Improved one-dimensional area law for frustration-free systems

Itai Arad

School of Computer Science and Engineering, The Hebrew University, 91904 Jerusalem, Israel

Zeph Landau and Umesh Vazirani

University of California at Berkeley, Berkeley, California 94720, USA

(Received 12 November 2011; published 29 May 2012)

We present a new proof for the 1D area law for frustration-free systems with a constant gap, which exponentially improves the entropy bound in Hastings’ 1D area law and which is tight to within a polynomial factor. For particles of dimension \( d \), spectral gap \( \epsilon > 0 \), and interaction strength at most \( J \), our entropy bound is \( S_{1D} \leq O(1) \cdot X^d \log^8 X \), where \( X \equiv (J \log d)/\epsilon \). Our proof is completely combinatorial, combining the detectability lemma with basic tools from approximation theory. In higher dimensions, when the bipartitioning area is \( |\partial L| \), we use additional local structure in the proof and show that \( S \leq O(1) \cdot |\partial L|^2 \log^6 |\partial L| \cdot X^d \log^8 X \). This is at the cusp of being nontrivial in the 2D case, in the sense that any further improvement would yield a subvolume law.

DOI: 10.1103/PhysRevB.85.195145 PACS number(s): 71.27.+a, 03.65.Ud

I. INTRODUCTION

One of the striking differences between quantum and classical systems is the number of parameters that are needed to describe them. A classical system of \( n \) particles can be generally described by \( O(n) \) parameters, whereas an arbitrary state of a similar quantum system would generally require \( 2^{O(n)} \) parameters. This exponential gap is directly related to the phenomena of entanglement; quantum states do not have to be simple product states but can be an arbitrary superposition of such states.

But how genuine is this exponential gap? Is it an artifact of the fact that we are considering arbitrary quantum states, or is it an inherent characteristic of physical states that occur in nature, which are, of course, a much more restricted set of states? Among the best physical systems one may consider with respect to this question are quantum many-body systems on a lattice, which are ubiquitous in condensed matter physics. These systems are generally described by a local Hamiltonian that models the local interaction between neighboring particles. In particular, one is interested in the entanglement properties of the ground state of these systems.

If the system has a spectral gap \( \epsilon > 0 \), we can approach this state by cooling the system to temperatures below that gap. Such cold systems would be in an almost perfectly coherent state by cooling the system to temperatures below that gap.

Area laws constitute one of the most important tools for bounding entanglement in such systems.\(^1\) Starting from Bekenstein’s seminal result that the entropy of a black hole is proportional to the surface of its horizon,\(^2\) it was later hypothesized that the origin of this entropy is the quantum entanglement between the inner part of the black hole and its surrounding.\(^3,4\) This conjecture has led researchers to study the scaling of entanglement entropy in the ground (vacuum) states in models of quantum field theories,\(^5,6\) where it was demonstrated that also in there, entropy scales like the surface area of the region rather than its volume, albeit with some logarithmic corrections in critical cases. The same behavior was then demonstrated also in the context of spin chains by Vidal et al.,\(^7\) which has led to the formulation of the area-law conjecture that we now describe.

Consider a local-Hamiltonian system of \( d \)-dimensional particles that sit on a \( D \)-dimensional grid, as shown in Fig. 1. We say that a state \( |\psi \rangle \) of this system obeys an area law if, for any contiguous region \( L \) of the grid, the entanglement between the particles inside \( L \) and the particles outside \( L \) is upper bounded by a constant times the surface area of \( L \) as follows:

\[
S_L(\psi) \leq O(|\partial L|).
\]

The area-law conjecture then states that if the system has a constant spectral gap \( \epsilon > 0 \), its ground state necessarily obeys an area law. This is clearly much stronger than the trivial bound on the entropy (known as a volume law), which is proportional to the number of particles inside \( L \).

On the face of it, there are several reasons to believe that ground states of gapped local Hamiltonians obey an area law. One might intuitively expect that the entanglement is local because it is generated by local interactions. More formally, it has been shown by Hastings that in the ground state of gapped systems, the correlation between two local observables decays exponentially in their lattice distance.\(^8\) It is, therefore, tempting to conclude that only the degrees of freedom near the boundary are entangled to those outside the region. However, the existence of data hiding states\(^9\) implies that such simple reasoning is probably insufficient to prove an area law.\(^10\)

Indeed, even though the area-law conjecture was shown to be true in many specific models, mostly in 1D (see, for example, Refs. 1, 7,11, and 12 and references therein), it was not until a few years ago, in a seminal paper,\(^13\) that Hastings proved that it holds for all 1D systems with a spectral gap. In this case, an area law says that ground-state entropy across any cut in the 1D chain is bounded by a constant, independent...
of the system size. From this, one can deduce that, for any practical purposes, the ground state of such systems can be described by a polynomial number of parameters (say, by a matrix product state), instead of an exponential number (see Ref. 13).

Unfortunately, Hastings’ 1D proof does not scale up to higher dimensions. One reason lies in the form of the entropy bound. For a 1D system with a bounded interaction strength \( J \), a spectral gap \( \epsilon > 0 \), and a particle dimension \( d \), the upper bound is \( 2^{O(X)} \), where \( X \overset{\text{def}}{=} \frac{\log d}{\epsilon} \). This exponential dependence on \( \log d \) is catastrophic if we want to use this formula in higher dimensions by fusing together particles along surfaces parallel to \( \partial L \). It implies an entropy bound of \( 2^{O(|\partial L|)} \) instead of \( O(|\partial L|) \). This is exponentially larger than the trivial volume-law bound of \( O(L) \). Proving area laws in two or higher dimensions therefore is still a wide open problem and a central goal in quantum Hamiltonian complexity.

A combinatorial approach to proving the area law for 1D frustration-free systems was introduced in Ref. 14. The proof replaced Hastings’ analytical machinery, including the Lieb-Robinson bound and spectral Fourier analysis with the detectability lemma,\(^{15}\) a combinatorial lemma about local Hamiltonians. However, the resulting bound was no better than Hastings’ bound because, at its heart, the argument followed the same outline as Hastings’, including the use of a “mutual information saturation”-type argument (or, in effect, a kind of “monogamy of entanglement”-type argument) that leads to an exponential slack. Nevertheless, the combinatorial nature of the detectability lemma opened up the possibility of a new inherently combinatorial proof of the area law.

A. Our results

In this paper we give a new proof of the area law in 1D for the case of gapped frustration-free Hamiltonians. Our proof yields an exponentially better bound on the entanglement entropy than the bound in Ref. 13. Specifically, we prove Theorem I.1.

**Theorem I.1.** Consider a 1D, frustration-free local Hamiltonian system \( H = \sum_i H_i \), over a system of \( n \) particles of local dimension \( d \), with \( H_i \) being 2-local nearest-neighbor interactions with a bounded strength \( \|H_i\| \leq J \). Assume, further, that the system has a unique ground state \( |\Omega\rangle \) and a spectral gap \( \epsilon > 0 \), and define

\[
X \overset{\text{def}}{=} \frac{J \log d}{\epsilon}.
\]

Then, along any cut in the chain, the von Neumann entanglement entropy between the two sides is bounded by

\[
S_{1D} \leq O(1) \cdot X^3 \log^3 X.
\]

The importance of this work is threefold. First, in 1D, we exponentially improve the upper bound as a function of \( \epsilon^{-1} \): The bound in Eq. (3) implies an upper bound of \( O(\epsilon^{-3} \log^8 \epsilon^{-1}) \). This is within a polynomial factor of recent lower bounds that were found by Hastings and Gottesman\(^{16}\) and Irani.\(^{17,18}\) The former showed a 1D system with a fixed \( d \) and \( J \), in which the entropy along a cut is at least \( O(\epsilon^{-1/4}) \), and the latter showed a lower bound of \( O(\epsilon^{-1/12}) \) for a translation-invariant system.

Second, and more important, is the relevance of this work for proving area laws in higher dimensions. A naïve application of our bound to higher-dimensional systems would yield an entropy bound of \( S \leq |\partial L|^3 \cdot \text{poly}(\log |\partial L|) \), which is still worse than a volume law but is much better than the previous exponential bound. Moreover, as we show in Sec. V, in two or more dimensions, one can further exploit the local properties of the system along the boundary \( \partial L \) and improve the bound to \( \partial L|^2 \cdot \text{poly}(\log |\partial L|) \). This bound is at the cusp of being nontrivial; any further improvement that would bound the entropy by \( \partial L|^2 \cdot \text{poly}(\log |\partial L|) \) for any \( \partial > 0 \), would prove a subvolume law for 2D. In fact, this gives a subvolume law for the case of fractals with dimension \( 1 < D < 2 \), but we shall not pursue that direction here.

Finally, our approach differs in fundamental ways from Hastings’ original proof. Here, the “monogamy of entanglement”-type argument is replaced with an iterative procedure to find product states with increasingly higher overlap with the ground state. The central quantity used to analyze the progress of this procedure is Schmidt rank, rather than more advanced tools like relative entropy and its monotonicity. While the iterative procedure is based on the detectability lemma, a much more intricate combinatorial structure is necessary to ensure that the procedure makes progress. We now give a high-level overview of the proof.

B. High-level overview of the proof

Consider a 1D chain of \( n \) \( d \)-dimensional particles with nearest-neighbor interactions, described by the gapped, frustration-free Hamiltonian \( H = \sum_i Q_i \) with a unique ground state \( |\Omega\rangle \). For the sake of simplicity, we assume that \( Q_i \) are projections and, therefore, \( P_i \overset{\text{def}}{=} I - Q_i \) are projections to the local ground spaces of the different terms.

The key to proving an area law across a cut is to find a product state \( |\phi\rangle = |\phi_L\rangle \otimes |\phi_R\rangle \) with respect to the bipartitioning of the system, which has a large overlap with \( |\Omega\rangle \). Our approach to finding such a product state is to start with any product state with a nonzero overlap with \( |\Omega\rangle \) and act on it with an operator that increases its overlap with \( |\Omega\rangle \), without increasing its Schmidt rank much. Specifically, we construct an operator \( K \) with the following property: \( K \) fixes \( |\Omega\rangle \), but when applied to any state \( |\psi\rangle \), it shrinks the component
orthogonal to $\ket{\Omega}$ by a factor of $\Delta$ while increasing the Schmidt rank of $\ket{\psi}$ by, at most, a factor of $D$. Clearly, there is a race between these two factors $D$ and $\Delta$. It turns out that when $D \cdot \Delta < 1/2$, we can amplify the overlap with $\ket{\Omega}$ by replacing $\ket{\phi} = \ket{\phi_L} \otimes \ket{\phi_T}$ with one of the Schmidt vectors of $K$.$\ket{\phi}$. This amplification continues all the way until the overlap is $\sqrt{T}(2D)$. A few more applications of $K$ to this product state yield a state with Schmidt rank $D^{O(1)}$, which has constant ($D$ independent) overlap with $\ket{\Omega}$. Further applications of $K$ give rise to Schmidt coefficients with vanishing mass, and, therefore, the entanglement entropy of $\ket{\Omega}$ can be bounded by $O(\log D)$.

The task of proving an area law therefore is reduced to the task of finding an operator $K$ with $D \cdot \Delta < 1/2$. Our starting point is the detectability lemma (DL). Denote the spectral gap by $\epsilon > 0$. We can partition the projections $\{P_i\}$ into two subsets of even and odd projections, which are called “layers” (see Fig. 2). Inside each layer, the projections commute because they are nonintersecting. Consequently, $\Pi_{\text{odd}} \overset{\text{def}}{=} P_1 \cdot P_3 \cdot P_5 \cdot \cdots$ and $\Pi_{\text{even}} \overset{\text{def}}{=} P_2 \cdot P_4 \cdot P_6 \cdots$ are the projections into the common ground spaces of the odd and even layers, respectively. Then, by the DL, the operator $A \overset{\text{def}}{=} \Pi_{\text{even}} \Pi_{\text{odd}}$ is an approximation to the ground-state projection. It preserves the ground state while shrinking its perpendicular subspace by an $n$-independent factor $\Delta_0(\epsilon) \simeq 1 - \epsilon \epsilon$, where $c$ is a geometrical factor. Moreover, each application of $A$ increases the Schmidt rank of our state by a constant factor of, at most, $D_0 \overset{\text{def}}{=} \Delta^2$ (due to the projection that intersects with the cut in the chain; see Fig. 2). Unfortunately, we generally expect $D_0 \cdot \Delta_0 > 1$, so the operator $A$ does not, by itself, suffice to carry out our plan.

To construct the operator $K$ we need several new ideas. First, we observe that $D_0$ and $\Delta_0$ can be replaced by $D_0^k$ and $\Delta_0^k$, respectively, by coarse graining: Fuse $k$ adjacent particles, making them a single particle of dimension $d^k$. Although this only increases the value of the product, it creates room for the next step, which is to modify the operator $A$ to decrease the factor by which it blows up the Schmidt rank. For concreteness, assume that the even layer contains the projection that intersects with the cut. We will focus on a segment of $m$ projections around the cut and denote their product by $\Pi_m$, so $\Pi_{\text{even}} = \Pi_m \Pi_{\text{rest}}$. We will replace the operator $\Pi_{\text{even}}$ with $\Pi_m \Pi_{\text{rest}}$ that closely approximates $\Pi_{\text{even}}$ while increasing the Schmidt rank by much less than $D_0^k$ (when amortized over several applications).

One of the great benefits of using the DL is that is the all projections in a given layer commute, and, hence, much of the following analysis becomes almost classical. Indeed, the $m$ projections around the cut $\{P_i\}_{i=1}^m$ define a decomposition of the Hilbert space of the system into a direct sum of $2^m$ eigenspaces, called sectors. Each sector is defined by a string $s = (s_1, \ldots, s_m)$, such that if $\ket{\psi_s}$ is in the $s$ sector, $P_i \ket{\psi_s} = (1 - s_i) \ket{\psi_s}$. A site with $s_i = 1$ is called a violation, since it corresponds to a nonzero energy of the corresponding local Hamiltonian term, and $\sum_{i=1}^m s_i$ is the total number of violations in the sector $s$.

Now, an arbitrary state $\ket{\psi}$ can be decomposed as $\ket{\psi} = \ket{\psi_0} + \ket{\psi_1}$, where $\ket{\psi_0}$ is its projection on the zero-violations sector and $\ket{\psi_1}$ is its projection on the violation sectors. Clearly $\Pi_m \ket{\psi} = \ket{\psi_0}$. To approximate this behavior, we will use the $\{P_i\}$ projections to construct an operator $\tilde{A}_m$ that is diagonal in the sectors decomposition, and, in addition, $\Pi_m \ket{\psi} = \ket{\psi_0} + \ket{\psi_1}$, with $\ket{\psi_1}$ in the violation sectors and $\|\psi_1\|^2 \leq \delta \|\psi_1\|^2$. It follows that the operator $A \overset{\text{def}}{=} \Pi_{m} \Pi_{\text{rest}} \Pi_{\text{odd}}$ approximates $\Pi_{\text{even}} \Pi_{\text{odd}}$ in the sense that $\tilde{A}(\Omega) = \ket{\Omegam}, \tilde{A}(\Omegam) \in H_1$, and $\|\tilde{A}(\Omegam)\|^2 \leq (\Delta_0^k + \delta)\|\psi_1\|^2$. Let $\tilde{A} \overset{\text{def}}{=} \Delta_0^k + \delta$.

To construct the operator $\Pi_m$, first consider the operator $\Pi = \sum_{i=1}^m (1 - P_i)$. The operator $\Pi$ counts the number of violations in a sector: If $\ket{\psi_s}$ belongs to the $s$ sector, $\Pi \ket{\psi_s} = \ket{s} \cdot \ket{\psi_s}$. The operator $\Pi_m$ will be a polynomial in $\Pi$, with the polynomial evaluating to 1 on $|s| = 0$, and less than $\delta$ on input with $|s|$ between 1 and $m$. Three ideas play a critical role in the construction of this polynomial and in bounding the increase in Schmidt rank. The first is the use of a Chebyshev polynomial, which achieves the desired behavior at the $m + 1$ points with a degree of only $j = O(m \log \delta^{-1})$. The second idea is that it suffices to bound the entanglement across any of the $m$ cuts and then to pay a further penalty of, at most, $D_1 \overset{\text{def}}{=} (D_0^k)^m$ to bound the entanglement across the cut of interest. So, if we consider the operator $A^\ell$, each term has degree $j \ell$ (i.e., is a product of $j \ell$ of the $P_s$), so the typical cut is crossed $j \ell / m$ times, resulting in a Schmidt rank increase by $(D_0^k)^j / m \simeq D_0^{j/\sqrt{m}}$. This means that the incremental Schmidt rank per application of a term of $A$ is $D_0^{j/\sqrt{m}}$, which can be made arbitrarily small by choosing $m$ to be large enough. Finally, a recursive grouping argument shows that we do not have to pay a price in Schmidt rank proportional to the number of terms in the polynomial (which would have been catastrophic); instead, we can decompose the operator $A^\ell$ as a sum of only $2^{O(\log^j)}$ operators, for each of which there is a (possibly different) cut with entanglement increase of $\simeq D_0^{j/\sqrt{m}}$.

Putting it all together, we have an operator $K = \tilde{A}^\ell$, which increases the Schmidt rank by $D = D_1 \tilde{A}^\ell$, where $\tilde{A} = 2^{O(\log^j)} D_0^{j/\sqrt{m}}$, and $j = O(\sqrt{m \log \delta^{-1}})$ and achieves a shrinking factor of $\Delta = \Delta'^k$ for $\Delta = \Delta_0^k + \delta$. It is now a matter of simple algebra to fix the parameters $m$, $\delta$, $k$, and $\ell$ such that $D \cdot \Delta < 1/2$. The end results turns out to be $\log D = O(1) \cdot X^3 \cdot \log X$ for $X = (\log D_0) / \epsilon$, and this completes the proof.

---

FIG. 2. (Color online) The settings of the detectability lemma for a 1D system with 12 particles $H = \sum_{i=1}^{12} Q_i$. The local interaction terms, $Q_i$, are divided into two layers: the even layer and the odd layer. The dashed red line denotes a cut in the system between particles 6 and 7.
C. Organization of paper

We begin with some preliminary definitions and known results from mathematics and quantum information in Sec. II.

In Sec. III we give the proof of our main result, Theorem I.1. The heart of the proof, which is its most technical part, is the diluting lemma. It is proved separately in Sec. IV. In Sec. V we provide an outline for our entanglement bound in 2D and beyond (the “almost volume law” result), and in Sec. VI we sketch the implication of our area law to the existence of a unique ground state beyond (the “almost volume law” result), and in Sec. VII we offer our summary and conclusions.

II. NOTATION AND PRELIMINARIES

Throughout this paper \log(\cdot) will denote the base 2 logarithmic function.

A. Local Hamiltonians

We consider a \(k\)-local Hamiltonian \(H\) acting on \(\mathcal{H} = (\mathbb{C}^d)^\otimes_n\), the Hilbert space of \(n\) particles (spins, qudits) of dimension \(d\) that sit on a \(D\)-dimensional grid. Our main result concerns the 1D case with \(D = 1\), but in Sec. V we will consider higher dimensions. We assume \(H = \sum_i H_i\), where each \(H_i\) is a non-negative bounded operator that acts nontrivially on a constant number of \(k\) particles (hence the term local Hamiltonian). We further assume that \(H\) has a unique ground state \(|\Omega\rangle\) with ground energy 0 and has constant spectral gap \(\epsilon > 0\). Since the \(H_i\) are all non-negative, the ground state must be a common zero eigenstate for each of the \(H_i\); a Hamiltonian with this feature is known as frustration free. We denote by \(\mathcal{H}_L \subset \mathcal{H}\) the orthogonal complement of the ground space of \(H\). Thus, \(\mathcal{H}_L\) is an invariant subspace for \(H\) and

\[
H|_{\mathcal{H}_L} \geq \epsilon \mathbb{1}.
\] (4)

We further assume that the \(H_i\) terms are projections, and, henceforth, we will denote them by \(Q_i\) to remind the reader. We define \(P_i\) to be the projection on the ground space of \(Q_i\), \(P_i \overset{\text{def}}{=} \mathbb{1} - Q_i\). The assumption that \(H\) is made of projections is not actually a restriction, as is demonstrated, for example, in Sec. 2 of Ref. 14. Indeed, given a general frustration-free system with a spectral gap \(\epsilon > 0\) and interaction strength bounded by \(\|H_i\| \leq J\), one can always pass to an equivalent system that shares the same ground space, which is made of projections and has a spectral gap \(\epsilon/J\). Therefore, throughout the paper, we will drop the \(J\) dependence; we will assume \(J = 1\) and work in dimensionless units.

B. The Schmidt decomposition

Given a state \(|\phi\rangle\) and a bipartition of the system to two noninteracting sets, \(R\) and \(L\), with corresponding Hilbert spaces \(\mathcal{H}_R, \mathcal{H}_L\) such that \(\mathcal{H} = \mathcal{H}_L \otimes \mathcal{H}_R\), we can consider the Schmidt decomposition of the state along this cut: \(|\phi\rangle = \sum_j \lambda_j |L_j\rangle \otimes |R_j\rangle\). Here \(\lambda_1 \geq \lambda_2 \geq \ldots\) are the Schmidt coefficients. Their squares sum to 1 (if \(|\phi\rangle\) is normalized) and are equal to the nonzero eigenvalues of the reduced density matrices of either sides of the cut.

The number of nonzero Schmidt coefficients in the Schmidt decomposition of \(|\phi\rangle\) is called the Schmidt rank (SR), which we shall denote as \(SR(\phi)\). The usefulness of the SR stems from it being a “worst case” estimate for the entanglement. As such, it is often easy to bound. The following facts are easy to verify.

**Fact II.1.**

1. \(SR(\phi + \psi) \leq SR(\phi) + SR(\psi)\).
2. If \(O\) is a \(k\)-local operator whose support intersects both \(\mathcal{H}_L\) and \(\mathcal{H}_R\), then it can increase the SR (with respect to the bipartitioning) of any state by, at most, a factor of \(d^k\): \(SR(O\phi) \leq d^kSR(\phi)\). If \(O\) intersects only one part of the system, its action cannot increase the SR.
3. Consider a 1D system. If \(r_i\) and \(r_j\) are the SR of \(|\phi\rangle\) that correspond to cuts between particles \(i, i + 1\) and \(j, j + 1\), then \(d^{-|i-j|} r_i \leq r_i \leq d^{|i-j|} r_j\).

An important fact about the SR is the following corollary of the Eckart-Young theorem,19 which states that the truncated matrices of either sides of the cut.

\[
A|_{\mathcal{H}_L} \leq \Delta_0(\epsilon) \overset{\text{def}}{=} \frac{1}{(\epsilon/2 + 1)^{2/3}}.
\] (6)

For brevity, we will often drop the \(\epsilon\) dependence and simply write \(\Delta_0\). We note that the DL is not restricted to the 1D case
and can be easily generalized to any dimension by using more than two layers.

In the commuting case, \( \Pi_{\text{even}} \) and \( \Pi_{\text{odd}} \) commute with each other, and their product, \( A \overset{\text{def}}{=} \Pi_{\text{odd}} \cdot \Pi_{\text{even}} \), is the projection into the ground state of the system. Generally, however, they do not commute, and, as a result, \( A \) is only an approximation to the ground space projection: It leaves the ground space invariant while shrinking its perpendicular space by some factor. The DL quantifies this approximation: It tells us that the shrinking factor is bounded away from 1 by a constant \( \Delta_0(\epsilon) \) that depends on the spectral gap and not on the system size.

### D. Chebyshev polynomials of the first kind

We recall some basic facts about Chebyshev polynomials of the first kind. More information can be found in Chapter 22 of Ref. 20 or in any standard text book on approximation theory. This is a family of polynomials \( \{T_n(x)\} \), where \( n \) denotes the degree of the polynomial. They are solutions of the ODE

\[
(1 - x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + n^2y = 0.
\]

(7)

For \( |x| \leq 1 \) they are given explicitly by

\[
T_n(x) \overset{\text{def}}{=} \cos(n \cos^{-1} x),
\]

(8)

and, for general \( x \), they are given recursively by

\[
T_0(x) = 1,
\]

\[
T_1(x) = x,
\]

\[
T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).
\]

(11)

By the recursion relation, the coefficient of \( x^n \) in \( T_n(x) \) is \( 2^{n-1} \), and by Eq. (8), it has \( n \) simple roots (Chebyshev nodes) in \((-1, 1), x_k = \cos(\frac{2k-1}{n})\), \( k = 1, \ldots, n \). Therefore,

\[
T_n(x) = 2^{n-1} \prod_{k=1}^{n}(x - x_k).
\]

(12)

We conclude this section with an inequality about the behavior \( T_n(x) \) near \( x = -1 \) that will be used in the proof of the diluting lemma. From Eq. (8) it follows that for all \( 1 \leq x \leq 1 \), \( |T_n(x)| \leq 1 \) and that \( T_n(1) = 1; T_n(-1) = (-1)^n \). Finally, let us look at the derivative of \( T_n(x) \) at \( x = -1 \): substituting \( x = -1 \) at Eq. (7) gives us \( T'_n(-1) + n^2T_n(-1) = 0 \) and so \( T'_n(-1) = -n^2T_n(-1) = (-1)^n n^2 \). Since \( T_n(x) \) has all its minimas/maximas inside \([-1, 1]\), its derivative \( T'_n(x) \) outside that region can only grow in absolute value, and, therefore, for all \( \delta > 0 \):

\[
|T_n(-1 - \delta)| \geq |T_n(-1)| + \delta |T'_n(-1)| \geq 1 + n^2\delta.
\]

(13)

### III. PROOF OF THE MAIN THEOREM

Recall from the outline of the proof presented in the introduction, that our goal is to construct an operator \( K \) whose effect is to rapidly increase the overlap with the ground state, while only slowly increasing the Schmidt rank. We formalize this property of \( K \) in the notion of a \((D, \Delta)\)-AGSP below and then show that if the trade-off between \( D \) and \( \Delta \) if favorable, i.e., \( D \cdot \Delta < 1/2 \), then we can show that there is a product state that has large overlap with the ground state, which, in turn, leads to a bound on the entanglement entropy of the ground state. Once this is established, we can move on to the central construction of this work, which is performed in the diluting lemma III.5. This is where an operator with the proper trade-off of \( D \) and \( \Delta \) is constructed.

We begin with a quantitative definition of an operator that moves any vector toward the ground state.

**Definition III.1 An approximate ground-space projection (AGSP).** Consider a local Hamiltonian system \( H = \sum_i H_i \) on a 1D chain, together with a cut between particles \( i^* \) and \( i^* + 1 \) that bipartitions the system. We say that an operator \( K \) is a \((D, \Delta)\)-Approximate Ground Space Projection (with respect to the cut) if the following holds:

(a) **Ground space invariance:** for any ground state \( \Omega \), \( K|\Omega\rangle = |\Omega\rangle \).

(b) **Shrinking:** for any state \( |\phi\rangle \in \mathcal{H}^1 \), also \( K|\Omega\rangle \in \mathcal{H}^1 \), and \( \|K|\Omega\rangle\|^2 \leq \Delta \).

(c) **Entangling:** for any state \( |\phi\rangle \), \( \text{SR}(K|\phi\rangle) \leq D \cdot \text{SR}(\phi) \), where \( \text{SR}(\cdot) \) is evaluated with respect to the bipartitioning.

We refer to \( D \) as the SR factor and \( \Delta \) as the shrinking factor.

We note that with this definition, the DL combined with Fact II.1 implies that \( A = \Pi_{\text{even}}\Pi_{\text{odd}} \) is a \((D_0, \Delta_0) \overset{\text{def}}{=} (D^2, 1 + \epsilon/2)^{1/2})\)-AGSP. In its bare form, however, this operator is not useful to us since its SR factor is too large with respect to its shrinking factor. Specifically, it turns that the important quantity to consider is the product \( D \cdot \Delta \). The following lemma shows how having a \((D, \Delta)\)-AGSP with \( D \cdot \Delta < \frac{1}{4} \) implies the existence of a product state whose overlap with the ground state is at least \( 1/\sqrt{2D} \).

**Lemma III.2.** If there exists an \((D, \Delta)\)-AGSP with \( D \cdot \Delta < \frac{1}{4} \) then there is a product state \( |\phi\rangle = |L\rangle \otimes |R\rangle \) whose overlap with the ground state is at least \( \frac{1}{\sqrt{2D}} \).

**Proof.** Let \( K \) be a \((D, \Delta)\)-AGSP with \( D \cdot \Delta \leq \frac{1}{4} \) and \( |\phi\rangle \) a product state \( |\phi\rangle \overset{\text{def}}{=} |L\rangle \otimes |R\rangle \) whose overlap with the ground state is at least \( \mu = \langle \Omega |\phi\rangle \geq 1/\sqrt{2D} \). Below we show that there exists another product state with a larger overlap with the ground state.

By assumption, we can expand \( |\phi\rangle \) as \( |\phi\rangle = \mu|\Omega\rangle + (1 - \mu^2)^{1/2}|\phi\rangle \), with \( |\phi\rangle \in \mathcal{H}^1 \). Let \( |\phi_1\rangle \overset{\text{def}}{=} K|\phi\rangle \). Since \( K \) is an \((D, \Delta)\)-AGSP, it follows that \( |\phi_1\rangle = \mu|\Omega\rangle + \delta_1|\phi_1\rangle \) with \( |\phi_1\rangle \in \mathcal{H}^1 \), \( \delta_1^2 \leq \Delta \), and \( \text{SR}(\phi_1) \leq D \).

Setting \( |v\rangle = \frac{1}{\sqrt{\text{SR}(\phi_1)}} |\phi_1\rangle \) to be the normalization of \( |\phi_1\rangle \), we have \( \text{SR}(v) = \text{SR}(\phi_1) \leq D \). Therefore, its Schmidt decomposition can be written as \( |v\rangle = \sum_{i=1}^{D} \lambda_i|L_i\rangle|R_i\rangle \). It follows that

\[
|\langle \Omega |v\rangle| = \frac{\mu}{\|\phi_1\|} \leq \sum_{i=1}^{D} \lambda_i|\langle \Omega |L_i\rangle|R_i\rangle| \leq \sqrt{\sum_{i=1}^{D} |\langle \Omega |L_i\rangle|R_i\rangle|^2},
\]

where the last inequality follows from Cauchy-Schwarz and the fact that \( \sum \lambda_i^2 = 1 \). Therefore, there must be an \( i \) for which

\[
|\langle \Omega |L_i\rangle|R_i\rangle|^2 \geq \frac{D \cdot \Delta}{\mu^2 + \delta_1^2} \geq \frac{\mu^2}{2D}.
\]
But since $D \cdot \Delta < 1/2$, and, by assumption $\mu < \frac{1}{\sqrt{29}}$, it follows that $D(\mu^2 + \Delta) < 1$, and so the overlap of $|L_1\rangle|R_i\rangle$ with the ground state is larger than $\mu$.

With this bound in place, we start from the product state with the maximal overlap with the ground state, and use any AGSP to obtain controlled approximations of the ground state, from which an upper bound on its entropy can be found. A very similar argument was used in Hastings’ proof of the 1D area law.13

**Lemma III.3.** If there exists a product state whose overlap with the ground state is at least $\mu$, together with a $(D, \Delta)$-AGSP, then the entanglement entropy of $\Omega$ is bounded by

$$S \leq O(1) \cdot \frac{\log \mu^{-1}}{\log \Delta^{-1}} \log D.$$

(14)

The proof can be found in the appendix. The brief overview is that we begin with the asserted product state and repeatedly apply the AGSP to it. The result is a series of vectors with increasing SR by a factor $D, \Delta$, and, by assumption a little, we relabel them by $P_1, \ldots, P_m$ and define

$$\Pi_m \equiv \prod_{i=1}^{m} P_i,$$

(21)

the projection into the common ground state of these projections. Note that in this notation the cut intersects with $P_i$, where $i^* = \lfloor m/2 \rfloor$ (see Fig. 3). Denoting by $\Pi_{\text{rest}}$ the product of all the remaining projections in $\Pi_{\text{even}}$, we have $\Pi_{\text{even}} = \Pi_m \cdot \Pi_{\text{rest}}$. We will approximate $\Pi_m$ by an operator $\hat{\Pi}_m$ and then define $\hat{A}$ by

$$\hat{A} \equiv \Pi_m \cdot \Pi_{\text{rest}} \cdot \Pi_{\text{odd}}.$$

(22)

Analysis of the amount of entanglement created by powers of $\hat{A}$ will focus exclusively on the structure of $\hat{\Pi}_m$ since $\Pi_{\text{rest}}, \Pi_{\text{odd}}$ do not increase the SR along the cut.

One of the great benefits of using the DL is that the all projections in a given layer commute, and, hence, the analysis becomes almost classical. Indeed, as we said in the outline of the proof, the projections in $\Pi_m$ define a decomposition of the Hilbert space of the system into a direct sum of $2^m$ eigenspaces, called sectors. Each sector is defined by a string $s = (s_1, \ldots, s_m)$, such that if $|\psi_s\rangle$ is in the $s$ sector, $P_i|\psi_s\rangle = (1 - s_i)|\psi_s\rangle$. A site with $s_i = 1$ is called a violation, since it corresponds to a nonzero energy of the corresponding local Hamiltonian term. We denote by $|\delta\rangle = \sum_{i=1}^{m} s_i$ the total number of violations in the sector $s$. Finally, we will also consider a coarse-grained decomposition in which we group together all the $\left(\frac{m}{2}\right)$ sectors with $k$ violations. The direct sum of these subspaces is called the $k$-violations sector.

Consider now an arbitrary state $|\psi\rangle$ and expand it as $|\psi\rangle = |\psi_0\rangle + |\psi_1\rangle$, where $|\psi_0\rangle$ is its projection on the zero-violations sector and $|\psi_1\rangle$ is its projection on the violating sectors. Clearly, $\Pi_m|\psi\rangle = |\psi_0\rangle$. To approximate this behavior, we will use the $P_{\ell i}$ projections to construct an operator $\Pi_m$ that is diagonal in the sectors decomposition and, in addition, $\Pi_m|\psi_0\rangle = |\psi_0\rangle + |\psi_1\rangle$, with $|\psi_1\rangle$ in the violating sectors and $\|\psi_1\|^2 \leq \delta\|\psi_0\|^2$ for some error parameter $\delta > 0$. It then is easy to verify that $\hat{A}|\Omega\rangle = |\Omega\rangle, A|\Omega^+\rangle = \hat{H}^+,$ and

$$K \equiv \hat{A}^\dagger (D, \Delta) = (D, \hat{D}^\dagger, \hat{A}^\dagger),$\

(20)

with $(D, \Delta) = (D, \hat{D}^\dagger, \hat{A}^\dagger).$ It satisfies $D \cdot \Delta < 1/2$ and

$$\log D = O(1) \cdot \log D_1 \cdot \log \hat{D} \leq O(1) \cdot \chi^3 \log^8 X.$$

We now proceed to define the general form of $\hat{A}$. 

### A. General settings

Without loss of generality, we assume that the bipartitioning cut in the chain intersects with an even projection (see Fig. 2). As a result, when applying $A$, only the $\Pi_{\text{even}}$ portion of the operator increases the SR. Therefore, our construction of $\hat{A}$ will modify $\Pi_{\text{even}}$, leaving $\Pi_{\text{odd}}$ intact.

We begin by considering the piece of $\Pi_{\text{even}}$ consisting of the $m$ even projections closest to the cut. We denote this set of projections by $I_m$ and, abuse the notation a little, we replace them by $P_1, \ldots, P_m$ and define

$$\Pi_m \equiv \prod_{i=1}^{m} P_i,$$

(21)

for any state $|\psi\rangle$. Then, for any cut $|\psi\rangle = |\psi_0\rangle + |\psi_1\rangle$.

(17)

The parameters $D_j, \hat{D}$ satisfy

$$\hat{D} \cdot \hat{A} = \frac{1}{2},$$

$$\log \hat{D} \leq \log (\hat{A}^{-1}) \leq O(1) \cdot \log^2 X,$$

(18)

$$\log D \leq O(1) \cdot \chi^3 \log^8 X.$$

(19)
\[ \| \hat{A}^{\ell} \| \overset{\Delta}{=} (\Delta + \delta)\Omega_{\ell} \|. \]  This gives us
\[ \hat{A} = \Delta_0 + \delta. \]  (23)

B. Constructing \( \hat{A}_{\ell m} \): A general discussion

Before actually constructing the operator \( \hat{A}_{\ell m} \), it might be useful to describe what we hope to accomplish. Recall that we would like to show that the increase in SR due to application of \( \hat{A}^\ell \) is bounded by \( D_\ell \hat{D}^\ell \), where \( D \cdot \hat{D} \leq \frac{1}{2} \). The rough idea is to show that there is some cut between particles \( i \) and \( i+1 \) within the support of \( \hat{A}^\ell \), such that the SR across this cut does not grow much due to application of \( \hat{A}^\ell \). This would account for the factor \( D^\ell \) in the above bound. Moreover, by Fact II.1, the SR across the middle cut \( i^*, i^* + 1 \) can only be larger than the SR between \( i, i+1 \) by a factor of, at most, \( d^{i-r+1} \leq D^\ell_0 \leq D_\ell \).

There are several approaches to constructing \( \hat{A}_{\ell m} \). Perhaps the most obvious one is by “diluting”; instead of using a product of \( m \) projections, we may use a product of \( r m \) randomly chosen projections for some \( 0 < r < 1 \). After applying such \( \hat{A}_{\ell m} \) for \( \ell \) layers, there would be columns with fewer than \( r \ell \) entangling projections, and so the SR along these columns will be bounded by \( D^\ell_0 \). Applying Fact II.1 as described above, the SR in the middle cut would be, at most, \( D^\ell_0 \leq D^\ell \), implying \( D = D^\ell \). One naive approach is to take \( D_\ell = 1 \) for every integer \( 1 \leq x \leq m \).

C. Constructing \( \hat{A}_{\ell m} \) using Chebyshev polynomials

To construct \( \hat{A}_{\ell m} \) entirely from \( \mathbb{N} \), we want a polynomial of minimal degree \( P(x) \) such that \( P(0) = 1 \) and \( |P(x)| \leq \delta < 1 \) for every integer \( 1 \leq x \leq m \). In such a case, \( \hat{A}_{\ell m} \) will leave the zero-violations sector intact, while shrinking the other sectors by \( \delta \). One naive approach is to take \( P(x) = 1 - x \cdot (1 - x/2) \cdots (1 - x/m) \). This gives \( \delta = 0 \) at the price of a polynomial of degree \( m \), the same degree as the original \( \Pi_{\ell m} \) (thereby creating too much entanglement). For a lower-degree polynomial with the desired properties we turn to the Chebyshev polynomial, a central object in approximation theory. As noted in Sec. II.D, the Chebyshev polynomial of the first kind \( T_n(x) \) oscillates between \(-1\) and \(1\) in the region \([-1, 1]\) and then increases rapidly outside that region. The idea, therefore, is to use a polynomial that is the mapping of the Chebyshev polynomial from \([-1, 1]\) to \([1, m]\) and rescaled to be \(1\) at \(x = 0\).

Definition IV.2 (the \( C_m(x) \) polynomial). The \( C_m(x) \) polynomial is a \( \sqrt{m} \)-order polynomial that is defined by
\[ C_m(x) \overset{\text{def}}{=} T_{\sqrt{m}} \left( \frac{x - \frac{m+1}{2}}{\frac{m+1}{2}} \right). \]  (25)
\[ C_m(0) \overset{\text{def}}{=} \frac{1}{C_m(0)}. \]  (26)

Figure 4 shows \( C_m(x) \) for \( m = 36 \). It is easy to verify that (1) \( C_m(0) = 1 \) and (2) \( |C_m(x)|^2 \leq \frac{1}{2} \) for every \( 1 \leq x \leq m \). The first claim follows from the definition, while the second follows from the fact that \( |\hat{C}_m(0)| \geq 3 \), which follows from Eq. (13):
\[ |\hat{C}_m(0)| = \left| T_{\sqrt{m}} \left( -1 - \frac{2}{m-1} \right) \right| \geq 1 + \frac{2}{m-1} (\sqrt{m})^2 \geq 3. \]  (27)

Note that \( C_m(x) \) is a polynomial of degree \( \sqrt{m} \), a huge improvement over the naive construction of degree \( m \). In fact,
it can be shown that this construction is optimal: As shown in Ref. 23, any polynomial that satisfies the above two properties must be of at least degree $\sqrt{m}$.

To improve the shrinking factor, we can apply $[C_m(x)]^q$, where $q$ is a parameter to be determined later. We can now define the operator $\hat{A}$ that appears in the diluting lemma.

**Definition IV.3 (the Chebyshev-based operator $\hat{A}$).** Given integers $m$ and $q$, the Chebyshev-based operator $\hat{A}$ is

$$\hat{A} \equiv \hat{\Pi}_m \cdot \Pi_{\text{rest}} \cdot \Pi_{\text{odd}},$$

where

$$\hat{\Pi}_m \equiv [C_m(N)]^q.$$

We note that $\hat{\Pi}_m$ is a degree $j \equiv q \sqrt{m}$ polynomial in $N$, which leaves the zero-violations sector intact and shrinks the violating sectors by at least $\delta = (1/9)^q$. Consequently, by Eq. (23), we get $\Delta = \Delta_0 + (1/9)^q$.

Let us now turn to the task of upper bounding the SR that is generated by $\hat{A}^\ell$.

**D. Upper bounding the SR factor of $\hat{A}^\ell$**

We bound SR generated by $\hat{A}^\ell$ in the following lemma:

**Lemma IV.4.** Set $j \equiv q \sqrt{m}$, the degree of $[C_m(x)]^q$. Then, for any state $|\phi\rangle$,

$$\text{SR}(\hat{A}^\ell \phi) \leq D_1 \hat{D}^\ell \cdot \text{SR}(\phi),$$

where

$$D_1 = D_0^m, \quad \hat{D} = 20 \cdot 2^{\frac{1}{2} \log j} + 2^{\frac{1}{2} \log j}, \quad D_0^{j/m}. \quad (31)$$

*Proof.* Recall that $\hat{\Pi}_m$ is a polynomial of degree $j$ in $N$ and, therefore, can be written as a sum of $j + 1$ terms:

$$\hat{\Pi}_m = \sum_{i=0}^{j} c_i N^i. \quad (33)$$

Consequently, $\hat{A}^\ell$ can be written as a superposition of $(j + 1)^\ell$ terms of the form

$$N^{i_1} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}) \cdot N^{i_2} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}) \cdot \ldots \cdot N^{i_j} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}),$$

with $i_1, \ldots, i_j$ between 0 and $j$. When applied to a general state $|\phi\rangle$, the term that potentially generates the highest SR is $(N^{i_1} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}})^j \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}})$, we will, therefore, upper bound the total SR by upper bounding its SR and multiplying the end result by $(j + 1)^\ell$.

**Definition IV.5 (a min-entangling operator).** We say that an operator $C$ is min-entangling with respect to some cut in the support of $I_m$ if it is of the form

$$M_{i_1} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}) \cdot M_{i_2} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}) \cdot \ldots \cdot M_{i_j} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}),$$

where $M_i$ are polynomials in the projections of $I_m$ such that

1. There exists a subset of, at most, $j \ell / m$ of the $M_i$’s that contains the projection $P \in I_m$ that intersects with the cut.
2. Each of the rest of the $M_i$ only contains projections that are either strictly to the left of $P$ or strictly to the right of $P$ (in $I_m$).

It follows that only those $M_i$ that contain the “entangling” projection $P$ increase the SR across the cut, and, therefore, the total SR increase is bounded by $D_0^{j/\ell m}$. The bound on the SR follows from the following decomposition of $\hat{A}^\ell$:

**FIG. 4.** (Color online) The polynomial $C_{36}(x)$, which is based on the Chebyshev polynomial of the first kind $T_6(x)$. For $x \in [1, 36]$ we have $|C_{36}(x)| \leq 1/3$.

**FIG. 5.** (Color online) An illustration of a $k$-coarse-grained system in 1D with $k = 4$. The elongated rectangles denote the coarse-grained projections and the ovals denote the original projections. Underneath even coarse-grained projections, one can “pull” a pyramid of original projections and, similarly, above an odd coarse-grained projection. Together, they form $k$ layers of the original projections. This shows that the coarse-grained shrinking exponent is actually $\Delta_0 = \Delta_k$. 

195145-8
Claim IV.6. \((N^j \cdot \Pi_{\text{rest}} \Pi_{\text{odd}})^f\) can be written as a sum of, at most, \([4(j + 1) \cdot (j/2 + 1) \cdot (j/4 + 1) \cdots (1 + 1)]^f\) min-entangling operators.

Proof. Say that an operator \(C\) is \(t\)-min-entangling if there is a contiguous interval \(I \subseteq I_m\) of \(m/2^t\) projections such that \(C\) is of the form

\[ M_{i_j} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}) \cdots M_{i_2} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}) \cdot M_{i_1} \cdot (\Pi_{\text{rest}} \Pi_{\text{odd}}), \]

where a subset of, at most, \(j/2^t\) of the \(M_i\)'s are equal to \(N_1 = \sum_{i=1}^n 1 - P_i\), while each of the other \(M_i\)'s is made of projections that are either to the left of \(I\) or to the right of \(I\) (in \(I_m\)) and, in particular, does not include any projection from \(I\).

We begin by splitting \(N\) into two terms: \(N = N_L + N_R\), where \(N_L = \sum_{i=1}^{n/2} (1 - P_i)\) and \(N_R = \sum_{i=n/2+1}^{n} (1 - P_i)\). Then \(N_i = (N_L + N_R) = \sum_{i=1}^{n/2} \left( \frac{N^j}{N_L} \right)^{i-1}\) \(N_L\). Expanding across all \(\ell\) layers, we end up with \((j + 1)^f\) terms. Let us focus on one such term and assume that at layer \(i\) it has the powers \(\frac{N_L^{j-n_i}}{N_R}\). Since the total degree in \(N_L\) and \(N_R\) across the \(\ell\) layers is \(j\ell/2\), it follows that the degree of one of the two across the \(\ell\) layers must be, at most, \(j\ell/2\), and, hence, every term is a 1-min-entangling operator. Consequently, \((N^j \cdot \Pi_{\text{rest}} \Pi_{\text{odd}})^f\) can be written as a sum of at \((j + 1)^f\) 1-min-entangling operators.

Proceeding recursively, we now pick one such term. It either contains, at most, \(j\ell/2\) \(N_L\) operators or, at most, \(j\ell/2\) \(N_R\) operators. Assume, without loss of generality, it is \(N_L\). We write \(N_L = N_{LL} + N_{LR}\), and, consequently, every \(N_{LL}\) can be written as the sum of \(n_1 + 1\) terms: \(N_{LL} = N_{LL} + n_1 N_{LR} + \cdots + N_{LR}\). Therefore, on opening the product, we obtain a sum of \(\prod_{i=1}^{n_1} (n_1 + 1)\) terms such that each term has either a maximal degree of \(j\ell/4\) of \(N_{LL}\) or of \(N_{LR}\). These are all 2-min-entangling operators. Moreover, it is now easy to verify subject to the constraint \(\sum_{i=1}^{n_1} n_i \leq j\ell/2\), the total number of terms \(\prod_{i=1}^{n_1} (n_1 + 1)\) is maximized when all \(n_i\) are equal: \(\prod_{i=1}^{n_1} (n_1 + 1) \leq (j/2 + 1)^f\). To summarize, we have just shown that \((N^j \cdot \Pi_{\text{rest}} \Pi_{\text{odd}})^f\) can be written as a sum of, at most, \((j + 1)(j/2 + 1)\) 2-min-entangling operators.

Continuing in this fashion for \(l\) log \(j\) rounds we end up with a total of, at most, \([((j + 1) \cdot (j/2 + 1) \cdot (j/4 + 1) \cdots (1 + 1)]^{f}\) log \(j\)-entangling operators. At this stage, the total degree of each such operator in \(N_L\) is, at most, \((j\ell/2)\). To bound the increase in the remaining \(\log(m/j)\) rounds, we observe that subject to the constraint \(\sum_{i=1}^{n_1} n_i \leq \ell/2^k\), the expression \(\prod_{i=1}^{n_1} (n_1 + 1)\) is maximized when \(\ell/2^k\) of the \(n_i\)'s are 1 and the rest 0: \(\prod_{i=1}^{n_1} (n_1 + 1) \leq 2^\ell/2^{2k}\). It follows that the total increase in the number of terms over these rounds is bounded by \(2^2 \cdot 2^2 / 2^{2k} \cdots \leq 4^\ell\). This completes the proof of the claim.

We can now finish off the proof of the main lemma of this section. Claim IV.6 gives a decomposition of the operator \((N^j \cdot \Pi_{\text{rest}} \Pi_{\text{odd}})^f\) as a sum of, at most, \([4(j + 1) \cdot (j/2 + 1) \cdot (j/4 + 1) \cdots (1 + 1)]^f\) min-entangling operators. As noted above, to apply this result to \(\tilde{A}\) we must further multiply this number by \((j + 1)^f\), so \(\tilde{A}^f\) can be written as a sum of no more than \([4(j + 1) \cdot (j/2 + 1) \cdot (j/4 + 1) \cdots (1 + 1)]^f\) terms. It is easy to verify that for \(j \geq 2\) (which is always the case),

\[ 4(j + 1)^2 \cdot (j/2 + 1) \cdot (j/4 + 1) \cdots (1 + 1) \leq 20j^{3/2} 2^{3/2} \log^2 j. \]

The SR contribution of each such min-entangling operator at that cut that passes through its diluted column is, at most, \(D_{0}^{1/m}\), and since this column is, at most, \(m\) particles away from the bipartitioning cut, it follows from Fact II.1 that its SR contribution to the bipartitioning cut is, at most, \(D_{0}^{1/m} \cdot D_{0}^m\).

Therefore, \(\text{SR}(\tilde{A}(\phi)) \leq D_{1} \cdot D_{1} \cdot \text{SR}(\phi)\), with

\[ D_{1} = D_{0}^m, \quad D_{1} = 20j^{3/2} 2^{3/2} \log^2 j D_{0}^{1/m} \cdot \Delta_{0} + (1/9)^{f}. \]

This concludes the proof of Lemma IV.4.

At this point it is worthwhile to pause and take inventory of where we are. We have shown that the operator \(\tilde{A}\) is an AGSP with characteristic factors \([D_{1} \cdot D_{1} \cdot \Delta_{0} + (1/9)^{f}]\). We are searching for an AGSP whose product of characteristic factors is below \(1/2\). This will be the case in our situation for some \(\ell\) as long as the product \(D_{1} \cdot \Delta_{0} + (1/9)^{f}\) is less than \(1\).

We work so far has shown

\[ D = [\Delta_{0} + (1/9)^{f}] = 20j^{3/2} 2^{3/2} \log^2 j D_{0}^{1/m} \cdot \Delta_{0} + (1/9)^{f}. \]

Recalling that \(j = q \sqrt{m}\) it is clear that we can find constants \(q, m\) (in \(D_{0}\)) such that \(D \cdot (1/9)^{f} < 1/2\); however, the term \(D \cdot \Delta_{0}\) may still be bigger than \(1\).

To solve this problem another idea is needed: coarse graining. As we shall see in the next subsection, fusing together \(k\) adjacent particles allows us to move to a new local Hamiltonian system with \(D_{0}\), \(\Delta_{0}\) replaced by \(D_{k}^{0}, \Delta_{k}^{0}\). Taking \(k = \mathcal{O}(q \log \Delta_{0}^{-1})\) would then yield \(\Delta_{k}^{0} \approx (1/9)^{f}\) and, consequently, lead to \(\tilde{D} \cdot \tilde{\Delta} < 1/2\).

E. Coarse graining

Consider a \(k\)-coarse-grained system, in which we fuse together \(k\) adjacent particles, making them a single particle of dimension \(d^k\). The new Hamiltonian of the system would now be a 2-local Hamiltonian on a chain, consisting of projections \(Q_{i} = \mathbb{1} - P_{i}^{r}\), where \(P_{i}^{r}\) denotes the projection into the common ground space of the 2\(k\) particles that form the coarse-grained particles \(i\) and \(i + 1\). We define the odd/even layers projections \(\Pi_{\text{odd}}, \Pi_{\text{even}}\) accordingly and note that every application of \(\Pi_{\text{even}} \Pi_{\text{odd}}\) increases the SR by, at most, a factor of \(D_{k}^{0}\). To estimate the shrinking factor, we may use the DL on the new Hamiltonian using the new spectral gap.

However, there is a much stronger bound that we can obtain by using the DL on the original system. Based on an idea that has already appeared in a related form in Ref. 14, we show the following:

Claim IV.7.

\[ \Pi_{\text{odd}} \Pi_{\text{even}} = \Pi_{\text{odd}} \cdot (\Pi_{\text{even}} \Pi_{\text{odd}} \Pi_{\text{odd}} \Pi_{\text{even}})^{k/2} \cdot \Pi_{\text{even}}. \]

This immediately implies the following:
Corollary IV.8. If $H = \sum_i H_i$ is the $k$-coarse-grained version of $H = \sum_i H_i$, then the SR and shrinking factors of the DL AGSP of $H'$ are related to those of $H$ by

$$D_i^0, \Delta_i^0 = \left(D_i^0, \Delta_i^0\right).$$

The claim follows from two observations:

1. For any coarse-grained constraint $P_i'$, we can always “pull” from it a product of the original projections $P_i$ that act on the support of $P_i'$. The reason being that $P_i'$ projects into the common ground space of all the original $2k$ particles in its support, and the original projections $P_i$ are trivial on that space.

In particular, let $P_{i|k}$ denote the “pyramid” or “light-cone” of original projections pulled from $P_i'$ as in the figure above. Then $P_{i|k} = \Pi_{odd} \cdot P_{i|k} \cdot P_{i|k} \cdot P_{i|k} \cdots$ and $P_{i|k} = \Pi_{even} \cdot P_{i|k} \cdot P_{i|k} \cdot P_{i|k} \cdots = (\Pi_{odd} \Pi_{even})^{k+1}$. This is due to their pyramid-like shape; they can be commuted past each other and rearranged as $k + 1$ layers of the original projections (see the figure above). We note that applying $k+1$ original layers of the DL shrinks the perpendicular space by a factor of $\Delta_i^0$.

F. Gluing it all together

If we, first, $k$-coarse-grain the system and then construct $\hat{A}$, we obtain the following factors:

$$D_t = D_0^{km},$$
$$\hat{D} = j^{3/2} \frac{1}{2} \log j \frac{D_0}{D_0^{m}},$$
$$\hat{\Delta} = \Delta_0^k + (1/9)^q.$$

We have three free parameters: $m, q$, and $k$ (recall that $j = q \sqrt{m}$). To finish the proof, we show how these can be chosen to obtain $\hat{D} \cdot \hat{\Delta} < 1/2$.

Our first step is to demand that $k$ is large enough such that $\Delta_0^k \leq (1/8)^q$. Looking at Eq. (6), it is easy to verify that as long as $\epsilon \leq 10$, this can achieved by defining, for example,

$$k \overset{\text{def}}{=} \frac{20q}{\epsilon}.$$

Then $\hat{\Delta} \leq 2(1/8)^q$. A sufficient condition for $\hat{D} \cdot \hat{\Delta} < 1/2$ is, therefore,

$$j^{3/2} \frac{1}{2} \log j \frac{D_0}{D_0^{m}} 2^{3q} \leq \frac{1}{80},$$

or, equivalently,

$$\frac{3}{2} \log j + \frac{1}{2} \log^2 j + 40X \frac{j}{m} q - 3q \leq -\log(80),$$

where we used the definition $X \overset{\text{def}}{=} \frac{\log j}{\epsilon} = \frac{\log D_0}{\epsilon}$ (see Theorem I.1).

To satisfy this equation, we demand that $40X \frac{j}{m} \leq 2$ so the leading term in the left-hand side of Eq. (44) would be $-q$. Substituting $j = q \sqrt{m}$ leads us to define

$$m \overset{\text{def}}{=} 20qX^2.$$
We first note that up to some unimportant geometrical factors, the DL works also in 2D (and in any dimension for that matter; see Ref. 14 for more details). For simplicity, let us assume that also in the present case we only have two layers and it is only one layer that increases the SR with respect to the cut, which will which we will still refer to as the “even layer.” The idea is then to mimic the 1D case and replace a segment of $m$ columns around the cut, which we denote by $I_m$ with the operator $\hat{N}_m$. The difference is that now $I_m$ contains $m \cdot I$ projections instead of $m$. Therefore, the polynomial that we would use would be $\{|C_{m,l}(x)|\}^m$, where $C_{m,l}(x)$ is based on the Chebyshev polynomial of degree $\sqrt{m}$. Just as in the 1D case, the shrinking factor $\hat{A}$ of $\hat{A}$ is given by $\hat{A} = \Delta_0 + (1/9)^t$.

What is the SR exponent $\hat{D}$ of this construction? A very similar analysis to that in Sec. IV D can be done here: We recursively divide $I_m$ by cutting it in parallel to the cut. Just as in the 1D case, there are $m$ such possible cuts in $I_m$. The difference is that now the SR contribution of the restriction of $\hat{A}$ to a certain cut is not $D_0$ but $O(1) \cdot I \cdot D_0$, because $\hat{A}$ contains a sum of $O(1) \cdot I$ projections along a cut that is parallel to the boundary. Therefore, the overall SR after $\ell$ layers is

$$SR(\hat{A}^\ell | \phi) \leq D_1 \cdot \hat{D}^\ell,$$  \hspace{1cm} (50)

with

$$\hat{D} \equiv 20 j^{3/2} 2^{j} \log^2 (I D_0) / j m,$$ \hspace{1cm} (51)

$$D_1 \equiv (D_0^m 1),$$ \hspace{1cm} (52)

and $j \equiv q \sqrt{m}$. In comparison with the 1D indices in Eqs. (34) and (35), we see that $D_0 \mapsto D_0^m$ in the formula for $D_1$, but $D_0 \mapsto I D_0$ in the formula for $\hat{D}$—an exponential saving in the latter.

From here, the analysis follows essentially the same steps as the 1D case. We perform an initial $k$ coarse grainings to drive $(D_0, \Delta_0) \mapsto (D_k, \Delta_k)$, Demanding that $\Delta_k \leq (1/8)^k$, we set $k \equiv 20 q / \epsilon$, as in Eq. (41). The analog of Eq. (43) then is

$$3 \log j + \frac{1}{2} \log^2 j + (\log I + 40 X q) J / m - 3q \leq -\log(80),$$ \hspace{1cm} (53)

and demanding that $(\log I + 20 X q) J / m \leq 2q$ leads us to set $m \equiv (1/2) (\log I + 20 X q)^2 \cdot I$. Plugging this back to Eq. (53) yields $q \leq \mathcal{O}(1) \cdot \log^2 (I \cdot X)$, and so $m \leq \mathcal{O}(1) \cdot I \cdot X^2 \log^4 (I \cdot X)$. This results in $\log D_1 \leq \mathcal{O}(1) \cdot I^2 \cdot X^3 \log^8 (I \cdot X)$, and $\log \hat{D} \leq \mathcal{O}(1) \cdot \log^2 (I \cdot X)$, which brings us to

$$S \leq \mathcal{O}(1) \cdot I^2 \cdot X^3 \cdot \log^8 (I \cdot X).$$ \hspace{1cm} (54)

VI. MATRIX PRODUCT STATES

We briefly sketch the implications of these results for the efficient approximation of the ground state via matrix product states. Specifically, Theorem I.1 implies the following corollary:

**Corollary VI.1.** Under the same conditions of Theorem I.1, for any integer $k > 0$ there exists a matrix product state (MPS) $|\psi_k\rangle$ with bond dimension $k$ such that

$$\|\Omega - |\psi_k\rangle\|^2 \leq 2\mathcal{O}(1) X^1 \log^8 X (n/k).$$ \hspace{1cm} (55)

This result follows from, first, noting that across any cut, we can use the properties of the AGSP $K$ to bound the norm of the tail of the Schmidt coefficients: $\sum_{i>j} \lambda_i^2 \leq 1/\mu \Delta' \leq 2D \Delta'$, where $\mu$ is the overlap of the initial product state with the ground state (see Lemma III.2 in Sec. III and the proof of Lemma III.3 in the appendix). Then, letting $k \equiv D'$, we find that if we truncate the Schmidt coefficient of a given cut at $k$, we introduce an error of $\delta \equiv \sum_{i>k} \lambda_i^2 \leq 2D$, where we used the fact that $D \cdot \Delta \leq 1/2$. Then, applying the MPS construction procedure of Vidal25 and truncating the SR across all cuts to $k$, yields an MPS of bond dimension $k$ that approximates the ground state to within the accumulated error $\delta = 2n \epsilon^2$. Finally, recalling from Eq. (16) in Lemma III.5 that $\log D \leq \mathcal{O}(1) \cdot X^1 \log^8 X$ gives Eq. (55).

VII. CONCLUSIONS

In conclusion, we have given a new approach to proving the area law for gapped and frustration-free 1D systems. Our proof uses the DL and the Chebyshev polynomials to upper bound the entropy by $\mathcal{O}(1) \cdot X^3 \log^8 X$, which is exponentially better than the bound in Ref. 13. It brings us much closer to the recent lower bound of $\mathcal{O}(1) \cdot \epsilon^{-1/4}$ (for fixed $d$) of Hastings and Gottesman16 and Irani.17

There are two immediate directions in which one might hope to improve this result. First, it seems very plausible that the proof can be extended to the frustrated case. Indeed, already in Hastings’ 1D proof,13 one of the first steps of the proof is to reduce the frustrated system into an almost frustration-free system by use of a coarse-graining procedure. It is possible that a similar technique can be also deployed here. Moreover, one might try to take a more direct approach and construct the AGSP directly from the Hamiltonian $H$ by replacing $\mathcal{N}$ with $H$. Both are sums of projections, and a similar SR analysis can be performed on operators of the form $\text{poly}(H)$.

The second direction, which is much more interesting, and probably more difficult, is to try to generalize the area law for 2D and beyond. In fact, any subvolume law for 2D would be an extremely interesting result. One possibility is to improve the $\log d$ dependence of the bound in Eq. (3). A bound linear in $\log d$ would imply an area law in all dimensions, whereas anything like $(\log d)^2$–3 would imply a subvolume law for low dimensions. However, it seems to us that the right approach is to better understand and exploit the locality of the problem in the direction parallel to the cut. This was done in a very weak way in Sec. V and led to an improved bound in the 2D case. We believe there are more local aspects of the problem that can be used. For example, in the current AGSP construction we do not assume anything about the underlying distribution of violations. Yet, these arise from a very specific local operation, namely the application of the previous AGSPs. If one could show that the distribution of violations decays exponentially in $k$, it may be enough to use a Chebyshev polynomial of a degree smaller than $\sqrt{m}$, which may lead to an area law.

More generally, one might want to prove some notion of independence, or decay of correlation, along the cut, thereby...
reducing the 2D problem to a stack of nearly independent 1D problems.

ACKNOWLEDGMENTS

We are grateful to Dorit Aharonov, Matt Hastings, Sandy Irani, Tobias Osborne, and Bruno Nachtergaele for inspiring discussions about the above and related topics. Itai Arad acknowledges support by Julia Kempe ERC Starting Grant QUOC and Julia Kempe Individual Research Grant of the Israel Science Foundation (Grant No. 759/07). Zeph Landau and Umesh Vazirani were supported in part by ARO Grant No. W911NF-09-1-0440 and NSF Grant No. CCF-0905626.

APPENDIX: BOUNDING THE ENTROPY OF A STEPLIKE PROBABILITY DISTRIBUTION

Proof of Lemma III.3. Let \(|\Omega| = \sum_{i=1}^{\lambda} |L_i| \otimes |R_i|\) be the Schmidt decomposition of the ground state \(|\Omega|\), and let \(|\phi\rangle = |L_i| \otimes |R_i|\) a product state such that \(|\langle \phi |\Omega\rangle| = \mu\). Define the sequence of states \(|\psi_i\rangle\) to be the normalization of the vectors \(K^i|\phi\rangle\). Since \(K\) is a \((D, \Delta)\)-AGSP, it follows that \(|\psi_i\rangle\) has the following properties:

1. \(S_R(\psi_i) \leq D^\ell\).
2. \(|\psi_i|\Omega|\rangle\geq \frac{\mu}{\mu^2 + \Delta^\ell(1-\mu^2)}\).

Then, by Fact II.2 (Eckart-Young theorem),

\[
\sum_{i>D^\ell} \lambda_i^2 \leq \frac{\mu^2}{\mu^2 + \Delta^\ell(1-\mu^2)} \leq 1 - \frac{\mu^2}{\mu^2 + \Delta^\ell} \leq \frac{1}{\mu^2} \Delta^\ell \text{ def } = p_\ell.
\]

We will use this bound to upper bound the entropy of the \(\{\lambda_i^2\}\). Choose \(\ell_0 = \lceil \log \frac{\mu^2}{\Delta^\ell} \rceil\) so that \(p_\ell < 1\). For \(\ell \geq 2\ell_0 + 1\), we will upper bound the entropy of the distribution with the bounds

\[
\sum_{j=D^\ell+1}^{D^{2\ell+1}} \lambda_j^2 \leq p_{2\ell} \leq p_\ell.
\]

This is maximized by spreading out the probability mass uniformly in each interval \([D^{2\ell} + 1, D^{2(\ell+1)}]\), which results in an upper bound on the contribution of the entropy on the interval \([D^{2\ell} + 1, D^{2(\ell+1)}]\) of

\[
p_\ell \log \frac{D^{2(\ell+1)} - D^{2\ell}}{p_\ell} \leq p_\ell \log \frac{D^{2(\ell+1)}}{p_\ell} \leq \Delta^{\ell-\ell_0} \log \frac{D^{4(\ell+1)}(\ell - \ell_0)}{\Delta}.
\]

(A1)

Above, the second inequality follows from \(p_\ell < 1\) along with the fact that the function \(x \log\frac{x}{\Delta}\) is increasing for \(x \leq 1\).

Summing Eq. (A1) over \(\ell \geq 2\ell_0 + 1\) then yields an upper bound of the entropy contribution of this tail as

\[
\sum_{\ell \geq 2\ell_0 + 1} \Delta^{\ell-\ell_0}(\ell - \ell_0) \log \frac{D^{2(\ell+1)}(\ell - \ell_0)}{\Delta} \\
\leq \sum_{\ell \geq 2\ell_0 + 1} \Delta^\ell \log \left(\frac{D^4}{\Delta}\right) \leq \frac{\Delta}{(1-\Delta)^2} \log \left(\frac{D^4}{\Delta}\right).
\]

Above, the first inequality follows from, first, noting that for the choice of \(\ell\), the exponent on the power of \(D\) is upper bounded by 4, then making the substitution of \(\ell' = \ell - \ell_0\). The second follows from using the series equality \(\sum_{i \geq 1} x^i = \frac{1}{1-x}\).

Combining this estimate with the maximal contribution of entropy over the first \(D^{2(\ell_0+1)}\) terms, namely \((2\ell_0 + 1) \log D = O(1) \cdot \log \left(\frac{\mu^{-1}}{\log \Delta} \right) \log D\), gives the following bound on the entropy:

\[
S \leq O(1) \cdot \frac{\log \left(\frac{\mu^{-1}}{\log \Delta} \right)}{\log \Delta} \log D + \frac{\Delta}{(1-\Delta)^2} \log \left(\frac{D^4}{\Delta}\right).
\]

To simplify this expression, note that for any integer \(k \geq 1\), \(K^k\) is an AGSP with characteristic factors \((D^k, \Delta^k)\). Choosing \(k = \lceil \log \frac{1}{\log \Delta} \rceil\) ensures that \(\frac{1}{4} \leq \Delta^k \leq \frac{1}{2}\). Substituting the parameters \(D^k\) for \(D\) and \(\Delta^k\) for \(\Delta\) in the above expression then yields

\[
S \leq O(1) \cdot \log \left(\frac{\mu^{-1}}{\log \Delta} \right) \log D + \log D^{4k} + 1
\leq O(1) \cdot k \log \left(\frac{\mu^{-1}}{\log \Delta} \right) \log D
\leq O(1) \cdot \log \left(\frac{\mu^{-1}}{\log \Delta} \right) \log D.
\]

\[\square\]
The examples in these papers are of frustrated Hamiltonians, whereas our results apply for frustration-free systems. However, the construction in Ref. 17 is translation invariant. It can be easily modified to be made frustration-free\(^2\) if the translation-invariance requirement is dropped, yielding a system in which the entanglement entropy \(S\) is lower bounded by \(\epsilon^{1/6}\), where \(\epsilon > 0\) is the spectral gap of the system.

Without loss of generality, we may absorb any phases of \(\mu\) into \(\phi\) and assume \(\mu\) is real and positive.

At first sight it seems that the shrinking should be of \(\Delta_0^{(k+1)/2}\), but actually we can do much better by “duplicating” the projections in the middle. For example, for \(k = 3\), we get \(\Pi_{\text{even}}\Pi_{\text{odd}}\Pi_{\text{even}}\Pi_{\text{odd}} = (\Pi_{\text{even}}\Pi_{\text{odd}})(\Pi_{\text{odd}}\Pi_{\text{even}})(\Pi_{\text{even}}\Pi_{\text{odd}}).\) Every bracket contributes a \(\Delta_0\) factor, so, overall, we get \(\Delta_0^n\).