properties.

Let $\mathcal{E}$ be the following distribution over subgroups S. Choose an $n/2 \times n$ matrix A by setting each entry to 0 or 1 uniformly and independently. Then let $S = \{x \mid Ax \equiv 0 \,(\mathrm{mod}\, 2)\}$. By Theorem 4, part (i), it suffices to lower-bound the multilinear formula size of the function $f_S(x)$, which is 1 if $x \in S$ and 0 otherwise.

Theorem 12 *If S is drawn from $\mathcal{E}$, then $\mathrm{MFS}(f_S) = n^{\Omega(\log n)}$ (and hence $\mathrm{TS}(|S\rangle) = n^{\Omega(\log n)}$), with probability $\Omega(1)$ over S.*

Proof. Let P be a uniform random partition of the inputs $x_1, \ldots, x_n$ of f_S into two sets $y = (y_1, \ldots, y_{n/2})$ and $z = (z_1, \ldots, z_{n/2})$. Let $M_{S|P}$ be the $2^{n/2} \times 2^{n/2}$ matrix whose (y,z) entry is $f_{S|P}(y,z)$; then we need to show that $\operatorname{rank}(M_{S|P})$ is large with high probability. Let A_y be the $n/2 \times n/2$ submatrix of the $n/2 \times n$ matrix A consisting of all rows that correspond to y_i for some $i \in \{1, \ldots, n/2\}$, and similarly let A_z be the $n/2 \times n/2$ submatrix corresponding to z. Then it is easy to see that, so long as A_y and A_z are both invertible, for all $2^{n/2}$ settings of y there exists a *unique* setting of z for which $f_{S|P}(y,z) = 1$. This then implies that $M_{S|P}$ is a permutation of the identity matrix, and hence that $\operatorname{rank}(M_{S|P}) = 2^{n/2}$. Now, the probability that a random $n/2 \times n/2$ matrix over $\mathbb{Z}_2$ is invertible is

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

So the probability that A_y and A_z are both invertible is at least 0.288^2. By Markov's inequality, it follows that for at least an 0.04 fraction of S's, $\operatorname{rank}(M_{S|P}) = 2^{n/2}$ for at least an 0.04 fraction of P's. Theorem 9 then yields the desired result. ■

Aaronson and Gottesman [3] show how to prepare any n-qubit subgroup state using a quantum circuit of size $O(n^2/\log n)$. So a corollary of Theorem 12 is that $\Psi\mathsf{P} \not\subset \mathsf{Tree}$. Since f_S clearly has a (non-multilinear) arithmetic formula of size $O(nk)$, a second corollary is the following.

Corollary 13 *There exists a family of functions $f_n : \{0,1\}^n \to \mathbb{R}$ that has polynomial-size arithmetic formulas, but no polynomial-size multilinear formulas.*

The reason Corollary 13 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 [15]).

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

Lemma 14 *Let M be an $N \times N$ complex matrix, and let I_N be the $N \times N$ identity matrix. Then $\|M - I_N\|_2^2 \geq N - \operatorname{rank}(M)$.*

Proof. The Hoffman-Wielandt inequality [29] (see also [6]) states that for any two $N \times N$ matrices M, P,

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

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

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

■

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

Theorem 15 *For all constants $\varepsilon \in [0,1)$, if S is drawn from $\mathcal{E}$, then $\mathrm{MFS}_\varepsilon\left(\widehat{f}_S\right) = n^{\Omega(\log n)}$ with probability $\Omega(1)$ over S.*

Proof. As in Theorem 12, we look at the matrix $M_{S|P}$ induced by a random partition $P = (y, z)$. We already know that for at least an 0.04 fraction of S's, the y and z variables are in one-to-one correspondence for at least an 0.04 fraction of P's. In that case $|S| = 2^{n/2}$, and therefore $M_{S|P}$ is a permutation of $I/\sqrt{|S|} = I/2^{n/4}$ where I is the identity. It follows from Lemma 14 that for all matrices M such that $\left\|M - M_{S|P}\right\|_2^2 \leq \varepsilon$,

$$\operatorname{rank}(M) \geq 2^{n/2} - \left\|\sqrt{|S|}\left(M - M_{S|P}\right)\right\|_2^2 \geq (1-\varepsilon)\, 2^{n/2}$$

and therefore $\operatorname{rank}_\varepsilon\left(M_{S|P}\right) \geq (1-\varepsilon)\, 2^{n/2}$. Hence

$$\Pr_{P \in \mathcal{P}}\left[\operatorname{rank}_\varepsilon\left(M_{f|P}\right) \geq 2^{n/2 - (n/2)^{1/8}/2}\right] \geq 0.04,$$

and the result follows from Corollary 10. ■

A corollary of Theorem 15 and of Theorem 4, part (iii), is that $\mathrm{TS}_\varepsilon(|S\rangle) = n^{\Omega(\log n)}$ with probability $\Omega(1)$ over S, for all $\varepsilon < 1$.

Finally, let us show how to derandomize the lower bound for subgroup states, using ideas pointed out to us by Andrej Bogdanov. In the proof of Theorem 12, all we used about the matrix A was that a random $k \times k$ submatrix has full rank with $\Omega(1)$ probability, where $k = n/2$. If we switch from the field $\mathbb{F}_2$ to $\mathbb{F}_{2^d}$ for some $d \geq \log_2 n$, then it is easy to construct explicit $k \times n$ matrices with this same property. For example, let

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

be the $n \times k$ Vandermonde matrix, where $1, \ldots, n$ are labels of elements in $\mathbb{F}_{2^d}$. *Any* $k \times k$ submatrix of V has full rank, because the Reed-Solomon (RS) code that V represents is a perfect erasure code.[9] Hence, there exists an explicit state of n "qupits" with $p = 2^d$ that has tree size $n^{\Omega(\log n)}$—namely the uniform superposition over all elements of the set $\{x \mid V^T x = 0\}$, where V^T is the transpose of V.

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

$$V_{\text{bin}} = \begin{pmatrix} Hm(1^0) & Hm(1^1) & \cdots & Hm(1^{k-1}) \\ Hm(2^0) & Hm(2^1) & \cdots & Hm(2^{k-1}) \\ \vdots & \vdots & & \vdots \\ Hm(n^0) & Hm(n^1) & \cdots & Hm(n^{k-1}) \end{pmatrix}.$$

For the remainder of the section, fix $k = n^\delta$ for some $\delta < 1/2$ and $d = O(\log n)$.

Lemma 16 *A $(kd + c) \times kd$ submatrix of V_{bin} chosen uniformly at random has rank kd (that is, full rank) with probability at least $2/3$, for c a sufficiently large constant.*

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

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

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

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

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

[9] In other words, because a degree-$(k-1)$ polynomial is determined by its values at any k points.

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

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

Theorem 17 *$\text{MFS}(f) = n^{\Omega(\log n)}$.*

Proof. Let V_y and V_z be two disjoint $kd \times (kd + c)$ submatrices of V_{bin}^T chosen uniformly at random. Then by Lemma 16 together with the union bound, V_y and V_z both have full rank with probability at least $1/3$. Letting $l = kd + c$, it follows that

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

by the same reasoning as in Theorem 12. Therefore $\text{MFS}(f) = n^{\Omega(\log n)}$ by Corollary 11. ■

Let $|S\rangle$ be a uniform superposition over all x such that $f(x) = 1$; then a corollary of Theorem 17 is that $\text{TS}(|S\rangle) = n^{\Omega(\log n)}$. Naturally, using the ideas of Theorem 15 one can also show that $\text{TS}_\varepsilon(|S\rangle) = n^{\Omega(\log n)}$ for all $\varepsilon < 1$.

As a final observation, it is not a coincidence that the states for which we succeeded in proving tree size lower bounds are the sort of states that arise in quantum error correction. For to show using Raz's method that $\text{MFS}(f) = n^{\Omega(\log n)}$ for various functions $f : \{0,1\}^n \to \{0,1\}$, we have basically argued that, if we randomly restrict most of f's inputs to 0 or 1, then with high probability there exists a unique (or nearly unique) input x^* consistent with the restriction such that $f(x^*) = 1$. This is close to saying that the set $S = \{x : f(x) = 1\}$ constitutes a binary erasure code with good parameters. For the requirement that x^* exist corresponds to $|S|$ being large (or equivalently, to the codeword length being small), and the requirement that x^* be nearly unique corresponds to decoding being nearly unique. There are admittedly some technical problems with making this correspondence more precise, which we will not consider here. Also, there is no theorem of the form "f corresponds to a good erasure code if and only if $\text{MFS}(f)$ is large." For a uniform random f has *exponential* multilinear formula size with overwhelming probability (by the proof of Theorem 7), yet does not correspond to any good erasure code.

5.2 Shor States

Since the motivation for our theory was to study possible Sure/Shor separators, an obvious question is, *do states arising in Shor's algorithm have superpolynomial tree size?* Unfortunately, we are only able to answer this question assuming a number-theoretic conjecture. To formalize the question, let

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

be a Shor state. It will be convenient for us to measure the second register, so that the state of the first register has the form

$$|a + p\mathbb{Z}\rangle = \frac{1}{\sqrt{I}} \sum_{i=0}^{I} |a + pi\rangle$$

for some integers $a < p$ and $I = \lfloor (2^n - a - 1)/p \rfloor$. Here $a + pi$ is written out in binary using n bits. Clearly a lower bound on $\text{TS}(|a + p\mathbb{Z}\rangle)$ would imply an equivalent lower bound for the joint state of the two registers. Also, to avoid some technicalities we assume p is prime. Since our goal is to prove a *lower* bound, this assumption is without loss of generality.

Given an n-bit string $x = x_{n-1} \ldots x_0$, let $f_{n,p,a}(x) = 1$ if $x \equiv a \pmod p$ and $f_{n,p,a}(x) = 0$ otherwise. Then $\text{TS}(|a + p\mathbb{Z}\rangle) = \Theta(\text{MFS}(f_{n,p,a}))$ by Theorem 4, so from now on we will focus attention on $f_{n,p,a}$.

Proposition 18

(i) Let $f_{n,p} = f_{n,p,0}$. *Then* $\text{MFS}(f_{n,p,a}) \leq \text{MFS}(f_{n+\log p,p})$, *meaning that we can set* $a = 0$ *without loss of generality.*

(ii) $\text{MFS}(f_{n,p}) = O(\min\{n2^n/p, np\})$.

Proof.

(i) Take the formula for $f_{n+\log p,p}$, and restrict the most significant $\log p$ bits to sum to a number congruent to $-a \bmod p$ (this is always possible since $x \to 2^n x$ is an isomorphism of $\mathbb{Z}_p$).

(ii) For $\text{MFS}(f_{n,p}) = O(n2^n/p)$, write out the x's for which $f_{n,p}(x) = 1$ explicitly. For $\text{MFS}(f_{n,p}) = O(np)$, use the Fourier transform, similarly to Theorem 25, part (v):

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

This immediately yields a sum-of-products formula of size $O(np)$.

■

We now state our number-theoretic conjecture.

Conjecture 19 *There exist constants* $\gamma, \delta \in (0,1)$ *and a prime* $p = \Omega\left(2^{n^\delta}\right)$ *for which the following holds. Let the set* A *consist of* n^δ *elements of* $\left\{2^0, \ldots, 2^{n-1}\right\}$ *chosen uniformly at random. Let* S *consist of all* 2^{n^δ} *sums of subsets of* A, *and let* $S \bmod p = \{x \bmod p : x \in S\}$. *Then*

$$\Pr_A\left[|S \bmod p| \geq (1+\gamma)\frac{p}{2}\right] = n^{-o(\log n)}.$$

Theorem 20 *Conjecture 19 implies that* $\text{MFS}(f_{n,p}) = n^{\Omega(\log n)}$ *and hence* $\text{TS}(|p\mathbb{Z}\rangle) = n^{\Omega(\log n)}$.

Proof. Let $f = f_{n,p}$ and $l = n^\delta$. Let R be a restriction of f that renames $2l$ variables $y_1, \ldots, y_l, z_1, \ldots, z_l$, and sets each of the remaining $n - 2l$ variables to 0 or 1. This leads to a new function, $f_R(y,z)$, which is 1 if $y+z+c \equiv 0 \pmod{p}$ and 0 otherwise for some constant c. Here we are defining $y = 2^{a_1}y_1 + \cdots + 2^{a_l}y_l$ and $z = 2^{b_1}z_1 + \cdots + 2^{b_l}z_l$ where $a_1, \ldots, a_l, b_1, \ldots, b_l$ are the appropriate place values. Now suppose $y \bmod p$ and $z \bmod p$ both assume at least $(1+\gamma)p/2$ distinct values as we range over all $x \in \{0,1\}^n$. Then by the pigeonhole principle, for at least γp possible values of $y \bmod p$, there exists a unique possible value of $z \bmod p$ for which $y + z + c \equiv 0 \pmod{p}$ and hence $f_R(y,z) = 1$. So $\text{rank}\left(M_{f|R}\right) \geq \gamma p$, where $M_{f|R}$ is the $2^l \times 2^l$ matrix whose (y,z) entry is $f_R(y,z)$. It follows that assuming Conjecture 19,

$$\Pr_{R \in \mathcal{R}_l}\left[\text{rank}\left(M_{f|R}\right) \geq \gamma p\right] = n^{-o(\log n)}.$$

Furthermore, $\gamma p \geq 2^{l - l^{1/8}/2}$ for sufficiently large n since $p = \Omega\left(2^{n^\delta}\right)$. Therefore $\text{MFS}(f) = n^{\Omega(\log n)}$ by Corollary 11. ■

Using the ideas of Theorem 15, one can show that under the same conjecture, $\text{MFS}_\varepsilon(f_{n,p}) = n^{\Omega(\log n)}$ and $\text{TS}_\varepsilon(|p\mathbb{Z}\rangle) = n^{\Omega(\log n)}$ for all $\varepsilon < 1$—in other words, there exist Shor states that cannot be approximated by polynomial-size trees.

In an earlier version of this paper, Conjecture 19 was stated without any restriction on how the set S is formed. The resulting conjecture was far more general than we needed, and indeed was falsified by Carl Pomerance (personal communication).

5.3 Tree Size and Persistence of Entanglement

In this section we pursue a deeper understanding of our lower bounds, by discussing a *physical* property of quantum states that is related to error-correction as well as superpolynomial tree size. Dür and Briegel [18] among others call a state "persistently entangled," if (roughly speaking) it remains highly entangled even after a limited amount of interaction with the environment. As an example, the Schrödinger cat state $\left(|0\rangle^{\otimes n}+|1\rangle^{\otimes n}\right)/\sqrt{2}$ is in some sense highly entangled, but it is *not* persistently entangled, since measuring a single qubit in the standard basis destroys all entanglement.

By contrast, consider the "cluster states" defined by Briegel and Raussendorf [13]. These states have attracted a great deal of attention because of their application to quantum computing via 1-qubit measurements only [39]. For our purposes, a two-dimensional cluster state is an equal superposition over all settings of a $\sqrt{n}\times\sqrt{n}$ array of bits, with each basis state having a phase of $(-1)^r$, where r is the number of horizontally or vertically adjacent pairs of bits that are both '1'. Dür and Briegel [18] showed that such states are persistently entangled in a precise sense: one can distill n-partite entanglement from them even after each qubit has interacted with a heat bath for an amount of time independent of n.

Persistence of entanglement seems related to how one shows tree size lower bounds using Raz's technique. For to apply Corollary 11, one basically "measures" most of a state's qubits, then partitions the unmeasured qubits into two subsystems of equal size, and argues that with high probability those two subsystems are still almost maximally entangled. The connection is not perfect, though. For one thing, setting most of the qubits to 0 or 1 uniformly at random is not the same as measuring them. For another, Theorem 9 yields $n^{\Omega(\log n)}$ tree size lower bounds without the need to trace out a subset of qubits. It suffices for the *original* state to be almost maximally entangled, no matter how one partitions it into two subsystems of equal size.

But what about 2-D cluster states—do *they* have tree size $n^{\Omega(\log n)}$? We strongly conjecture that the answer is 'yes.' However, proving this conjecture will almost certainly require going beyond Theorem 9. One will want to use random restrictions that respect the 2-D neighborhood structure of cluster states—similar to the restrictions used by Raz [40] to show that the permanent and determinant have multilinear formula size $n^{\Omega(\log n)}$.

We end this section by showing that there exist states that are persistently entangled in the sense of Dür and Briegel [18], but that have polynomial tree size. In particular, Dür and Briegel showed that even *one*-dimensional cluster states are persistently entangled. On the other hand:

Proposition 21 *Let*

$$|\psi\rangle=\frac{1}{2^{n/2}}\sum_{x\in\{0,1\}^n}(-1)^{x_1x_2+x_2x_3+\cdots+x_{n-1}x_n}|x\rangle .$$

Then $\mathrm{TS}(|\psi\rangle)=O\left(n^4\right)$.

Proof. Given bits i,j,k, let $\left|P_n^{ijk}\right\rangle$ be an equal superposition over all n-bit strings $x_1\ldots x_n$ such that $x_1=i$, $x_n=k$, and $x_1x_2+\cdots+x_{n-1}x_n\equiv j\,(\mathrm{mod}\,2)$. Then

$$\left|P_n^{i0k}\right\rangle=\frac{1}{\sqrt{8}}\left(\begin{array}{c}\left|P_{n/2}^{i00}\right\rangle\left|P_{n/2}^{00k}\right\rangle+\left|P_{n/2}^{i10}\right\rangle\left|P_{n/2}^{01k}\right\rangle+\left|P_{n/2}^{i00}\right\rangle\left|P_{n/2}^{10k}\right\rangle+\left|P_{n/2}^{i10}\right\rangle\left|P_{n/2}^{11k}\right\rangle+\\ \left|P_{n/2}^{i01}\right\rangle\left|P_{n/2}^{00k}\right\rangle+\left|P_{n/2}^{i11}\right\rangle\left|P_{n/2}^{01k}\right\rangle+\left|P_{n/2}^{i01}\right\rangle\left|P_{n/2}^{11k}\right\rangle+\left|P_{n/2}^{i11}\right\rangle\left|P_{n/2}^{10k}\right\rangle\end{array}\right),$$

$$\left|P_n^{i1k}\right\rangle=\frac{1}{\sqrt{8}}\left(\begin{array}{c}\left|P_{n/2}^{i00}\right\rangle\left|P_{n/2}^{01k}\right\rangle+\left|P_{n/2}^{i10}\right\rangle\left|P_{n/2}^{00k}\right\rangle+\left|P_{n/2}^{i00}\right\rangle\left|P_{n/2}^{11k}\right\rangle+\left|P_{n/2}^{i10}\right\rangle\left|P_{n/2}^{10k}\right\rangle+\\ \left|P_{n/2}^{i01}\right\rangle\left|P_{n/2}^{01k}\right\rangle+\left|P_{n/2}^{i11}\right\rangle\left|P_{n/2}^{00k}\right\rangle+\left|P_{n/2}^{i01}\right\rangle\left|P_{n/2}^{10k}\right\rangle+\left|P_{n/2}^{i11}\right\rangle\left|P_{n/2}^{11k}\right\rangle\end{array}\right).$$

Therefore $\mathrm{TS}\left(\left|P_n^{ijk}\right\rangle\right) \leq 16\,\mathrm{TS}\left(\left|P_{n/2}^{ijk}\right\rangle\right)$, and solving this recurrence relation yields $\mathrm{TS}\left(\left|P_n^{ijk}\right\rangle\right) = O\left(n^4\right)$. Finally observe that

$$|\psi\rangle = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes n} - \frac{\left|P_n^{010}\right\rangle + \left|P_n^{011}\right\rangle + \left|P_n^{110}\right\rangle + \left|P_n^{111}\right\rangle}{\sqrt{2}}.$$

■

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 $\mathsf{TreeBQP}$ be the class of languages accepted by such a machine, does $\mathsf{TreeBQP} = \mathsf{BPP}$? 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 $\mathsf{TreeBQP}$ 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 $\mathsf{TreeBQP}$ 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 $\mathsf{TreeBQP} \neq \mathsf{BQP}$. (On the other hand, we do not yet have oracle evidence even for $\mathsf{BQP} \not\subset \mathsf{AM}$, though not for lack of trying [2].)

Definition 22 $\mathsf{TreeBQP}$ *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 (as we will show in Theorem 23, part (i), the choice of gate set will not 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 $\left|\psi^{(t)}\right\rangle$ 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.[10] 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.

[10] 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.

Theorem 23

(i) The definition of TreeBQP *is invariant under the choice of gate set.*

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

(iii) BPP $\subseteq$ TreeBQP $\subseteq$ BQP.

Proof.

(i) The Solovay-Kitaev Theorem [31, 36] 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 $O(\text{polylog}(1/\varepsilon))$. 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 TreeBQP. Furthermore, since the approximation circuit for $g^{(t)}$ acts only on k qubits, any intermediate state $|\varphi\rangle$ it produces satisfies $\mathrm{TS}_{1/2^{\Omega(n)}}(|\varphi\rangle) \leq k4^k\, \mathrm{TS}_{1/2^{\Omega(n)}}\left(\left|\psi^{(t-1)}\right\rangle\right)$ by Proposition 2.

(ii) 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 $\mathrm{MFS}\left(T_k^h\right) \leq 2h \max_i \mathrm{MFS}\left(T_{\lceil k/2 \rceil}^h\right) + O(1)$, and solving this recurrence yields $\mathrm{MFS}\left(T_k^{k/2}\right) = k^{O(\log k)}$. 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.

(iii) 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.

■

Theorem 24 TreeBQP $\subseteq \Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$.

Proof. Since TreeBQP is closed under complement, it suffices to show that 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 $\mathsf{AM} \subseteq \Pi_2^{\mathsf{P}}$ [26], 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.

■

In the proof of Theorem 24, 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 TreeBQP except that any states in AmpP are allowed, then $\mathsf{AmpP\text{-}BQP} \subseteq \Sigma_3^{\mathsf{P}} \cap \Pi_3^{\mathsf{P}}$ as well.

7 The Experimental Situation

The results of this paper suggest an obvious challenge for experimenters: *prepare non-tree states in the lab*. For were this challenge met, it would rule out one way in which quantum mechanics could fail, just as the Bell inequality experiments of Aspect et al. [8] did twenty years ago. If they wished, quantum computing skeptics could then propose a new candidate Sure/Shor separator, and experimenters could try to rule out *that* one, and so on. The result would be to divide the question of whether quantum computing is possible into a series of smaller questions about which states can be prepared. In our view, this would aid progress in two ways: by helping experimenters set clear goals, and by forcing theorists to state clear positions.

However, our experimental challenge raises some immediate questions. In particular, what would it *mean* to prepare a non-tree state? How would we know if we succeeded? Also, have non-tree states already been prepared (or observed)? The purpose of this section is to set out our thoughts about these questions.

First of all, when discussing experiments, it goes without saying that we must convert asymptotic statements into statements about specific values of n. The central tenet of computational complexity theory is that this is possible. Thus, instead of asking whether n-qubit states with tree size $2^{\Omega(n)}$ can be prepared, we ask whether 200-qubit states with tree size at least (say) 2^{80} can be prepared. Even though the second question does not logically imply anything about the first, the second is closer to what we ultimately care about anyway. Admittedly, knowing that $\mathrm{TS}(|\psi_n\rangle) = n^{\Omega(\log n)}$ tells us little about $\mathrm{TS}(|\psi_{100}\rangle)$ or $\mathrm{TS}(|\psi_{200}\rangle)$, especially since in Raz's paper [40], the constant in the exponent $\Omega(\log n)$ is taken to be 10^{-6} (though this can certainly be improved). Thus, proving tight lower bounds for small n is one of the most important problems left open by this paper. In Appendix 10 we solve the problem for the case of manifestly orthogonal tree size.

A second common objection is that our formalism applies only to pure states, but in reality all states are mixed. However, there are several natural ways to extend the formalism to mixed states. Given a mixed state ρ, we could minimize tree size over all purifications of ρ, or minimize the expected tree size $\sum_i |\alpha_i|^2 \mathrm{TS}(|\psi_i\rangle)$, or maximum $\max_i \mathrm{TS}(|\psi_i\rangle)$, over all decompositions $\rho = \sum_i \alpha_i |\psi_i\rangle \langle\psi_i|$.

A third objection is a real quantum state might be a "soup" of free-wandering fermions and bosons, with no localized subsystems corresponding to qubits. How can one determine the tree size of such a state? The answer is that one cannot. Any complexity measure for particle position and momentum states would have to be quite different from the measures considered in this paper. On the other hand, the states of interest for quantum computing usually *do* involve localized qubits. Indeed, even if quantum information is stored in particle positions, one might force each particle into two sites (corresponding to $|0\rangle$ and $|1\rangle$), neither of which can be occupied by any other particle. In that case it again becomes meaningful to discuss tree size.

But how do we verify that a state with large tree size was prepared? Of course, if $|\psi\rangle$ is preparable by a polynomial-size quantum circuit, then *assuming quantum mechanics is valid* (and assuming our gates behave as specified), we can always test whether a given state $|\varphi\rangle$ is close to $|\psi\rangle$ or not. Let U map $|0\rangle^{\otimes n}$ to $|\psi\rangle$; then it suffices to test whether $U^{-1}|\varphi\rangle$ is close to $|0\rangle^{\otimes n}$. However, in the experiments under discussion, the validity of quantum mechanics is the very point in question. And once we allow Nature to behave in arbitrary ways, a skeptic could explain *any* experimental result without having to invoke states with large tree size.

The above fact has often been urged against us, but as it stands, it is no different from the fact that one could explain any astronomical observation without abandoning the Ptolemaic system. The issue is not one of mathematical proof, but of accumulating observations that are consistent with the hypothesis of large tree size, and inconsistent with alternative hypotheses if we disallow special pleading. So for example, to test whether the subgroup state

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

was prepared, we might use CNOT gates to map $|x\rangle$ to $|x\rangle \left|v^T x\right\rangle$ for some vector $v \in \mathbb{Z}_2^n$. Based on our knowledge of S, we could then predict whether the qubit $\left|v^T x\right\rangle$ should be $|0\rangle$, $|1\rangle$, or an equal mixture of $|0\rangle$ and $|1\rangle$ when measured. Or we could apply Hadamard gates to all n qubits of $|S\rangle$, then perform the same test for the subgroup dual to S. In saying that a system is in state $|S\rangle$, it is not clear if we *mean* anything more than that it responds to all such tests in expected ways. Similar remarks apply to Shor states and cluster states.

In our view, tests of the sort described above are certainly *sufficient*, so the interesting question is whether they are *necessary*, or whether weaker and more indirect tests would also suffice. This question rears its head when we ask whether non-tree states have already been observed. For as pointed out to us by Anthony Leggett, there exist systems studied in condensed-matter physics that are strong candidates for having superpolynomial tree size. An example is the magnetic salt $\mathrm{LiHo}_x\mathrm{Y}_{1-x}\mathrm{F}_4$ studied by Ghosh et al. [24], which, like the cluster states of Briegel and Raussendorf [13], basically consists of a lattice of spins subject to pairwise nearest-neighbor Hamiltonians. The main differences are that the salt lattice is 3-D instead of 2-D, is tetragonal instead of cubic, and is irregular in that not every site is occupied by a spin. Also, there are weak interactions even between spins that are not nearest neighbors. But none of these differences seems likely to change a superpolynomial tree size into a polynomial one.

For us, the main issues are (1) how precisely can we characterize[11] the quantum state of the magnetic salt, and (2) how strong the evidence is that that *is* the state. What Ghosh et al. [24] did was to calculate bulk properties of the salt, such as its magnetic susceptibility and specific heat, with and without taking into account the quantum entanglement generated by the nearest-neighbor Hamiltonians. They found that including entanglement yielded a better fit to the experimentally measured values. However, this is clearly a far cry from preparing a system in a state of one's choosing by applying a known pulse sequence, and then applying any of a vast catalog of tests to verify that the state was prepared. So it would be valuable to have more direct evidence that states qualitatively like cluster states can exist in Nature.

In summary, our results underscore the importance of current experimental work on large, persistently entangled quantum states; but they also suggest a new motivation and perspective for this work. They suggest that we reexamine known condensed-matter systems with a new goal in mind: understanding the complexity of their associated quantum states. They also suggest that 2-D cluster states and random subgroup states are interesting in a way that 1-D spin chains and Schrödinger cat states are not. Yet when experimenters try to prepare states of the former type, they often see it as merely a stepping stone towards demonstrating error-correction or another quantum computing benchmark. Thus, Knill et al. [32] prepared[12] the 5-qubit state

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

for which $\mathrm{MOTS}(|\psi\rangle) = 40$ from the decomposition

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

and for which we conjecture $\mathrm{TS}(|\psi\rangle) = 40$ as well. However, the sole motivation of the experiment was to demonstrate a 5-qubit quantum error-correcting code. In our opinion, whether states with large tree size can be prepared is a fundamental question in its own right. Were that question studied directly, perhaps we could address it for larger numbers of qubits.

[11] By "characterize," we mean give an explicit formula for the amplitudes at a particular time t, in some standard basis. If a state is characterized as the ground state of a Hamiltonian, then we first need to solve for the amplitudes before we can prove tree size lower bounds using Raz's method.

[12] Admittedly, what they really prepared is the 'pseudo-pure' state $\rho = \varepsilon |\psi\rangle \langle\psi| + (1-\varepsilon) I$, where I is the maximally mixed state and $\varepsilon \approx 10^{-5}$. Braunstein et al. [11] have shown that, if the number of qubits n is less than about 14, then such states cannot be entangled. That is, there exists a representation of ρ as a mixture of pure states, each of which is separable and therefore has tree size $O(n)$. This is a well-known limitation of the liquid NMR technology used by Knill et al. Thus, a key challenge is to replicate the successes of liquid NMR using colder qubits.

Let us end by stressing that, in the perspective we are advocating, there is nothing sacrosanct about tree size as opposed to other complexity measures. This paper concentrated on tree size because it is the subject of our main results, and because it is better to be specific than vague. On the other hand, Section 4, Appendix 9, and Appendix 10 contain numerous results about orthogonal tree size, manifestly orthogonal tree size, Vidal's χ complexity, and other measures. Readers dissatisfied with *all* of these measures are urged to propose new ones, perhaps motivated directly by experiments. We see nothing wrong with having multiple ways to quantify the complexity of quantum states, and much wrong with having no ways.

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 large-scale 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. Furthermore, 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 should instead be some sort of *complexity*. That is, Sure states should 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 classical to quantum 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 tree size lower bounds?

(2) Can we prove Conjecture 19, implying an $n^{\Omega(\log n)}$ tree size lower bound for Shor states?

(3) 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?

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

(5) 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?

(6) 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.

(7) Is there a practical method to compute the tree size of, say, 10-qubit states? Such a method would have great value in interpreting experimental results.

Acknowledgments

I thank Ran Raz for fruitful correspondence and for sharing an early version of his paper; the anonymous reviewers for detailed comments that improved the paper enormously; and Andrej Bogdanov, Don Coppersmith, Viatcheslav Dobrovitski, Oded Goldreich, Ray Laflamme, Anthony Leggett, Leonid Levin, Mike Mosca, Ashwin Nayak, Carl Pomerance, John Preskill, Alexander Razborov, Peter Shor, Rob Spekkens, Barbara Terhal, Luca Trevisan, 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*, 2(5):410–423, 2002. quant-ph/0206089.

[2] S. Aaronson. Quantum lower bound for recursive Fourier sampling. *Quantum Information and Computation*, 3(2):165–174, 2003. ECCC TR02-072, quant-ph/0209060.

[3] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. Lett.*, 70(052328), 2004. quant-ph/0406196.

[4] 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.

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

[6] A. Ambainis, L. J. Schulman, A. Ta-Shma, U. V. Vazirani, and A. Wigderson. The quantum communication complexity of sampling. *SIAM J. Comput.*, 32:1570–1585, 2003.

[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 J. Comput.*, 26(5):1411–1473, 1997. First appeared in ACM STOC 1993.

[10] M. L. Bonet and S. R. Buss. Size-depth tradeoff for Boolean formulae. *Inform. 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] H. J. Briegel and R. Raussendorf. Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.*, 86:910–913, 2001. quant-ph/0004051.

[14] N. H. Bshouty, R. Cleve, and W. Eberly. Size-depth tradeoffs for algebraic formulae. *SIAM J. Comput.*, 24(4):682–705, 1995.

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

[16] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1098–1105, 1996. quant-ph/9512032.

[17] J. Cronin. CP symmetry violation - the search for its origin. Nobel Lecture, December 8, 1980.

[18] W. Dür and H. J. Briegel. Stability of macroscopic entanglement under decoherence. *Phys. Rev. Lett.*, 92, 2004. quant-ph/0307180.

[19] V. Fitch. The discovery of charge-conjugation parity asymmetry. Nobel Lecture, December 8, 1980.

[20] 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.

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

[22] 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.

[23] G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D*, 34:470–491, 1986.

[24] S. Ghosh, T. F. Rosenbaum, G. Aeppli, and S. N. Coppersmith. Entangled quantum state of magnetic dipoles. *Nature*, 425:48–51, 2003. cond-mat/0402456.

[25] O. Goldreich. On quantum computing. www.wisdom.weizmann.ac.il/˜oded/on-qc.html, 2004.

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

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

[28] F. Green, S. Homer, C. Moore, and C. Pollett. Counting, fanout, and the complexity of quantum ACC. *Quantum Information and Computation*, 2(1):35–65, 2002. quant-ph/0106017.

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

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

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

[32] E. Knill, R. Laflamme, R. Martinez, and C. Negrevergne. Implementation of the five qubit error correction benchmark. *Phys. Rev. Lett.*, 86:5811–5814, 2001. quant-ph/0101034.

[33] A. J. Leggett. Testing the limits of quantum mechanics: motivation, state of play, prospects. *J. Phys. Condensed Matter*, 14:R415–451, 2002.

[34] L. A. Levin. Polynomial time and extravagant models, in The tale of one-way functions. *Problems of Information Transmission*, 39(1):92–103, 2003. cs.CR/0012023.

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

[36] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[37] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994.

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

[39] R. Raussendorf, D. E. Browne, and H. J. Briegel. Measurement-based quantum computation on cluster states. *Phys. Rev. A*, 68, 2003. quant-ph/0301052.

[40] R. Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. In *Proc. ACM STOC*, pages 633–641, 2004. ECCC TR03-067.

[41] R. Raz. Multilinear-$NC_1 \neq$ Multilinear-NC_2. In *Proc. IEEE FOCS*, pages 344–351, 2004. ECCC TR04-042.

[42] A. Steane. Multiple particle interference and quantum error correction. *Proc. Roy. Soc. London*, A452:2551–2577, 1996. quant-ph/9601029.

[43] G. 't Hooft. Quantum gravity as a dissipative deterministic system. *Classical and Quantum Gravity*, 16:3263–3279, 1999. gr-qc/9903084.

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

[45] L. G. Valiant and V. V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Comput. Sci.*, 47(3):85–93, 1986.

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

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

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

9 Appendix: Relations Among Quantum State Classes

This appendix presents some results about the quantum state hierarchy introduced in Section 3. Theorem 25 shows simple inclusions and separations, while Theorem 26 shows that separations higher in the hierarchy would imply major complexity class separations (and vice versa).

Theorem 25

(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,* $\otimes_3 \neq \Sigma_4 \cap \otimes_4$.

Proof.

(i) $\mathsf{Tree} \subseteq \mathsf{Circuit}$ since any multilinear tree is also a multilinear circuit. $\mathsf{Circuit} \subseteq \mathsf{AmpP}$ since the circuit yields a polynomial-time algorithm for computing the amplitudes. For $\mathsf{Vidal} \subseteq \mathsf{Circuit}$, we use an idea of Vidal [46]: given $|\psi_n\rangle \in \mathsf{Vidal}$, for all $j \in \{1, \ldots, n\}$ we can express $|\psi_n\rangle$ as

$$\sum_{i=1}^{\chi(|\psi\rangle)} \alpha_{ij} \left|\phi_i^{[1\ldots j]}\right\rangle \otimes \left|\phi_i^{[j+1\ldots n]}\right\rangle$$

where $\chi(|\psi_n\rangle)$ is polynomially bounded. Furthermore, Vidal showed that each $\left|\phi_i^{[1\ldots j]}\right\rangle$ can be written as a linear combination of states of the form $\left|\phi_i^{[1\ldots j-1]}\right\rangle \otimes |0\rangle$ and $\left|\phi_i^{[1\ldots j-1]}\right\rangle \otimes |1\rangle$—the point being that the set of $\left|\phi_i^{[1\ldots j-1]}\right\rangle$ states is the same, independently of $\left|\phi_i^{[1\ldots j]}\right\rangle$. This immediately yields a polynomial-size multilinear circuit for $|\psi_n\rangle$.

(ii) Given $|\psi_n\rangle \in \mathsf{Vidal}$, we can decompose $|\psi_n\rangle$ as

$$\sum_{i=1}^{\chi(|\psi\rangle)} \alpha_i \left|\phi_i^{[1\ldots n/2]}\right\rangle \otimes \left|\phi_i^{[n/2+1\ldots n]}\right\rangle.$$

Then $\chi\left(\left|\phi_i^{[1\ldots n/2]}\right\rangle\right) \leq \chi(|\psi_n\rangle)$ and $\chi\left(\left|\phi_i^{[n/2+1\ldots n]}\right\rangle\right) \leq \chi(|\psi_n\rangle)$ for all i, so we can recursively decompose these states in the same manner. It follows that $\mathrm{TS}(|\psi_n\rangle) \leq 2\chi(|\psi\rangle)\,\mathrm{TS}\left(\left|\psi_{n/2}\right\rangle\right)$; solving this recurrence relation yields $\mathrm{TS}(|\psi_n\rangle) \leq (2\chi(|\psi\rangle))^{\log n} = n^{O(\log n)}$.

(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: $2^{-n/4}(|00\rangle + |11\rangle)^{\otimes n/2}$.

(iv) $\otimes_2 \subseteq \mathsf{MOTree}$ is obvious, while $\mathsf{MOTree} \not\subset \otimes_2$ follows from the example of $\left|P_n^i\right\rangle$, an equal superposition over all n-bit strings of parity i. The following recursive formulas imply that

$\mathrm{MOTS}\left(\left|P_n^i\right\rangle\right) \leq 4\,\mathrm{MOTS}\left(\left|P_{n/2}^i\right\rangle\right) = O\left(n^2\right)$:

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

On the other hand, $|P_n\rangle \notin \otimes_2$ follows from $|P_n\rangle \notin \Sigma_1$ together with 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—for example,

$$\frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes n} + \left(\frac{|01\rangle + |10\rangle}{\sqrt{2}}\right)^{\otimes n/2}\right).$$

$\otimes_2 \neq \otimes_3$ 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 $\Sigma_3 \subseteq \Sigma_4 \cap \otimes_4$.

■

Theorem 26

(i) $\mathsf{BQP} = \mathsf{P}^{\#\mathsf{P}}$ *implies* $\mathsf{AmpP} \subseteq \Psi\mathsf{P}$.

(ii) $\mathsf{AmpP} \subseteq \Psi\mathsf{P}$ *implies* $\mathsf{NP} \subseteq \mathsf{BQP/poly}$.

(iii) $\mathsf{P} = \mathsf{P}^{\#\mathsf{P}}$ *implies* $\Psi\mathsf{P} \subseteq \mathsf{AmpP}$.

(iv) $\Psi\mathsf{P} \subseteq \mathsf{AmpP}$ *implies* $\mathsf{BQP} \subseteq \mathsf{P/poly}$.

Proof.

(i) First, $\mathsf{BQP} = \mathsf{P}^{\#\mathsf{P}}$ implies $\mathsf{BQP/poly} = \mathsf{P}^{\#\mathsf{P}}/\mathsf{poly}$, since given a $\mathsf{P}^{\#\mathsf{P}}/\mathsf{poly}$ machine M, the language consisting of all (x, a) such that M accepts on input x and advice a is clearly in BQP. So assume $\mathsf{BQP/poly} = \mathsf{P}^{\#\mathsf{P}}/\mathsf{poly}$, and consider a state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$ with $|\psi\rangle \in \mathsf{AmpP}$. By the result of Bernstein and Vazirani [9] that $\mathsf{BQP} \subseteq \mathsf{P}^{\#\mathsf{P}}$, for all b there exists a quantum circuit of size polynomial in n and b that approximates $p_0 = \sum_{y\in\{0,1\}^{n-1}} |\alpha_{0y}|^2$, or the probability that the first qubit is measured to be 0, to b bits of precision. So by uncomputing garbage, we can prepare a state close to $\sqrt{p_0}\,|0\rangle + \sqrt{1-p_0}\,|1\rangle$. Similarly, given a superposition over length-k prefixes of x, we can prepare a superposition over length-$(k+1)$ prefixes of x by approximating the conditional measurement probabilities. We thus obtain a state close to $\sum_x |\alpha_x|\,|x\rangle$. The last step is to approximate the phase of each $|x\rangle$, apply that phase, and uncompute to obtain a state close to $\sum_x \alpha_x |x\rangle$.

(ii) Given a *SAT* instance, first use Valiant-Vazirani [45] to produce a formula φ with either 0 or 1 satisfying assignments. Then let $\alpha_x = 1$ if x is a satisfying assignment for φ and $\alpha_x = 0$ otherwise; clearly $|\psi\rangle = \sum_x \alpha_x |x\rangle$ is in AmpP. By the assumption $\mathsf{AmpP} \subseteq \Psi\mathsf{P}$, there exists a polynomial-size quantum circuit that approximates $|\psi\rangle$, and thereby finds the unique satisfying assignment for φ if it exists.

(iii) As in part (i), $\mathsf{P} = \mathsf{P}^{\#\mathsf{P}}$ implies $\mathsf{P/poly} = \mathsf{P}^{\#\mathsf{P}}/\mathsf{poly}$. The containment $\Psi\mathsf{P} \subseteq \mathsf{AmpP}$ follows since we can approximate amplitudes to polynomially many bits of precision in $\#\mathsf{P}$.

(iv) As is well known [9], any quantum computation can be made 'clean' in the sense that it accepts if and only if a particular basis state (say $|0\rangle^{\otimes n}$) is measured. The implication follows easily.

■

10 Appendix: Manifestly Orthogonal Tree Size

This appendix studies the manifestly orthogonal tree size of coset states:[13] states having the form

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

where $C = \{x \mid Ax \equiv b\}$ is a coset in $\mathbb{Z}_2^n$. In particular, we present a *tight* characterization of $\mathrm{MOTS}(|C\rangle)$, which enables us to prove *exponential* lower bounds on it, in contrast to the $n^{\Omega(\log n)}$ lower bounds for ordinary tree size. This characterization also yields a separation between orthogonal and manifestly orthogonal tree size; and an algorithm for computing $\mathrm{MOTS}(|C\rangle)$ whose complexity is only singly exponential in n. Our proof technique is independent of Raz's, and is highly tailored to take advantage of manifest orthogonality. However, even if our technique finds no broader application, the fact that it gives tight bounds makes it almost unique—and thus, we hope, of interest to complexity theorists.

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

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

be a uniform superposition over the elements of S, and let $M(S)$ be a shorthand for $\mathrm{MOTS}(|S\rangle)$.

Let $C = \{x : Ax \equiv b\}$ be a subgroup in $\mathbb{Z}_2^n$, for some $A \in \mathbb{Z}_2^{k \times n}$ and $b \in \mathbb{Z}_2^k$. Let $[n] = \{1, \ldots, n\}$, and let (I, J) be a nontrivial partition of $[n]$ (one where I and J are both nonempty). Then clearly there exist distinct cosets $C_I^{(1)}, \ldots, C_I^{(H)}$ in the I subsystem, and distinct cosets $C_J^{(1)}, \ldots, C_J^{(H)}$ in the J subsystem, such that

$$C = \bigcup_{h \in [H]} C_I^{(h)} \otimes C_J^{(h)}.$$

[13] All results apply equally well to the subgroup states of Section 5.1; the greater generality of coset states is just for convenience.

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

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

Lemma 27 *For all cosets C,*

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

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

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

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

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

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

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

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

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

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

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

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

Theorem 28 *If $n \geq 2$, then*

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

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

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

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

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

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

where $M\left(S^{(j)}\right) = M(S_I) + M(S_J)$ follows from the minimality of T. As in Lemma 27, there must be an h such that $S_I \subseteq C_I^{(h)}$ and $S_J \subseteq C_J^{(h)}$. But Lemma 27 then implies that $M(C_I) / |C_I| \leq M(S_I) / |S_I|$ and that $M(C_J) / |C_J| \leq M(S_J) / |S_J|$. Combining these bounds with $|C_I| \geq |S_I|$ and $|C_J| \geq |S_J|$, we obtain by a harmonic mean inequality that

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

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

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

■

We can express Theorem 28 directly in terms of the matrix A as follows. Let $M(A) = M(C) = \mathrm{MOTS}(|C\rangle)$ where $C = \{x : Ax \equiv b\}$ (the vector b is irrelevant, so long as $Ax \equiv b$ is solvable). Then

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

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

Corollary 29 *There exists a deterministic $O(n3^n)$-time algorithm that computes $M(A)$, given A as input.*

Proof. First compute $\operatorname{rank}(A^*)$ for all 2^{n-1} matrices A^* that are formed by choosing a subset of the columns of A. This takes time $O\left(n^3 2^n\right)$. Then compute $M(A^*)$ for all A^* with one column, then for all A^* with two columns, and so on, applying the formula (*) recursively. This takes time

$$\sum_{t=1}^{n} \binom{n}{t} t 2^t = O\left(n 3^n\right).$$

■

Another easy consequence of Theorem 28 is that the language $\{A : M(A) \leq s\}$ is in NP. We do not know whether this language is NP-complete but suspect it is.

As we mentioned, our characterization lets us prove exponential lower bounds on the manifestly orthogonal tree size of coset states.

Theorem 30 *Suppose the entries of $A \in \mathbb{Z}_2^{k \times n}$ are drawn uniformly and independently at random, where $k \in \left[4 \log_2 n, \frac{1}{2}\sqrt{n \ln 2}\right]$. Then $M(A) = \left(n/k^2\right)^{\Omega(k)}$ with probability $\Omega(1)$ over A.*

Proof. Let us upper-bound the probability that certain "bad events" occur when A is drawn. The first bad event is that A contains an all-zero column. This occurs with probability at most $2^{-k} n = o(1)$. The second bad event is that there exists a $k \times d$ submatrix of A with $d \geq 12k$ that has rank at most $2k/3$. This also occurs with probability $o(1)$. For we claim that, if A^* is drawn uniformly at random from $\mathbb{Z}_2^{k \times d}$, then

$$\Pr_{A_I}\left[\operatorname{rank}(A^*) \leq r\right] \leq \binom{d}{r} \left(\frac{2^r}{2^k}\right)^{d-r}.$$

To see this, imagine choosing the columns of A^* one by one. For $\operatorname{rank}(A^*)$ to be at most r, there must be at least $d - r$ columns that are linearly dependent on the previous columns. But each column is dependent on the previous ones with probability at most $2^r/2^k$. The claim then follows from the union bound. So the probability that *any* $k \times d$ submatrix of A has rank at most r is at most

$$\binom{n}{d}\binom{d}{r}\left(\frac{2^r}{2^k}\right)^{d-r} \leq n^d d^r \left(\frac{2^r}{2^k}\right)^{d-r}.$$

Set $r = 2k/3$ and $d = 12k$; then the above is at most

$$\exp\left\{12k \log n + \frac{2k}{3}\log(12k) - \left(12k - \frac{2k}{3}\right)\frac{k}{3}\right\} = o(1)$$

where we have used the fact that $k \geq 4 \log n$.

Assume that neither bad event occurs, and let $\left(A_I^{(0)}, A_J^{(0)}\right)$ be a partition of the columns of A that minimizes the expression (*). Let $A^{(1)} = A_I^{(0)}$ if $\left|A_I^{(0)}\right| \geq \left|A_J^{(0)}\right|$ and $A^{(1)} = A_J^{(0)}$ otherwise, where $\left|A_I^{(0)}\right|$ and $\left|A_J^{(0)}\right|$ are the numbers of columns in $A_I^{(0)}$ and $A_J^{(0)}$ respectively (so that $\left|A_I^{(0)}\right| + \left|A_J^{(0)}\right| = n$). Likewise, let $\left(A_I^{(1)}, A_J^{(1)}\right)$ be an optimal partition of the columns of $A^{(1)}$, and let $A^{(2)} = A_I^{(1)}$ if $\left|A_I^{(1)}\right| \geq \left|A_J^{(1)}\right|$ and $A^{(2)} = A_J^{(1)}$ otherwise. Continue in this way until an $A^{(t)}$ is reached such that $\left|A^{(t)}\right| = 1$. Then an immediate consequence of (*) is that $M(A) \geq Z^{(0)} \cdot \cdots \cdot Z^{(t-1)}$ where

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

and $A^{(0)} = A$.

Call l a "balanced cut" if $\min\left\{\left|A_I^{(l)}\right|, \left|A_J^{(l)}\right|\right\} \geq 12k$, and an "unbalanced cut" otherwise. If l is a balanced cut, then $\operatorname{rank}\left(A_I^{(l)}\right) \geq 2k/3$ and $\operatorname{rank}\left(A_J^{(l)}\right) \geq 2k/3$, so $Z^{(l)} \geq 2^{k/3}$. If l is an unbalanced cut, then call l a "freebie" if $\operatorname{rank}\left(A_I^{(l)}\right) + \operatorname{rank}\left(A_J^{(l)}\right) = \operatorname{rank}\left(A^{(l)}\right)$. There can be at most k freebies, since for each one, $\operatorname{rank}\left(A^{(l+1)}\right) < \operatorname{rank}\left(A^{(l)}\right)$ by the assumption that all columns of A are nonzero. For the other unbalanced cuts, $Z^{(l)} \geq 2$.

Assume $\left|A^{(l+1)}\right| = \left|A^{(l)}\right|/2$ for each balanced cut and $\left|A^{(l+1)}\right| = \left|A^{(l)}\right| - 12k$ for each unbalanced cut. Then if our goal is to minimize $Z^{(0)} \cdot \cdots \cdot Z^{(t-1)}$, clearly the best strategy is to perform balanced cuts first, then unbalanced cuts until $\left|A^{(l)}\right| = 12k^2$, at which point we can use the k freebies. Let B be the number of balanced cuts; then

$$Z^{(0)} \cdot \cdots \cdot Z^{(t-1)} = \left(2^{k/3}\right)^B 2^{\left(n/2^B - 12k^2\right)/12k}.$$

This is minimized by taking $B = \log_2\left(\frac{n \ln 2}{4k^2}\right)$, in which case $Z^{(0)} \cdot \cdots \cdot Z^{(t-1)} = \left(n/k^2\right)^{\Omega(k)}$. ■

A final application of our characterization is to separate orthogonal from manifestly orthogonal tree size.

Corollary 31 *There exist states with polynomially-bounded orthogonal tree size, but manifestly orthogonal tree size* $n^{\Omega(\log n)}$. *Thus* $\mathsf{OTree} \neq \mathsf{MOTree}$.

Proof. Set $k = 4\log_2 n$, and let $C = \{x : Ax \equiv 0\}$ where A is drawn uniformly at random from $\mathbb{Z}_2^{k \times n}$. Then by Theorem 30,

$$\operatorname{MOTS}(|C\rangle) = \left(n/k^2\right)^{\Omega(k)} = n^{\Omega(\log n)}$$

with probability $\Omega(1)$ over A. On the other hand, if we view $|C\rangle$ in the Fourier basis (that is, apply a Hadamard to every qubit), then the resulting state has only $2^k = n^{16}$ basis states with nonzero amplitude, and hence has orthogonal tree size at most n^{17}. So by Proposition 2, part (i), $\operatorname{OTS}(|C\rangle) \leq 2n^{17}$ as well. ■

Indeed, the orthogonal tree states of Corollary 31 are superpositions over polynomially many separable states, so we also obtain that $\Sigma_2 \not\subset \mathsf{MOTree}$.