2018 IEEE 59th Annual Symposium on Foundations of Computer Science

Quantum algorithm for simulating real time evolution of lattice Hamiltonians*

Jeongwan Haah*, Matthew B. Hastings*†, Robin Kothari*, Guang Hao Low*
*Quantum Architectures and Computation, Microsoft Research, Redmond, Washington, USA
†Station Q, Microsoft Research, Santa Barbara, California, USA

***Abstract*—We study the problem of simulating the time evolution of a lattice Hamiltonian, where the qubits are laid out on a lattice and the Hamiltonian only includes geometrically local interactions (i.e., a qubit may only interact with qubits in its vicinity). This class of Hamiltonians is very general and encompasses essentially all physical Hamiltonians.**

Our algorithm simulates the time evolution of such a Hamiltonian on n qubits for time T up to error ϵ using $O(nT\,\mathrm{polylog}(nT/\epsilon))$ gates with depth $O(T\,\mathrm{polylog}(nT/\epsilon))$. Our algorithm is the first simulation algorithm that achieves gate cost quasilinear in nT and polylogarithmic in $1/\epsilon$. Our algorithm also readily generalizes to time-dependent Hamiltonians and yields an algorithm with similar gate count for any piecewise slowly varying time-dependent bounded local Hamiltonian.

We also prove a matching lower bound on the gate count of such a simulation, showing that any quantum algorithm that can simulate a piecewise constant bounded local Hamiltonian in one dimension to constant error requires $\widetilde{\Omega}(nT)$ gates in the worst case. The lower bound holds even if we only require the output state to be correct on local measurements. To our best knowledge, this is the first nontrivial lower bound on the gate complexity of the simulation problem.

Our algorithm is based on a decomposition of the time-evolution unitary into a product of small unitaries using Lieb-Robinson bounds. In the appendix, we prove a Lieb-Robinson bound tailored to Hamiltonians with small commutators between local terms, giving zero Lieb-Robinson velocity in the limit of commuting Hamiltonians. This improves the performance of our algorithm when the Hamiltonian is close to commuting.

***Keywords*-Quantum algorithms; Hamiltonian simulation; Local Hamiltonians; Lieb-Robinson bound**

I. Introduction

1) Background: The problem of simulating the time evolution of a quantum system is perhaps the most important application of quantum computers. Indeed, this was the reason Feynman proposed quantum computing [1], and it remains an important practical application since a significant fraction of the world's supercomputing power is used to solve instances of this problem that arise in materials science, condensed matter physics, high energy physics, and chemistry [2].

All known classical algorithms (i.e., algorithms that run on traditional non-quantum computers) for this problem run in exponential time. On the other hand, from the early days of quantum computing [1], [3] it was known that quantum computers can solve this problem in polynomial time. More precisely, when formalized as a decision problem, the problem of simulating the time evolution of a quantum system is in the complexity class BQP, the class of problems solved by a quantum computer to bounded error in polynomial time. Furthermore, the problem is complete for BQP [4], [5], which means we do not expect there to be efficient classical algorithms for the problem, since that would imply BPP = BQP, which in turn would imply polynomial-time algorithms for problems such as integer factorization and discrete log [6].

2) Hamiltonian simulation problem: The Hamiltonian simulation problem is a standard formalization of the problem of simulating the time evolution[1] of a quantum system. In this problem, we assume the quantum system whose time evolution we wish to simulate consists of n qubits and we want to simulate its time evolution for time T, in the sense that we are provided with the initial state $|\psi(0)\rangle$ and we want to compute the state of the system at time T, $|\psi(T)\rangle$. The goal of an efficient simulation is to solve the problem in time polynomial in n and T.

The state of a system of n qubits can be described by a complex vector of dimension 2^n of unit norm. Since we are studying quantum algorithms for the problem, we are given the input as an n-qubit quantum state, and have to output an n-qubit quantum state. The relation between the output state at time T and the initial state at time 0 is given by the Schrödinger equation

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \tag{1}$$

where the Hamiltonian H, a $2^n \times 2^n$ complex Hermitian matrix, has entries which may also be functions of time. The Hamiltonian captures the interaction between the constituents of the system and governs time dynamics. In the special case where the Hamiltonian is independent

* The full version of this paper is available at https://arxiv.org/abs/1801.03922

[1]This is sometimes referred to as "real time evolution", to distinguish it from "imaginary time evolution" which we will not talk about in this paper.

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00041

of time, the Schrödinger equation can be solved to yield $|\psi(T)\rangle = e^{-iHT}|\psi(0)\rangle$.

More formally, the input to the Hamiltonian simulation problem consists of a Hamiltonian H (or $H(t)$ in the time-dependent case), a time T, and an error parameter ϵ. The goal is to output a quantum circuit that approximates the unitary matrix that performs the time evolution above (e.g., for time-independent Hamiltonians, the quantum circuit should approximate the unitary e^{-iHT}). The notion of approximation used is the spectral norm distance between the ideal unitary and the one performed by the circuit. The cost of a quantum circuit is measured by the number of gates used in the circuit, where the gates come from some fixed universal gate set. Note that it is important to describe how the Hamiltonian in the input is specified, since it is a matrix of size $2^n \times 2^n$. This will be made precise when talking about specific classes of Hamiltonians that we would like to simulate.

3) Geometrically local Hamiltonian simulation: The most general class of Hamiltonians that is commonly studied in the literature is the class of sparse Hamiltonians [7]–[15]. A Hamiltonian on n qubits is sparse if it has only $\mathrm{poly}(n)$ nonzero entries in any row or column. For such Hamiltonians, we assume we have an efficient algorithm to compute the nonzero entries in each row or column, and the input Hamiltonian is specified by an oracle that can be queried for this information. In this model, recent quantum algorithms have achieved optimal complexity in terms of the queries made to the oracle [13].

A very important special case of this type of Hamiltonian is a "local Hamiltonian." Confusingly, this term is used to describe two different kinds of Hamiltonians in the literature. We distinguish these two definitions of "local" by referring to them as "non-geometrically local" and "geometrically local" in this introduction. A non-geometrically local Hamiltonian is a Hamiltonian $H(t)$ that can be written as a sum of polynomially many terms $H_j(t)$, each of which acts nontrivially on only k qubits at a time (i.e., the matrix acts as identity on all other qubits). A geometrically local Hamiltonian is similar, except that each term $H_j(t)$ must act on k adjacent qubits. Since we refer to "adjacent" qubits, the geometry of how the qubits are laid out in space must be specified. In this paper we will deal with qubits laid out in a D-dimensional lattice in Euclidean space. I.e., qubits are located at points in $\mathbb{Z}^D$ and are adjacent if they are close in Euclidean distance.

Lattice Hamiltonians in D-dimensions (with $D \leq 3$) already model all the physical systems we are interested in, are of fundamental importance by the principle of locality. From a practical perspective, this case captures a large fraction of all physical systems we are interested in.[2]

4) Prior best algorithms: To describe the known algorithms for this problem, we need to formally specify the problem. Although our results apply to very general time-dependent Hamiltonians, while comparing to previous work we assume the simpler case where the Hamiltonian is time independent.

We assume our n qubits are laid out in a D-dimensional lattice Λ in $\mathbb{R}^D$, where $D = O(1)$, and every unit ball contains $O(1)$ qubits. We assume our Hamiltonian H is given as a sum of terms $H = \sum_{X \subseteq \Lambda} h_X$, where each h_X only acts nontrivially on qubits in X (and acts as identity on the qubits in $\Lambda \setminus X$), such that $h_X = 0$ if $\mathrm{diam}(X) > 1$, which enforces geometric locality. (More formally, we rescale the metric in such a way that $\mathrm{diam}(X) > 1$ implies $h_X = 0$.) We normalize the Hamiltonian by requiring $\|h_X\| \leq 1$.

We consider a quantum circuit simulating the time evolution due to such a Hamiltonian efficient if it uses $\mathrm{poly}(n, T, 1/\epsilon)$ gates. To get some intuition for what we should hope for, notice that in the real world, time evolution takes time T and uses n qubits. Regarding "Nature" as a quantum simulator, we might expect that there is a quantum circuit that uses $O(n)$ qubits, $O(T)$ circuit depth, and $O(nT)$ total gates to solve the problem. It is also reasonable to allow logarithmic overhead in the simulation since such overheads are common even when one classical system simulates the time evolution of another (e.g., when one kind of Turing machine simulates another).

However, previous algorithms for this problem fall short of this expectation. The best Hamiltonian simulation algorithms for sparse Hamiltonians [11], [13], [14] have query complexity $O(nT \,\mathrm{polylog}(nT/\epsilon))$, but the assumed oracle for the entries requires $O(n)$ gates to implement, yielding an algorithm that uses $O(n^2 T \,\mathrm{polylog}(nT/\epsilon))$ gates. This was also observed in a recent paper of Childs, Maslov, Nam, Ross, and Su [16], who noted that for $T = n$, all the sparse Hamiltonian simulation algorithms had gate cost proportional to n^3 (or worse). A standard application of high-order Lie-Trotter-Suzuki expansions [3], [9], [17], [18] yields gate complexity $O(n^2 T (nT/\epsilon)^\delta)$ for any fixed $\delta > 0$. It has been argued [19, Sec. 4.3] that this in fact yields an algorithm with gate complexity $O(nT(nT/\epsilon)^\delta)$ for any fixed $\delta > 0$. We believe this analysis is correct, but perhaps some details need to be filled in to make the analysis rigorous. In any case, this algorithm still performs worse than desired, and in particular does not have polylogarithmic dependence on $1/\epsilon$.

[2]There are some physical situations where we do care about more general Hamiltonians. Even though the system we are given may be described by a geometrically local Hamiltonian, it is sometimes computationally advantageous to represent a given system with a non-geometrically local (or sparse) Hamiltonian.

A. Results

We exhibit a quantum algorithm that simulates the time evolution due to a time-dependent lattice Hamiltonian with a circuit that uses $O(nT\,\mathrm{polylog}(nT/\epsilon))$ geometrically local 2-qubit gates (i.e., the gates in our circuit also respect the geometry of the qubits), with depth $O(T\,\mathrm{polylog}(nT/\epsilon))$ using only $\mathrm{polylog}(nT/\epsilon)$ ancilla qubits. We then also prove a matching lower bound, showing that no quantum algorithm can do better (up to logarithmic factors), even if we relax the output requirement significantly. We now describe our results more formally.

1) Algorithmic results: We consider a more general version of the problem with time-dependent Hamiltonians. In this case we will have $H(t) = \sum_{X\subseteq\Lambda} h_X(t)$ with the locality and norm conditions as before. However, now the operators $h_X(t)$ are functions of time and we need to impose some reasonable constraints on the entries to obtain polynomial-time algorithms.

First we need to be able to compute the entries of our Hamiltonian efficiently at a given time t. We say that a function $\alpha : [0,T] \ni t \mapsto \alpha(t) \in \mathbb{R}$ is *efficiently computable* if there is an algorithm that outputs $\alpha(t)$ to precision ϵ for any given input t specified to precision ϵ in running time $\mathrm{polylog}(T/\epsilon)$. Note that any complex-valued analytic function on a nonzero neighborhood of a closed real interval in the complex plane is efficiently computable (see the full version of this paper for a proof). We will assume that each entry in a local term $h_X(t)$ is efficiently computable.

In addition to being able to compute the entries of the Hamiltonian, we require that the entries do not change wildly with time; otherwise, a sample of entries at discrete times may not predict the behavior of the entries at other times. We say that a function α on the interval $[0,T]$ ($T \geq 1$) is *piecewise slowly varying* if there are $M = O(T)$ intervals $[t_{j-1}, t_j]$ with $0 = t_0 < t_1 < \cdots < t_M = T$ such that $\frac{\mathrm{d}}{\mathrm{d}t}\alpha(t)$ exists and is bounded by $1/(t_j - t_{j-1})$ for $t \in (t_{j-1}, t_j)$. In particular, a function is piecewise slowly varying if it is a sum of $O(T)$ pieces, each of which has derivative at most $O(1)$. We will assume that each entry in a term $h_X(t)$ is piecewise slowly varying.

We are now ready to state our main result, which is proved in Section II

Theorem 1. *Let $H(t) = \sum_{X\subseteq\Lambda} h_X(t)$ be a time-dependent Hamiltonian on a lattice Λ of n qubits, embedded in the Euclidean metric space $\mathbb{R}^D$. Assume that every unit ball contains $O(1)$ qubits and $h_X = 0$ if $\mathrm{diam}(X) > 1$. Also assume that every local term $h_X(t)$ is efficiently computable (e.g., analytic), piecewise slowly varying on time domain $[0,T]$, and has $\|h_X(t)\| \leq 1$ for any X and t. Then, there exists a quantum algorithm that can approximate the time evolution of H for time T to accuracy ϵ using $O(Tn\,\mathrm{polylog}(Tn/\epsilon))$ 2-qubit local gates, and has depth $O(T\,\mathrm{polylog}(Tn/\epsilon))$.*

Our algorithm uses $O(1)$ ancillas per system qubit on which H is defined. The ancillas are interspersed with the system qubits, and all the gates respect the locality of the lattice.

2) Lower bounds: We also prove a lower bound on the gate complexity of problem of simulating the time evolution of a time-dependent lattice Hamiltonian. This lower bound matches, up to logarithmic factors, the gate complexity of the algorithm presented in Theorem 1. Note that unlike previous lower bounds on Hamiltonian simulation [9], [10], [12], which prove lower bounds on query complexity, this is a lower bound on the number of gates required to approximately implement the time-evolution unitary. To our best knowledge, this is the first nontrivial lower bound on the gate complexity of the simulation problem. For concreteness, we focus on a 1-dimensional time-dependent local Hamiltonian in this section, although the lower bound extends to other constant dimensions with minor modifications. The lower bounds are proved in Section III.

Before stating the result formally, let us precisely define the class of Hamiltonians for which we prove the lower bound. We say a Hamiltonian $H(t)$ acting on n qubits is a "piecewise constant 1D Hamiltonian" if $H(t) = \sum_{j=1}^{n-1} H_j(t)$, where $H_j(t)$ is only supported on qubits j and $j+1$ with $\max_t \|H_j(t)\| = O(1)$, and there is a time slicing $0 = t_0 < t_1 < \cdots < t_M = T$ where $t_m - t_{m-1} \leq 1$ and $M = O(T)$ such that $H(t)$ is time-independent within each time slice.

For such Hamiltonians, the time evolution operator for time T can be simulated with error at most ϵ using Theorem 1 with $O(Tn\,\mathrm{polylog}(Tn/\epsilon))$ 2-qubit local gates (i.e., the 2-qubit gates only act on adjacent qubits). In particular, for any constant error, the simulation only requires $\widetilde{O}(Tn)$ 2-qubit local gates. We prove a matching lower bound, where the lower bound even holds against circuits that may use non-geometrically local (i.e., acting on non-adjacent qubits) 2-qubit gates from a possibly infinite gate set and unlimited ancilla qubits.

Theorem 2. *For any integers n and $T \leq 4^n$, there exists a piecewise constant bounded 1D Hamil-tonian $H(t)$ on n qubits, such that any quantum circuit that approximates the time evolution due to $H(t)$ for time T to constant error must use $\widetilde{\Omega}(Tn)$ 2-qubit gates. The quantum circuit may use unlimited ancilla qubits and the gates may be non-local and come from a possibly infinite gate set.*

Note that this lower bound only holds for $T \leq 4^n$, because any unitary on n qubits can be implemented with $\widetilde{O}(4^n)$ 2-qubit local gates [20], [21].

We can also strengthen our lower bound to work in the situation where we are only interested in measuring a local observable at the end of the simulation. The simulation algorithm presented in Theorem 1 provides a strong guarantee: the output state is ϵ-close to the ideal output state in trace

distance. Trace distance captures distinguishability with respect to arbitrary measurements, but for some applications it might be sufficient for the output state to be close to the ideal state with respect to local measurements only. We show that even in this limited measurement setting, it is not possible to speed up our algorithm in general. In fact, our lower bound works even if the only local measurement performed is a computational basis measurement on the first output qubit.

Theorem 3. *For any integers n and T such that $1 \leq n \leq T \leq 2^n$, there exists a piecewise constant bounded 1D Hamiltonian $H(t)$ on n qubits, such that any quantum circuit that approximates the time evolution due to $H(t)$ for time T to constant error on any local observable must use $\widetilde{\Omega}(Tn)$ 2-qubit gates. If $T \leq n$, we have a lower bound of $\widetilde{\Omega}(T^2)$ gates. (The quantum circuit may use unlimited ancilla qubits and the gates may be non-local and come from a possibly infinite gate set.)*

Note that the fact that we get a weaker lower bound of $\widetilde{\Omega}(T^2)$ when $T \leq n$ is not a limitation, but reflects the fact that small time evolutions are actually easier to simulate when the measurement is local. To see this, consider first simulating the time evolution using the algorithm in Theorem 1. This yields a circuit with $\widetilde{O}(Tn)$ 2-qubit local gates. But if we only want the output of a local measurement after time T, qubits that are far away from the measured qubits cannot affect the output, since the circuit only consists of 2-qubit local gates. Hence we can simply remove all gates that are more than distance equal to the depth of the circuit, $\widetilde{O}(T)$, away from the measured qubits. We are then left with a circuit that uses $\widetilde{O}(T^2)$ gates, matching the lower bound in Theorem 3.

3) Other results: In the full version of this paper, we perform concrete resource estimation on our algorithm against a Hamiltonian that has been considered in the past as a benchmark for such algorithms. Not only is our algorithm asymptotically superior to prior algorithms, but our numerical estimates show that it quickly outperforms prior algorithms.

In the full version we also describe two extensions of our main algorithmic result. First we show how our algorithm can work well even when the Hamiltonian has some terms with large norm without paying a cost proportional to the largest norm of any term in the Hamiltonian. Our second result shows how to improve the dimension dependence of the algorithm in Section II from exponential in the dimension to polynomial in the dimension.

Finally, in the full version we also prove an improved Lieb-Robinson bound tailored to Hamiltonians with small commutators between local terms. This bound yields zero Lieb-Robinson velocity in the limit of commuting Hamiltonians. This improves the performance of our algorithm when the Hamiltonian is close to commuting.

B. Techniques

1) Algorithm: Our algorithm is based on a decomposition of the time evolution unitary using Lieb-Robinson bounds [22]–[26], that was made explicit by Osborne [27] (see also Michalakis [28, Sec. III]), which when combined with recent advances in Hamiltonian simulation [11], [13], [14], yields Theorem 1.

Lieb-Robinson bounds are theorems that informally state that information travels at a constant speed in geometrically local Hamiltonians. For intuition, consider a 1-dimensional lattice of qubits and a geometrically local Hamiltonian that is evolved for a short amount of time. If the time is too short, no information about the first qubit can be transmitted to the last qubit. Lieb-Robinson bounds make this intuition precise, and show that the qubit at position n is only affected by the qubits and operators at position 1 after time $\Omega(n)$. Note that if this were a small-depth unitary circuit of geometrically local 2-qubit gates such a statement would follow using a "lightcone" argument. In other words, after one layer of geometrically local 2-qubit gates, the influence of qubit 1 can only have spread to qubit 2. Similarly, after k layers of 2-qubit gates, the influence of qubit 1 can only have spread up to qubit k. The fact that this extends to geometrically local Hamiltonians is nontrivial, and is only approximately true. See Lemma 5 for a formal statement of a Lieb-Robinson bound.

We use these ideas to chop up the large unitary that performs time evolution for the full Hamiltonian H into many smaller unitaries that perform time evolution for a small portion of the Hamiltonian. Quantitatively, we break Hamiltonian simulation for H for time $O(1)$ into $O(n/\log(nT/\epsilon))$ pieces, each of which is a Hamiltonian simulation problem for a Hamiltonian on an instance of size $O(\log(nT/\epsilon))$ to exponentially small error. At this point we can use any Hamiltonian simulation algorithm for the smaller piece as long as it has polynomial gate cost and has exponentially good dependence on ϵ. While Hamiltonian simulation algorithms based on product formulas do not have error dependence that is $\mathrm{polylog}(1/\epsilon)$, recent Hamiltonian simulation algorithms, such as [10]–[14] have $\mathrm{polylog}(1/\epsilon)$ scaling. Thus our result crucially uses the recent advances in Hamiltonian simulation with improved error scaling.

2) Lower bound: As noted before, we lower bound the gate complexity (or total number of gates) required for Hamiltonian simulation, which is different from prior work which proved lower bounds on the query complexity of Hamiltonian simulation. As such, our techniques are completely different from those used in prior work. Informally, our lower bounds are based on a refined circuit-size hierarchy theorem for quantum circuits, although we are technically comparing two different resources in two different models, which are simulation time for Hamiltonians versus gate cost for circuits.

As a simple motivating example, consider circuit-size hierarchy theorems for classical or quantum circuits more generally. Abstractly, a hierarchy theorem generally states that a computational model with X amount of a resource (e.g., time, space, gates) can do more if given more of the same resource. For example, it can be shown that for every $G \ll 2^n/n$, there exists a Boolean function on n bits that cannot be computed by a circuit of size G, but can be computed by a circuit of size $G + O(n)$. We show similar hierarchy theorems for quantum circuit size, except that we show that the circuit of larger size that computes the function actually comes from a weaker family of circuits. Informally, we are able to show that there are functions that can be computed by a larger circuit that uses only geometrically local 2-qubit gates from a fixed universal gate set that cannot be a computed by a smaller circuit, even if we allow the smaller circuit access to unlimited ancilla bits and non-geometrically local 2-qubit from an infinite gate set. We then leverage this asymmetric circuit size hierarchy theorem to show that there is a Hamiltonian whose evolution for time T cannot be simulated by a circuit of size $\ll nT$, by embedding the result of any quantum circuit with geometrically local 2-qubit gates into a piecewise constant Hamiltonian with time proportional to the depth of the circuit.

II. Algorithm and analysis

In this section we establish our main algorithmic result, restated below for convenience:

Theorem 1. *Let $H(t) = \sum_{X \subseteq \Lambda} h_X(t)$ be a time-dependent Hamiltonian on a lattice Λ of n qubits, embedded in the Euclidean metric space $\mathbb{R}^D$. Assume that every unit ball contains $O(1)$ qubits and $h_X = 0$ if $\operatorname{diam}(X) > 1$. Also assume that every local term $h_X(t)$ is efficiently computable (e.g., analytic), piecewise slowly varying on time domain $[0, T]$, and has $\|h_X(t)\| \leq 1$ for any X and t. Then, there exists a quantum algorithm that can approximate the time evolution of H for time T to accuracy ϵ using $O(Tn \operatorname{polylog}(Tn/\epsilon))$ 2-qubit local gates, and has depth $O(T \operatorname{polylog}(Tn/\epsilon))$.*

The algorithm is depicted in Figure 1. Before showing why this algorithm works, we provide a high-level overview of the algorithm and the structure of the proof. Since a time evolution unitary $U(T;0)$ is equal to $U(T = t_M; t_{M-1})U(t_{M-1}, t_{M-2}) \cdots U(t_2; t_1)U(t_1; t_0 = 0)$, we will focus on a time evolution operator $U(t;0)$ where $t = O(1)$, generated by a slowly varying bounded Hamiltonian. The key idea, as shown in Figure 1, is that the time-evolution operator, e^{-itH} due to the full Hamiltonian $\sum_{X \subseteq \Lambda} h_X$ can be approximately written as a product

$$e^{-itH} \approx \left(e^{-itH_A}\right)\left(e^{+itH_Y}\right)\left(e^{-itH_{Y \cup B}}\right). \tag{2}$$

Here $A \cup B = \Lambda$ and we think of A and B as large regions, but Y as a small subset of A. The error in the approximation is exponentially small as long as Y is large enough. This is formally proved in Lemma 6, which is supported by Lemma 4 and Lemma 5. Applying this twice, using

$$e^{-itH_{Y \cup B}} \approx \left(e^{-itH_B}\right)\left(e^{+itH_Z}\right)\left(e^{-itH_{Y \cup Z}}\right) \tag{3}$$

leads to a symmetric approximation as depicted at the bottom left of Figure 1. This procedure can then be repeated for the large operators supported on A and B to reduce the size of all the operators involved, leading to the pattern in Figure 1 (a). This reduces the problem of implementing the time-evolution operator for H into the problem of implementing smaller time-evolution operators, which can be implemented using known quantum algorithms. We now establish the lemmas needed to prove the result.

Lemma 4. *Let A_t and B_t be continuous time-dependent Hermitian operators, and let U_t^A and U_t^B with $U_0^A = U_0^B = \mathbf{1}$ be the corresponding time evolution unitaries. Then the following hold:*

(i) *$W_t = (U_t^B)^\dagger U_t^A$ is the unique solution of $i\partial_t W_t = \left((U_t^B)^\dagger (A_t - B_t) U_t^B\right) W_t$ and $W_0 = \mathbf{1}$.*

(ii) *If $\|A_s - B_s\| \leq \delta$ for all $s \in [0, t]$, then $\left\|U_t^A - U_t^B\right\| \leq t\delta$.*

Proof: (i) Differentiate. The solution to the ordinary differential equation is unique. (ii) Apply Jensen's inequality for $\|\cdot\|$ (implied by the triangle inequality for $\|\cdot\|$) to the equation $W_t - W_0 = \int_0^t \mathrm{d}s \partial_s W_s$. Then, invoke (i) and the unitary invariance of $\|\cdot\|$. ∎

The next lemma is proved in the full version of this paper.

Lemma 5 (Lieb-Robinson bound [22]–[25])**.** *Let $H = \sum_X h_X$ be a local Hamiltonian and O_X be any operator supported on X, and put $\ell = \lfloor \operatorname{dist}(X, \Lambda \setminus \Omega) \rfloor$. Then*

$$\left\|(U_t^H)^\dagger O_X U_t^H - (U_t^{H_\Omega})^\dagger O_X U_t^{H_\Omega}\right\| \leq |X| \, \|O_X\| \, \frac{(2\zeta_0 |t|)^\ell}{\ell!}, \tag{4}$$

where $\zeta_0 = \max_{p \in \Lambda} \sum_{Z \ni p} |Z| \, \|h_Z\| = O(1)$. In particular, there are constants $v_{LR} > 0$, called the Lieb-Robinson velocity,[3] and $\mu > 0$, such that for $\ell \geq v_{LR}|t|$, we have

$$\left\|(U_t^H)^\dagger O_X U_t^H - (U_t^{H_\Omega})^\dagger O_X U_t^{H_\Omega}\right\| \leq O(|X| \, \|O_X\| \exp(-\mu\ell)). \tag{5}$$

We are considering strictly local interactions (as in Theorem 1), where $h_X = 0$ if $\operatorname{diam}(X) > 1$, but similar results hold with milder locality conditions such as $\|h_X\| \leq e^{-\operatorname{diam}(X)}$ [22]–[26]; see the appendix for a detailed proof. Below we will only use the result that the error is at most

[3] Strictly speaking, the Lieb-Robinson velocity is defined to be the infimum of any v_{LR} such that Eq. (5) holds.

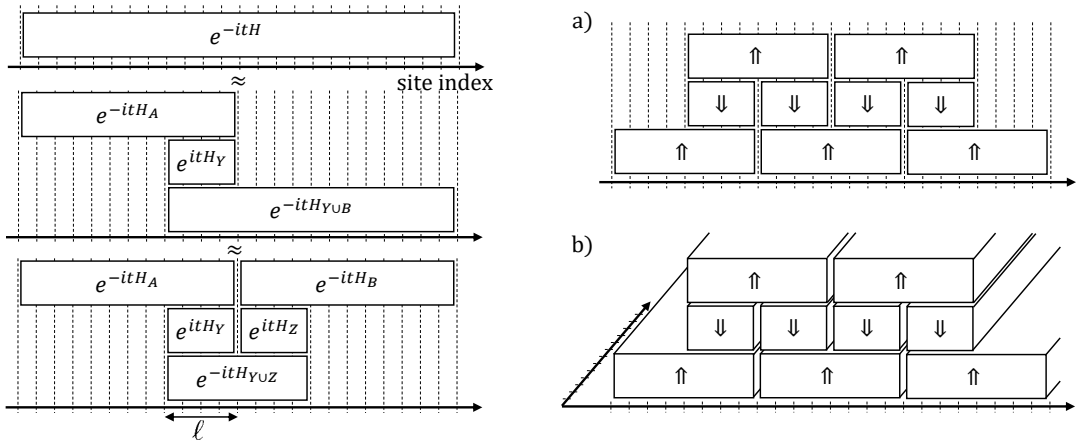

Figure 1. Decomposition of time evolution operator for time $t = O(1)$. The time is going upwards. Each block $\square$ represents the forward time evolution, $e^{-itH_\square}$, if the arrow is upward, and backward time evolution, $e^{+itH_\square}$, if the arrow is downward. Here, $H_\square$ is the sum of local terms in the Hamiltonian supported within the block. The overlap has size ℓ. (a) shows a one-dimensional setting, but a generalization to higher D dimensions is readily achieved by regarding each block as a $(D-1)$-dimensional hyperplane so that the problem reduces to lower dimensions. (b) shows a two-dimensional setting. The approximation error from the depicted decomposition is $\epsilon = O(e^{-\mu\ell}L^D/\ell)$ where L is the linear system size, ℓ is the width of the overlap between blocks, and $\mu > 0$ is a constant that depends only on the locality of the Hamiltonian. One can use any algorithm to further decompose the resulting "small" unitaries on $O(\log(L/\epsilon))$ qubits into elementary gates. To achieve gate count that is linear (up to logarithmic factors) in spacetime volume, the algorithm for simulating the blocks needs to be polynomial in the block size and polylogarithmic in accuracy.

$O(e^{-\mu\ell})$ for some $\mu > 0$ and fixed t. For slower decaying interactions, the bound is weaker and the overlap size ℓ in Figure 1 will have to be larger.

The Lieb-Robinson bound implies the following decomposition.

Lemma 6. *Let $H = \sum_X h_X$ be a local Hamiltonian (as in Theorem 1, or a more general definition for which Lemma 5 still holds). Then there are constants $v, \mu > 0$ such that for any disjoint regions A, B, C, we have*

$$\left\| U_t^{H_{A\cup B}} (U_t^{H_B})^\dagger U_t^{H_{B\cup C}} - U_t^{H_{A\cup B\cup C}} \right\| \leq O(e^{vt-\mu\,\mathrm{dist}(A,C)}) \sum_{X:bd(AB,C)} \|h_X\| \tag{6}$$

where $X : bd(AB, C)$ means that $X \subseteq A \cup B \cup C$ and $X \not\subseteq A \cup B$ and $X \not\subseteq C$.

Proof: We omit "$\cup$" for the union of disjoint sets. The following identity is trivial but important:

$$U_t^{H_{ABC}} = U_t^{H_{AB}+H_C} \underbrace{(U_t^{H_{AB}+H_C})^\dagger U_t^{H_{ABC}}}_{=W_t}. \tag{7}$$

By Lemma 4 (i), W_t is generated by [27], [28]

$$(U_t^{H_{AB}+H_C})^\dagger (\underbrace{H_{ABC} - H_{AB} - H_C}_{H_{\mathrm{bd}}}) U_t^{H_{AB}+H_C} = (U_t^{H_B+H_C})^\dagger H_{\mathrm{bd}} U_t^{H_B+H_C} + \underbrace{O(\|H_{\mathrm{bd}}\|\, e^{vt-\mu\ell})}_{=\delta} \tag{8}$$

where ℓ is the distance from the support of the boundary terms H_{bd} to A, and the estimate of δ is due to Lemma 5 applied to local terms in H_{bd}. Since H_{bd} contains terms that cross between AB and C, the distance ℓ is at least $\mathrm{dist}(A, C)$ minus 2. By Lemma 4 (i) again, the unitary generated by the first term of (8) is $(U_t^{H_B+H_C})^\dagger U_t^{H_{BC}}$, which can be thought of as the "interaction picture" time-evolution operator of the Hamiltonian in (8). This is our simplification of the "patching" unitary, which is $t\delta$-close to W_t by Lemma 4 (ii). ∎

Proof of Theorem 1: The circuit for simulating the Hamiltonian is described in Figure 1. The decomposition of time evolution unitary in Figure 1 is a result of iterated application of Lemma 6. For a one-dimensional chain, let L be the length of the chain, so there are $O(L)$ qubits. Take a two contiguous blocks Y and Z of the chain that overlaps by length $\ell \ll L$. Under periodic boundary conditions there are two components in the intersection, and under open boundary conditions, there is one component in the intersection. Applying Lemma 6, we decompose the full unitary into two blocks on Y and Z, respectively, and one or two blocks in the intersection. Every block unitary in the decomposition is a time evolution operator with respect to the sum of Hamitonian terms within the block, and we can recursively apply the decomposition. Making the final blocks as small as possible, we end up with a layout of small unitaries as shown in Figure 1 (a). The error from this decomposition is $O(\delta L/\ell)$, which is exponentially small in ℓ for $t = O(1)$.

Going to higher dimensions $D > 1$, we first decompose the full time evolution into unitaries on $O(L/\ell)$ hyperplanes (codimension 1). This entails error $O(e^{-\mu\ell}L^D/\ell)$ since the boundary term has norm at most $O(L^{D-1})$. For each hyperplane the decomposition into $O(L/\ell)$ blocks of codimension

2 gives error $O(e^{-\mu\ell}(\ell L^{D-2})(L/\ell))$. Summing up all the hyperplanes, we get $O(e^{-\mu\ell}L^D/\ell)$ for the second round of decomposition. After D rounds of the decomposition the total error is $O(e^{-\mu\ell}DL^D/\ell)$, and we are left with $O((L/\ell)^D)$ blocks of unitaries for $t = O(1)$. For longer times, apply the decomposition to each factor of $U(T = t_M; t_{M-1})U(t_{M-1}, t_{M-2}) \cdots U(t_2; t_1)U(t_1; t_0 = 0)$

It remains to implement the unitaries on $m = O(TL^D/\ell^D)$ blocks $\square$ of $O(\ell^D)$ qubits where $\ell = O(\log(TL/\epsilon))$ to accuracy ϵ/m. All blocks have the form $U_t^{H_\square}$, and can be implemented using any known Hamiltonian simulation algorithm. For a time-independent Hamiltonian, if we use an algorithm that is polynomial in the spacetime volume and polylogarithmic in the accuracy such as those based on signal processing [13], [14] or linear combination of unitaries [10]–[12], then the overall gate complexity is $O(TL^D \operatorname{polylog}(TL/\epsilon)) = O(Tn \operatorname{polylog}(Tn/\epsilon))$, where the exponent in the polylog factor depends on the choice of the algorithm.[4] For a slowly varying time-dependent Hamiltonian, we can use the fractional queries algorithm [10] or the Taylor series approach [11], [29], [30] to achieve the same gate complexity. The Taylor series approach uses a subroutine $|t\rangle \mapsto |t\rangle \left(\sum_j |\alpha_j(t)|\right)^{-1/2} \sum_j \sqrt{\alpha_j(t)}\,|j\rangle$, where $\alpha_j(t)$ is the real coefficient of Pauli operator P_j in the Hamiltonian, which must be efficiently evaluated. ∎

For not too large system sizes L, it may be reasonable to use a bruteforce method to decompose the block unitaries into elementary gates [31, Chap. 8].

III. Optimality

In this section we prove a lower bound on the gate complexity of problem of simulating the time evolution of a time-dependent local Hamiltonian. (Recall that throughout this paper we use *local* to mean geometrically local.) The results are restated below for the reader's convenience.

Theorem 2. *For any integers n and $T \leq 4^n$, there exists a piecewise constant bounded 1D Hamil-tonian $H(t)$ on n qubits, such that any quantum circuit that approximates the time evolution due to $H(t)$ for time T to constant error must use $\widetilde{\Omega}(Tn)$ 2-qubit gates. The quantum circuit may use unlimited ancilla qubits and the gates may be non-local and come from a possibly infinite gate set.*

Theorem 3. *For any integers n and T such that $1 \leq n \leq T \leq 2^n$, there exists a piecewise constant bounded 1D Hamiltonian $H(t)$ on n qubits, such that any quantum circuit that approximates the time evolution due to $H(t)$ for time T to constant error on any local observable must use $\widetilde{\Omega}(Tn)$ 2-qubit gates. If $T \leq n$, we have a lower bound of $\widetilde{\Omega}(T^2)$ gates. (The quantum circuit may use unlimited ancilla qubits and the gates may be non-local and come from a possibly infinite gate set.)*

[4] If we use the quantum signal processing based algorithms [13], [14] to implement the blocks of size $O(\ell^D)$, then we need $O(\log \ell)$ ancilla qubits for a block. Thus, if we do not mind implementing them all in serial, then it follows that the number of ancillas needed is $O(\log \log(Tn/\epsilon))$, which is much smaller than what would be needed if the quantum signal processing algorithm was directly used to simulate the full system.

A. Lower bound proofs

We now prove Theorem 2 and Theorem 3, starting with Theorem 3. This lower bound follows from the following three steps. First, in Lemma 7, we observe that for every depth-T quantum circuit on n qubits that uses local 2-qubit gates, there exists a Hamiltonian $H(t)$ of the above form such that time evolution due to $H(t)$ for time T is equal to applying the quantum circuit. Then, in Lemma 8 we show that the number of distinct Boolean functions on n bits computed by such quantum circuits is at least exponential in $\widetilde{\Omega}(Tn)$, where we say a quantum circuit has computed a Boolean function if its first output qubit is equal to the value of the Boolean function with high probability. Finally, in Lemma 9 we observe that the maximum number of Boolean functions that can be computed (to constant error) by the class of quantum circuits with G arbitrary non-local 2-qubit gates from any (possibly infinite) gate set is exponential in $\widetilde{O}(G \log n)$. Since we want this class of circuits to be able to simulate all piecewise constant bounded 1D Hamiltonians for time T, we must have $G = \widetilde{\Omega}(Tn)$.

Lemma 7. *Let U be a depth-T quantum circuit on n qubits that uses local 2-qubit gates from any gate set. Then there exists a piecewise constant bounded 1D Hamiltonian $H(t)$ such that the time evolution due to $H(t)$ for time T exactly equals U.*

Proof: We first prove the claim for a depth-1 quantum circuit. This yields a Hamiltonian $H(t)$ that is defined for $t \in [0, 1]$, whose time evolution for unit time equals the given depth-1 circuit. Then we can apply the same argument to each layer of the circuit, obtaining Hamiltonians valid for times $t \in [1, 2]$, and so on, until $t \in [T - 1, T]$. This yields a Hamiltonian $H(t)$ defined for all time $t \in [0, T]$ whose time evolution for time T equals the given unitary. If the individual terms in a given time interval have bounded spectral norm, then so will the Hamiltonian defined for the full time duration.

For a depth 1 circuit with local 2-qubit gates, since the gates act on disjoint qubits we only need to solve the problem for one 2-qubit unitary and sum the resulting Hamiltonians. Consider a unitary U_j that acts on qubits j and $j + 1$. By choosing $H_j = i \log U_j$, we can ensure that $e^{-iH_j} = U_j$ and $\|H_j\| = O(1)$.

The overall Hamiltonian is now piecewise constant with T time slices. ∎

Note that it also possible to use a similar construction to obtain a Hamiltonian that is continuous (instead of being

piecewise constant) with a constant upper bound on the norm of the first derivative of the Hamiltonian.

Lemma 8. *For any integers n and T such that $1 \leq n \leq T \leq 2^n$, the number of distinct Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that can be computed by depth-T quantum circuits on n qubits that use local 2-qubit gates from a finite gate set is at least $2^{\widetilde{\Omega}(Tn)}$.*

Proof: We first divide the n qubits into groups of $k = \log_2 T$ qubits, which is possible since $T \leq 2^n$. On these k qubits, we will show that it is possible to compute $2^{\widetilde{\Omega}(T)}$ distinct Boolean functions with a depth T circuit that uses local 2-qubit gates. One way to do this is to consider all Boolean functions on $k' < k$ bits. Any Boolean function f_x that evaluates to $f_x(x) = 1$ on exactly one input x of size k' can be computed with a circuit of $\widetilde{O}(k')$ gates and $\widetilde{O}(k')$ depth using only 2-qubit gates and 1 ancilla qubit in addition to one output qubit [20, Corollary 7.4]. An arbitrary Boolean function $f : \{0,1\}^{k'} \to \{0,1\}$ is a sum of such functions: $f = \sum_{x \in f^{-1}(1)} f_x = \bigoplus_{x \in f^{-1}(1)} f_x$. Implementing all f_x for $x \in f^{-1}(1)$ in serial using a common output qubit, we obtain a circuit for the full function f. Since $f^{-1}(1)$ consists at most $2^{k'}$ bit strings, this will yield a circuit of size $\widetilde{O}(2^{k'})$ and depth $\widetilde{O}(2^{k'})$. Making all the gates local using SWAP gates does not change these expressions by more than a log factor in the exponent. By choosing $k' = k - \Theta(\log k)$, we can compute all Boolean functions on k' bits with depth at most T. Since there are $2^{2^{k'}} = 2^{\widetilde{\Omega}(T)}$ distinct Boolean functions on k' bits, we have shown that circuits with depth T using $k = \log_2 T$ qubits can compute at least $2^{\widetilde{\Omega}(T)}$ distinct Boolean functions.

We can compute $2^{\widetilde{\Omega}(T)}$ distinct Boolean functions on each of the n/k blocks of k qubits to obtain $(2^{\widetilde{\Omega}(T)})^{n/k} = 2^{\widetilde{\Omega}(Tn)}$ distinct Boolean functions with n/k outputs. I.e., we have computed a function $\{0,1\}^n \to \{0,1\}^{n/k}$. Since we want to obtain a single-output Boolean function, as the overall goal is to prove lower bounds against simulation algorithms correct on local measurements, we combine these Boolean functions into one. We do this by computing the parity of all the outputs of these n/k Boolean functions using CNOT gates. Computing the parity uses at most n 2-qubit local gates and has depth n. The circuit now has depth $T + n \leq 2T$ and by rescaling T we can make this circuit have depth T, while retaining the lower bound of $2^{\widetilde{\Omega}(Tn)}$ distinct Boolean functions.

Unfortunately, after taking the parity of these n/k functions, it is not true that the resulting functions are all distinct. For example, the parity of functions $f(x)$ and $g(y)$ is a new function $f(x) \oplus g(y)$, which also happens to be the parity of the functions $\neg f(x)$ and $\neg g(y)$. To avoid this overcounting of functions, we do not use all possible functions on k' bits in the argument above, but only all those functions that map the all-zeros input to 0. This only halves the total number of functions, which does not change the asymptotic expressions above. With this additional constraint, it is easy to see that if $f(x) \oplus g(y) = f'(x) \oplus g'(y)$, this implies that f and f' are the same, by fixing y to be the all-zeros input, and similarly that g and g' are the same. ∎

We say a quantum circuit U computes a Boolean function $f : \{0,1\}^n \to \{0,1\}$ with high probability if measuring the first output qubit of $U|x_1 x_2 \cdots x_n 0 \cdots 0\rangle$ yields $f(x)$ with probability at least $2/3$.

Lemma 9. *The number of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that can be computed with high probability by quantum circuits with unlimited ancilla qubits using G non-local 2-qubit gates from any gate set is at most $2^{\widetilde{O}(G \log n)}$.*

Proof: First we note that if a circuit U with G arbitrary 2-qubit gates from any gate set computes a Boolean function with probability at least $2/3$, then there is another circuit with $\widetilde{O}(G)$ gates from a finite 2-qubit non-local gate set that computes the same function with probability at least $2/3$. We do this by first boosting the success probability of the original circuit using an ancilla qubit to a constant larger than $2/3$ and then invoking the Solovay–Kitaev theorem [32] to approximate each gate in this circuit to error $O(1/G)$ with a circuit from a finite gate set of 2-qubit gates. This increases the circuit size to $\widetilde{O}(G)$ gates. Since each gate has error $O(1/G)$, the overall error is only a constant, and the new circuit approximates computes the Boolean function f with high probability.

We now have to show that the number of Boolean functions on n bits computed by a circuit with $\widetilde{O}(G)$ non-local 2-qubit gates from a finite gate set is at most $2^{\widetilde{O}(G \log n)}$. To do so, we simply show that the total number of distinct circuits with $\widetilde{O}(G)$ non-local 2-qubit gates from a finite gate set is at most $2^{\widetilde{O}(G \log n)}$.

First observe that a circuit with $\widetilde{O}(G)$ gates can only use $\widetilde{O}(G)$ ancilla qubits, since each 2-qubit gate can interact with at most 2 new ancilla qubits. Furthermore, the depth of a circuit cannot be larger than the number of gates in the circuit. Let us now upper bound the total number of quantum circuits of this form. Each such circuit can be specified by listing the location of each gate and which gate it is from the finite gate set. Specifiying the latter only needs a constant number of bits since the gate set is finite, and the location can be specified using the gate's depth, and the labels of the two qubits it acts on. The depth only requires $O(\log G)$ bits to specify, and since there are at most $n + \widetilde{O}(G)$ qubits, this only needs $O(\log n + \log G)$ bits to specify. In total, since there are $\widetilde{O}(G)$ gates, the entire circuit can be specified with $\widetilde{O}(G \log n)$ bits. Finally, since any such circuit can be specified with $\widetilde{O}(G \log n)$ bits, there can only be $2^{\widetilde{O}(G \log n)}$ such circuits. ∎

Proof of Theorem 3: Suppose that any piecewise constant bounded 1D Hamiltonian on n qubits can be

simulated for time T using G 2-qubit non-local gates from any (possibly infinite) gate set. Then using Lemma 7 and Lemma 8, we can compute at least $2^{\widetilde{\Omega}(Tn)}$ distinct Boolean functions using such Hamiltonians. By assumption, each of these Boolean functions can be approximately computed by a circuit with G gates. Now invoking Lemma 9, we know that such circuits can compute at most $2^{\widetilde{O}(G \log n)}$ n-bit Boolean functions. Hence we must have $G \log n = \widetilde{\Omega}(Tn)$, which yields $G = \widetilde{\Omega}(Tn)$.

The proof for $T \leq n$ follows in a black-box manner from the first part of the theorem statement by only using T out of the n available qubits. In this case the first part of the theorem applies and yields a lower bound of $\widetilde{\Omega}(T^2)$. ■

We now prove Theorem 2, which follows a similar outline. The first step is identical, and we can reuse Lemma 7. For the next step, instead of counting distinct Boolean functions, we count the total number of "distinct" unitaries. Unlike Boolean functions on n bits, there are infinitely many unitaries on n qubits. Hence we count unitaries that are "distinguishable." Formally, we say U and V are distinguishable if they are some constant, say 0.1 for concreteness, apart in diamond norm. This is equivalent to the existence of a state $|\psi\rangle$ such that $U|\psi\rangle$ and $V|\psi\rangle$ have trace distance 0.1. In Lemma 10 we show that the number of distinguishable unitaries computed by quantum circuits on n qubits with depth T is exponential in $\widetilde{\Omega}(Tn)$. As before, we then show that the maximum number of distinguishable unitaries that can be computed (to constant error) by the class of quantum circuits with G arbitrary non-local 2-qubit gates from any (possibly infinite) gate set is exponential in $\widetilde{O}(G \log n)$.

Lemma 10. *For any integer n and integer $n \leq T \leq 4^n$, there exists a set of unitaries of size $2^{\widetilde{\Omega}(Tn)}$, such that every unitary in the set can be computed by a depth-T quantum circuit on n qubits that uses local 2-qubit gates from a finite gate set, and the diamond norm distance between any $U \neq V$ from this set is at least 0.1.*

Proof: We divide the n qubits into groups of $k = \log_4 T$ qubits, which is possible since $T \leq 4^n$. On these k qubits, we will compute $2^{\widetilde{\Omega}(T)}$ distinguishable unitaries (i.e., unitaries that are at least distance 0.1 apart from each other) with a depth T circuit that uses local 2-qubit gates. We can do this by considering a maximal set of unitaries on k' qubits that is distinguishable. More precisely, on k' qubits there exist $2^{\Omega(4^{k'})}$ unitaries such that each pair of unitaries is at least distance 0.1 apart; see e.g. [33]. (This follows from the fact that in the group of $d \times d$ unitaries with metric induced by operator norm, a ball of radius 0.1 has volume exponentially small in d^2.) We know that any unitary on k' qubits can be exactly written as a product of $\widetilde{O}(4^{k'})$ arbitrary 2-qubit gates [20], [21]. Making these gates local and from a finite gate set only adds polynomial factors in k'. By choosing $k' = k - \Theta(\log k)$, we can compute this set of $2^{\Omega}(4^{k'}) = 2^{\widetilde{\Omega}(T)}$ distinguishable unitaries with depth at most T.

Since we can compute $2^{\widetilde{\Omega}(T)}$ pairwise-distant unitaries on each of the n/k blocks of k qubits, we can compute $(2^{\widetilde{\Omega}(T)})^{n/k} = 2^{\widetilde{\Omega}(Tn)}$ unitaries on all n qubits by considering all possible combinations of unitaries on the different blocks. Finally, if U and V are distinguishable, then so are $U \otimes X$ and $V \otimes Y$, since the distinguisher can simply ignore the second register. ■

Lemma 11. *Let S be a set of distinguishable unitaries (i.e., the diamond norm distance between any $U \neq V$ from this set is at least 0.1). Then if any unitary in S can be computed by a quantum circuit with G non-local 2-qubit gates from any gate set, then $|S| = 2^{\widetilde{O}(G \log n)}$.*

Proof: This proof is essentially the same as that of Lemma 9. We first observe that if a circuit over an arbitrary gate set computes a unitary U, we can approximate it to error less than 0.04 using the Solovay–Kitaev theorem and increase the circuit size to $\widetilde{O}(G)$ gates. Importantly, since the unitaries are a distance 0.1 apart, one circuit cannot simultaneously approximate two different unitaries to error 0.04. Then exactly the same counting argument as in Lemma 9 shows there can only be $2^{\widetilde{O}(G \log n)}$ such circuits. ■

Proof of Theorem 2: Suppose that any piecewise constant bounded local Hamiltonian on n-qubits could be simulated for time T using G 2-qubit non-local gates from any (possibly infinite) gate set. Then using Lemma 7 and Lemma 10, we can produce a set S of distinguishable unitaries of size $2^{\widetilde{\Omega}(Tn)}$. By assumption, each of these unitaries can be approximately computed by a circuit with G non-local 4-qubit gates. Now invoking Lemma 9, we know that such circuits can approximate at most $2^{\widetilde{O}(G \log n)}$ distinguishable unitaries on n qubits. Hence we must have $G \log n = \widetilde{\Omega}(Tn)$, which yields $G = \widetilde{\Omega}(Tn)$. ■

IV. Discussion

We have only analyzed local Hamiltonians on (hyper)cubic lattices embedded in some Euclidean space, but Lieb-Robinson bounds with exponential dependence on the separation distance hold more generally. However, on more general graphs, it may be more difficult to find an appropriate decomposition that gives a small error; this must be analyzed for each graph. One advantage of the method here is that the accuracy improves for smaller Lieb-Robinson velocity. This can occur if the terms in the Hamiltonian have a small commutator (see the appendix in the full version of this paper).

The decomposition based on Lieb-Robinson bounds looks very similar to higher order Lie-Trotter-Suzuki formulas. The difference is in the fact that the overlap is chosen to be larger and larger (though very slowly) as the simulated spacetime volume increases. If we want an algorithm that

does not use any ancilla qubits, similar to algorithms based on Lie-Trotter-Suzuki formulas, then we can simulate the small blocks from Lieb-Robinson bounds by high order Suzuki formulas [9], [18] where the accuracy dependence is polynomial (power-law) of arbitrarily small exponent $a > 0$. This combination results in an algorithm of total gate complexity $O(Tn(Tn/\epsilon)^a)$, similar to what is claimed to be achievable in Ref. [19, Sec. 4.3].

Application to fermions is straightforward but worth mentioning. Since Hamiltonian terms always have fermion parity even, Lieb-Robinson bounds hold without any change. Given the block decomposition based on the Lieb-Robinson bound, we should implement each small blocks in $\mathrm{polylog}(Tn/\epsilon)$ gates. The Jordan-Wigner transformation, a representation of Clifford algebra, is a first method one may consider:

$$\gamma_{2j-1} \mapsto \sigma_1^z \otimes \cdots \otimes \sigma_{j-1}^z \otimes \sigma_j^x, \tag{9}$$

$$\gamma_{2j} \mapsto \sigma_1^z \otimes \cdots \otimes \sigma_{j-1}^z \otimes \sigma_j^y, \tag{10}$$

where $\gamma_{2j-1}, \gamma_{2j}$ are Majorana (real) fermion operators, and the right-hand side is a tensor product of Pauli matrices. Often, the tensor factor of σ^z preceding σ^x or σ^y is called a *Jordan-Wigner string*. In one spatial dimension, the above representation where the ordering of γ is the same as the chain's direction gives a local qubit Hamiltonian, since in any term Jordan-Wigner strings cancel. The ordering of fermions is thus very important. (Under periodic boundary conditions, at most one block may be nonlocal; however, we can circumvent the problem by regarding the periodic chain, a circle, as a double line of finite length whose end points are glued: $\big([-1,+1] \times \{\uparrow,\downarrow\}\big)/\{(-1,\uparrow) \equiv (-1,\downarrow), (+1,\uparrow) \equiv (+1,\downarrow)\}$. This trick doubles the density of qubits in the system, but is applicable in any dimensions for periodic boundary conditions.)

In higher dimensions with fermions, a naive ordering of fermion operators turns most of the small blocks into nonlocal operators under the Jordan-Wigner transformation. However, fortunately, there is a way to get around this, at a modest cost, by introducing auxiliary fermions and let them mediate the interaction of a target Hamiltonian [34]. The auxiliary fermions are "frozen," during the entire simulation, by an auxiliary Hamiltonian that commutes with the target Hamiltonian. With a specific ordering for the fermions, one can represent all the new interaction terms as local qubit operators. The key is that if $c_j c_k$ is a fermion coupling whose Jordan-Wigner strings do not cancel, we instead simulate $c_j c_k \gamma_1 \gamma_2$, where $\gamma_{1,2}$ are auxiliary, such that the Jordan-Wigner strings of γ_1, γ_2 cancel those of c_j, c_k, respectively. The auxiliary γ's may be "reused" for other interaction terms if the interaction term involves fermions that are close in a given ordering of fermions. (Ref. [34] explains the manipulation for quadratic terms but it is straightforward that any higher order terms can be treated similarly. They also manipulate the Hamiltonian for the auxiliary γ to make it local after Jordan-Wigner transformation, but for our simulation purposes it suffices to initialize the corresponding auxiliary qubits.) In this approach, if we insert $O(1)$ auxiliary fermion per fermion in the target system, the gate and depth complexity is the same as if there were no fermions. Note that we can make the density of auxiliary fermions arbitrarily small by increasing the simulation complexity. Divide the system with non-overlapping blocks of diameter ℓ, which is e.g. $O(\log n)$, that form a hypercubic lattice. (These blocks have nothing to do with our decomposition by Lieb-Robinson bounds.) Put $O(1)$ auxiliary fermions per block, and order all the fermions lexicographically so that all the fermions in a block be within a consecutive segment of length $O(\ell^D)$ in the ordering. Interaction terms within a block have Jordan-Wigner string of length at most $O(\ell^D)$, and so do the inter-block terms using the prescription of [34]. The gate and depth complexity of this modified approach has $\mathrm{poly}(\ell)$ overhead.

References

[1] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, no. 6–7, pp. 467–488, 1982.

[2] National Energy Research Scientific Computing Center, "NERSC Annual Report 2016," https://www.nersc.gov/news-publications/publications-reports/nersc-annual-reports, 2016.

[3] S. Lloyd, "Universal quantum simulators," *Science*, vol. 273, no. 5278, pp. 1073–1078, 1996. [Online]. Available: http://www.jstor.org/stable/2899535

[4] R. P. Feynman, "Quantum mechanical computers," *Optics News*, vol. 11, no. 2, pp. 11–20, 1985.

[5] D. Nagaj, "Fast universal quantum computation with railroad-switch local hamiltonians," *Journal of Mathematical Physics*, vol. 51, p. 062201, 2010.

[6] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, 1997, preliminary version in FOCS 1994.

[7] D. Aharonov and A. Ta-Shma, "Adiabatic quantum state generation and statistical zero knowledge," in *Proceedings of the 35th ACM Symposium on Theory of Computing*, 2003, pp. 20–29.

[8] A. M. Childs, "Quantum information processing in continuous time," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[9] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, "Efficient quantum algorithms for simulating sparse Hamiltonians," *Communications in Mathematical Physics*, vol. 270, no. 2, pp. 359–371, 2007.

[10] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, "Exponential improvement in precision for simulating sparse Hamiltonians," 2014, pp. 283–292.

[11] ——, "Simulating Hamiltonian dynamics with a truncated Taylor series," *Phys. Rev. Lett.*, vol. 114, p. 090502, 2015.

[12] D. W. Berry, A. M. Childs, and R. Kothari, "Hamiltonian simulation with nearly optimal dependence on all parameters," in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, Oct 2015, pp. 792–809.

[13] G. H. Low and I. L. Chuang, "Optimal Hamiltonian simulation by quantum signal processing," *Phys. Rev. Lett.*, vol. 118, p. 010501, 2017.

[14] ——, "Hamiltonian simulation by Qubitization," *arXiv preprint arXiv:1610.06546*, 2016.

[15] ——, "Hamiltonian simulation by uniform spectral amplification," *arXiv preprint arXiv:1707.05391*, 2017.

[16] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, "Toward the first quantum simulation with quantum speedup," *arXiv preprint arXiv:1711.10980*, 2017.

[17] H. F. Trotter, "On the product of semi-groups of operators," *Proceedings of the American Mathematical Society*, vol. 10, no. 4, pp. 545–545, apr 1959.

[18] M. Suzuki, "General theory of fractal path integrals with applications to many-body theories and statistical physics," *Journal of Mathematical Physics*, vol. 32, no. 2, pp. 400–407, 1991.

[19] S. P. Jordan, K. S. M. Lee, and J. Preskill, "Quantum computation of scattering in scalar quantum field theories," *Quantum Information and Computation*, vol. 14, no. 11&12, pp. 1014–1080, 2014.

[20] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, "Elementary gates for quantum computation," *Phys. Rev. A*, vol. 52, pp. 3457–3467, Nov 1995.

[21] E. Knill, "Approximation by quantum circuits," *arXiv preprint quant-ph/9508006*, 1995.

[22] E. H. Lieb and D. W. Robinson, "The finite group velocity of quantum spin systems," *Communications in Mathematical Physics*, vol. 28, no. 3, pp. 251–257, sep 1972.

[23] M. B. Hastings, "Lieb-Schultz-Mattis in higher dimensions," *Phys. Rev. B*, vol. 69, p. 104431, Mar 2004.

[24] B. Nachtergaele and R. Sims, "Lieb-Robinson bounds and the exponential clustering theorem," *Commun. Math. Phys.*, vol. 265, pp. 119–130, 2006.

[25] M. B. Hastings and T. Koma, "Spectral gap and exponential decay of correlations," *Commun. Math. Phys.*, vol. 265, pp. 781–804, 2006.

[26] M. B. Hastings, "Locality in quantum systems," *arXiv preprint arXiv:1008.5137*, 2010.

[27] T. J. Osborne, "The dynamics of 1D quantum spin systems can be approximated efficiently," *Phys. Rev. Lett.*, vol. 97, p. 157202, 2006.

[28] S. Michalakis, "Stability of the area law for the entropy of entanglement," *arXiv preprint arXiv:1206.6900*, 2012.

[29] G. H. Low and N. Wiebe, "Hamiltonian simulation in the interaction picture," 2018.

[30] M. Kieferova, A. Scherer, and D. Berry, "Simulating the dynamics of time-dependent Hamiltonians with a truncated Dyson series," 2018.

[31] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*. American Mathematical Society, 2002, vol. GSM 47.

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

[33] S. J. Szarek, "Metric entropy of homogeneous spaces and finsler geometry of classical lie groups," 1997.

[34] F. Verstraete and J. I. Cirac, "Mapping local hamiltonians of fermions to local hamiltonians of spins," *J. Stat. Mech.*, vol. 0509, p. 09012, 2005.