The 1D Area Law and the Complexity of Quantum States: A combinatorial approach

Dorit Aharonov
School of Computer Science and Engineerig
The Hebrew University
Jerusalem, Israel
dorit.aharonov@gmail.com

Itai Arad
School of Computer Science and Engineerig
The Hebrew University
Jerusalem, Israel
arad.itai@gmail.com

Zeph Landau
Computer Science division
University of California at Berkeley
Berkeley, CA, USA
zeph.landau@gmail.com

Umesh Vazirani
Computer Science division
University of California at Berkeley
Berkeley, CA, USA
vazirani@eecs.berkeley.edu

Abstract— The classical description of quantum states is in general exponential in the number of qubits. Can we get polynomial descriptions for more restricted sets of states such as ground states of interesting subclasses of local Hamiltonians? This is the basic problem in the study of the complexity of ground states, and requires an understanding of multi-particle entanglement and quantum correlations in such states.

Area laws provide a fundamental ingredient in the study of the complexity of ground states, since they offer a way to bound in a quantitative way the entanglement in such states. Although they have long been conjectured for many body systems in arbitrary dimensions, a general rigorous was only recently proved in Hastings' seminal paper [8] for 1D systems. In this paper, we give a combinatorial proof of the 1D area law for the special case of frustration free systems, improving by an exponential factor the scaling in terms of the inverse spectral gap and the dimensionality of the particles. The scaling in terms of the dimension of the particles is a potentially important issue in the context of resolving the 2D case and higher dimensions, which is one of the most important open questions in Hamiltonian complexity.

Our proof is based on a reformulation of the detectability lemma, introduced by us in the context of quantum gap amplification [1]. We give an alternative proof of the detectability lemma, which is not only simpler and more intuitive than the original proof, but also removes a key restriction in the original statement, making it more suitable for this new context. We also give a one page proof of Hastings' proof that the correlations in the ground states of gapped Hamiltonians decay exponentially with the distance, demonstrating the simplicity of the combinatorial approach for those problems.

Keywords-area law, ground state, Hamiltonian, description complexity, detectability lemma, entanglement

1. INTRODUCTION

While the exponential description of quantum states (the fact that 2^n complex numbers are necessary to describe the general state of an n qubit system) leads to the extraordinary computational power of quantum systems, it is also a fundamental source of difficulty in simulating and analyzing quantum systems. Are there important classes of quantum states that have polynomial descriptions? A natural class to consider is ground states of local Hamiltonians, $H = H_1 + \ldots + H_m$, where each term H_i is a positive definite matrix which acts non-trivially on at most k qubits. If the ground state is unique (the minimum eigenvalue of H has multiplicity 1), then it can be efficiently specified by H, whose description is $O(mn^k 2^k)$. However, this is not a useful description since, for example, even the problem of computing the energy (eigenvalue) of the ground state, given H, is QMA-complete. What we want is a description of the quantum state that allows an efficient implementation of any k qubit measurement on the quantum state. The flip side of this problem occurs if we describe the ground state by specifying the reduced density matrices on each subset of k qubits. Now it is easy to implement any k qubit measurement. The problem is that it is hard to verify whether the $\binom{n}{k}$ density matrices are consistent with a quantum state on the entire system, i.e., whether they can be glued together into one state.

Formally, a set of n qubit quantum states S have an efficient description scheme, if there are two polynomial time procedures: CHECK and MEASURE, and a polynomial length bit string x encoding each quantum state in S, with the following property: if CHECK(x) = 1 then x is an encoding of a valid quantum state on n qubits. MEASURE(x, M) is the outcome of performing measurement M on the state represented by x, for any k qubit measurement M. Of course, the definition can be generalized to consider approximate versions of description schemes.

The challenge of classifying which quantum state sets have efficient description schemes is not only fundamental to quantum complexity theory, it is also of major importance to condensed matter physics. The question is of particular interest when attention is restricted to the set of states which are ground states of local Hamiltonians; in other words, for which sets of local Hamiltonians do the ground states have efficient description schemes, such that analysis and study of

these states becomes possible. What makes these questions particularly challenging is the phenomenon of multi-particle entanglement, which is far from understood quantitatively, and is responsible for the exponential description of general quantum states.

A local Hamiltonian is said to be gapped if its spectral gap – the difference between the lowest and second the lowest eigenvalues - is constant . This is a particularly important class of Hamiltonians since many naturally occurring Hamiltonians are gapped. The ground states of gapped local Hamiltonians are believed to obey *area laws*, which provide a way of bounding the entanglement in ground states of such Hamiltonians. Consider the interaction graph (hypergraph) associated with a local Hamiltonian — it has a vertex for each particle and an edge (hyper-edge) for each term of the Hamiltonian. Roughly, an area law says that entanglement is local in this interaction graph in the following sense: consider a subset of particles L. Then if the system is in its ground state, then the entanglement entropy across the cut $(L, \bar{L})$ is big-Oh of the number of edges crossing between L and $\bar{L}$.

A few years ago, in a seminal paper [8], Hastings proved that the area law holds for 1D systems, if the Hamiltonian is gapped. In this case the interaction graph is a path, and the area law says that ground state entanglement across any contiguous cut is bounded by a constant. Previously, Vidal has proved that when the entanglement in a 1D system is bounded by a constant along every contiguous cut, an efficient description scheme of the ground state exists, in the form of what is called Matrix Product states (MPSs) with bounded bond dimension [11]. Hastings' result thus implies, via the results of Vidal, that for ground states of 1D gapped Hamiltonians, efficient description schemes exist.

A major open problem is to extend Hastings' result to higher dimensions, as well as to extend the description schemes assuming such bounds on the entanglement. The problem of extending Hastings result to higher dimensions was tackled so far mainly by physicists. One reason is that Hastings' proof uses sophisticated analytic methods common to the condensed matter physics literature such as the Lieb-Robinson bound. This is a bound on the velocity at which disturbances propagate in local many-body quantum systems, which has proved to be extremely useful for understanding their structure.

1.1. Results

In this paper we introduce a combinatorial approach to the area law question, and to the related question regarding quantum correlations in ground states. Our approach is based on a reformulation (together with a much simplified proof), of the Detectability Lemma (DL) introduced by the authors in Ref. [1] in the different context of the quantum PCP challenge. The DL provides an approximation to the projection on the ground space, which has a simple product structure that is amenable to analyzing the generated entanglement. We believe that the combinatorial nature of the DL makes it a useful tool to tackle questions related to quantum correlations in ground states of local Hamiltonians.

As a first example of how the DL can be used to handle correlations, we provide a one page proof of Hastings' result [6] that the correlations in the ground states of Gapped Hamiltonians decay exponentially with the distance. This applies to d-dimensional grids for any d. More precisely, consider two local observables A and B, which act on sets of particles that are of distance ℓ on the grid; the decay of correlations means that the expectation value of their product is (almost) equal to the product of their expectation, up to an error which decays exponentially with the grid distance between those operators.

We then apply the reformulation of the DL to the much more involved problem of the area law in 1D. We provide a proof of the area law for 1D frustration-free gapped Hamiltonians. Using an intricate combinatorial analysis, we derive a result which is almost exponentially better than that of Hastings [8], in terms of the scaling in the inverse spectral gap and the dimensionality of the particles. Whereas Hastings' upper bound scales as $e^{\mathcal{O}(X)}$ for $X \stackrel{\text{def}}{=} \frac{1}{\epsilon} \log d$, our upper bound scales quasi-polynomially as $X^{\mathcal{O}(\log X)} = e^{\mathcal{O}(\log^2 X)}$.

For a fixed d, a simple lower bound was known via a state whose entropy scales logarithmically with those parameters [2], [3]. Recently, tighter examples were found by Gottesman and Hastings [5], and independently by Irani [9], in which the entanglement scales polynomially with the inverse spectral gap[1]. Our improved bounds therefore reduce an exponential mismatch (in terms of the spectral gap) between the upper and lower bounds, to a quasi-polynomial one

In addition, the resulting proof is combinatorial and completely bypasses the analytic techniques that seemed to be a major barrier for computer scientists' participation in this important aspect of Hamiltonian complexity. It also raises the hope that similar tools can be used to tackle the 2D case, which is one of the major open questions in condensed matter theory. Indeed, if the quasi-polynomial bound in $X = \frac{1}{\epsilon} \log d$ in the entanglement entropy for the 1D case were improved to linear in X, the area law for 2D would follow.

One might imagine that this exponential decay of correlations implies that the entanglement between a region L and its complement is (mostly) located at the boundary of L, and thus obeys an area law. However, such an implication is not

---

[1]The examples in these papers are of a frustrated Hamiltonians, whereas our results apply for frustration-free systems. However, the construction in Ref. [9] is translation-invariant. It can be easily modified to be made frustration-free [10] 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.

known, and indeed, a beautiful example by Hastings shows a matrix product state (MPS), constructed using quantum expanders, in which the entanglement entropy is exponential in the correlation length [7], [8].

Our results hold in the frustration-free case. We mention, however, that many of the quantum phenomenon are revealed already in the context of frustration-free Hamiltonians, and that the major open problems in quantum Hamiltonian complexity (e.g, quantum PCP and 2D area law) are wide open already for this case. Much is to be learned from studying frustration-free Hamiltonians, before we proceed to the more general case. Moreover, Hastings' original proof [8] essentially reduces the frustrated case into an approximately frustration-free system by coarse graining, and a similar approach may work to extend the results presented here to the general case.

We proceed to an overview of the DL and the proof of the area law.

1.2. The Detectability Lemma (DL)

Consider a gapped frustration-free local Hamiltonian $H = \sum_{i=1}^{m} H_i$ with $0 \leq H_i \leq 1$, in which the ground energy is 0, and the spectral gap is $\epsilon = \mathcal{O}(1)$. Denote the ground state by $|\Omega\rangle$. We would like to understand "local" properties of the ground state, such as behavior of correlations between distant particles in the ground state, propagation in space of perturbations, etc. To this end we would like to mimic the projection on the ground state, $\Pi_{gs} = |\Omega\rangle\langle\Omega|$, by a "local" operator (for some notion of locality to be clarified soon).

One way to approach this question is to ask: is there a local operator that fixes the ground state but shrinks all other eigenvectors significantly? A natural first guess for such an operator is the operator $G = (\mathbb{1} - \frac{1}{m}H)$. It fixes the ground space, and shrinks its orthogonal complement by a factor that can be as large as $(1 - \epsilon/m)$. As m increases the factor approaches 1 and the shrinking becomes negligible. This is not strong enough for various purposes, as we will soon see; our challenge is to do better. Suitably reformulated, the DL provides an operator which fixes the ground state but shrinks all other eigenstates by a *constant* factor, independent of m.

For simplicity of the current discussion, let us explain the exact statement under the assumption that the particles are on a 1D chain, and that H_i are projections Q_i that act on adjacent particles, so that $H = \sum_{i=1}^{n-1} Q_i$. Set P_i to be the projections into the local ground spaces of H_i, i.e., $P_i = \mathbb{1} - Q_i$. The $\{Q_i\}$ terms can be partitioned into two sets, the even and odd terms, which we call layers (see Fig. 1), and the projections into the common ground spaces of these layers are given by $\Pi_{odd} \stackrel{\text{def}}{=} P_1 P_3 P_5 \cdots$ and $\Pi_{even} \stackrel{\text{def}}{=} P_2 P_4 P_6 \cdots$. The DL states that the operator

$$A \stackrel{\text{def}}{=} \Pi_{even}\Pi_{odd} \tag{1}$$

is the "local" operator we want. We can now explain what we mean by "locality": when A is applied ℓ times to some
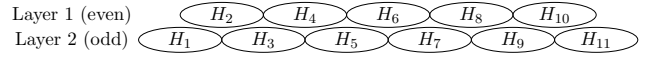

Figure 1. Partitioning the local terms of the 1D-, nearest-neighbor Hamiltonian $H = \sum_{i=1}^{n-1}$ into two subsets (layers): even layer, consisting of terms H_i that work on particles $i, i+1$ with even i, and a complementary odd layer.
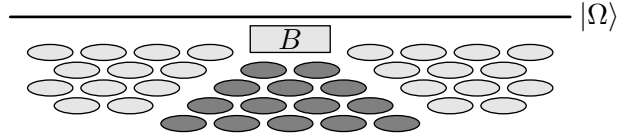

Figure 2. The "causality cone" of projections that is defined by a local operator B working on $\ell = 4$ layers. These are the local terms in the layers that are graph-connected to B (shaded ovals).

local perturbation B that acts on the ground state $|\Omega\rangle$, there is a pyramid-shaped "causality cone" of projections that is defined by B. These are simply all terms which are graph-connected to the operator in question (see Fig. 2). All the projections outside that cone commute with B and can therefore be absorbed in the ground state, leaving us with a *local* operator of size $\mathcal{O}(\ell)$.

The DL implies that A^ℓ can be viewed as an excellent approximation to the ground state projection Π_{gs}. Indeed, as A shrinks all the states that are perpendicular to the ground space by a constant factor, it follows that :

$$\|A^\ell - \Pi_{gs}\| \leq e^{-\mathcal{O}(\ell)} . \tag{2}$$

In our new simplified proof of this reformulation of the DL, we also manage to drop the additional assumption about the number of distinct types of terms of the Hamiltonian that was used in [1].

1.3. Overview of the area law in 1D

We consider some cut in the chain, and we would like to bound the entropy of $|\Omega\rangle$, the unique ground state of a local gapped 1D Hamiltonian, along that cut. As a warm up let's start by considering the commuting case, where proving the area law is simple. Indeed, if the Schmidt decomposition of $|\Omega\rangle$ along the cut is $|\Omega\rangle = \sum_i \lambda_i|L_i\rangle \otimes |R_i\rangle$, we can start with the product state $|\phi\rangle \stackrel{\text{def}}{=} |L_1\rangle \otimes |R_1\rangle$, which has an overlap of λ_1 with the ground state. We can then project it onto the ground state by acting on it with the product $\prod_{i=1}^{n-1} P_i$. In the commuting case all the local projections P_i commute and so this is indeed the projection to the ground state. We find that $|\Omega\rangle = \frac{1}{\lambda_1} \prod_i P_i|L_1\rangle \otimes |R_1\rangle$. But of all the projections in the product, only the projections whose support intersects with the cut actually change the Schmidt rank of the state. Moreover, as each one of those projections is a local projection over two particles of dimension d, it is easy to see that the resulting state, $|\Omega\rangle$, would have a maximal Schmidt rank of $D_0 \stackrel{\text{def}}{=} d^2$. From here, it is now

easy to deduce that the entropy of $|\Omega\rangle$ along the cut is upper bounded by $\log D_0 = 2 \log d$.

The starting point for the non-commuting case is the same. We start from the state $|\phi\rangle \overset{\text{def}}{=} |L_1\rangle \otimes |R_1\rangle$, and would like to project it onto the ground state. However, as the P_i no longer commute, their product is no longer the projection to the ground state. We could try to use the DL to approximate that projection. Applying the DL operator A ℓ times, we obtain a state $|\phi_\ell\rangle = A^\ell |L_1\rangle \otimes |R_1\rangle$, where A^ℓ is the approximation of the projection on the ground state, using Eq. (2). This provides an exponentially good approximation to the ground state; unfortunately, the growth of the Schmidt rank is also exponential in ℓ, and the tradeoff between the two is unfavorable and does not result in a useful bound. Instead, we must modify the operator A to one that approximates the projection onto the ground state while still producing significantly slower growth in Schmidt rank. The technical details are quite involved, and we defer further explanations to the proof presented in the body of the paper.

1.4. Open Questions

A technical open question is to tighten the quasi-polynomial gap (in terms of the spectral gap) between the 1D area law and the lower bounds provided by Gottesman-Hastings and Irani.

Of course, the most important open question is to extend the proofs to the 2D case and beyond. One approach is to improve the 1D area law in terms of dependence on the dimensionality of the particles d, to $O(\log d)$, so that when the 2D structure is viewed as 1D (for example, by fusing together particles along columns parallel to the cut), the 1D bound would suffice to prove a 2D area law. A more promising avenue would also take into account the local structure of the problem along the dimension of the cut. For example, it might be helpful to understand the decay of correlations *along* the cut.

It also remains open to clarify how area laws can be used to provide more efficient description schemes in dimensions higher than 1.

2. Notations and Preliminaries

We consider a local Hamiltonian H acting on $\mathcal{H} = (\mathbb{C}^d)^{\otimes n}$, the space of n qudits of dimension d. $H = \sum_i Q_i$ where each Q_i acts non-trivially on a constant number of qubits (hence the term local Hamiltonian). We assume that H has a unique ground state of energy 0, which is also a common zero eigenvector of all terms Q_i; this means that H is frustration free. We also assume that H is "gapped", i.e., H's smallest eigenspace is 1 dimensional and has eigenvalue 0, and all other eigenvalues are equal or larger than some constant $\epsilon > 0$. We denote by $\mathcal{H}' \subset \mathcal{H}$ the orthogonal complement to the one dimensional ground space of H.

Thus $\mathcal{H}'$ is an invariant subspace for H, and

$$H|_{\mathcal{H}'} \geq \epsilon \mathbb{1}. \tag{3}$$

Throughout this extended abstract we further assume that the Q_i's are projections. The projection restriction can be lifted easily to general positive semi-definite matrices. P_i is defined to be the projection on the ground space of Q_i and so we have in this case $P_i = \mathbb{1} - Q_i$.

Given a state $|\phi\rangle$ and a partition of the qubits to two non-intersecting sets, R and L, with corresponding Hilbert spaces $\mathcal{H}_L, \mathcal{H}_R$, we can consider the Schmidt decomposition of the state along this cut $|\phi\rangle = \sum_i \lambda_i |L_i\rangle \otimes |R_i\rangle$. We shall be interested in the non-zero Schmidt coefficients, which will be ordered in a descending order: $\lambda_1 \geq \lambda_2 \geq \ldots$. The *Schmidt rank* (SR) of $|\phi\rangle$ is then the number of non-zero coefficients λ_i, and will be denoted by $\text{SR}(\phi)$. The *entanglement entropy* is the entropy of the set $\{|\lambda_i|^2\}$, or equivalently, the Von Neumann entropy of the reduced density matrix $\rho_L(\phi)$.

Finally, we mention an important fact about the SR which follows from Eckart-Young theorem [4]. It states that the truncated Schmidt decomposition provides the best approximation to a vector in the following sense:

Fact 2.1: Let $|\phi\rangle$ be a vector on $\mathcal{H}_L \otimes \mathcal{H}_R$ with Schmidt coefficients $\lambda_1 \geq \lambda_2 \geq \ldots$. The largest inner product between $|\phi\rangle$ and a norm one vector with SR r is $\sqrt{\sum_{i=1}^r \lambda_i^2}$.

3. The detectability lemma: Statement and a new proof

For the sake of readability, we will state and prove the lemma under the assumption that the qudits are set on a line and that the interaction terms are two-local nearest neighbors. We partition the set of projections $\{P_i\}$ into two sets: $\{P_1, P_3, \ldots\}$ and $\{P_2, P_4, \ldots\}$ which we shall call the odd and even layers (see Fig. 1). Notice that all the projections within each layer commute. We shall denote $\Pi_{odd} = P_1 P_3 \cdots$, (respectively $\Pi_{even} = P_2 P_4 \cdots$) the product of the projections in the odd (respectively even) layer. Let $A = \Pi_{odd}\Pi_{even}$. Notice that A fixes the ground state of H. In this case the DL states:

Lemma 3.1 (The Detectability Lemma (DL) in $1D$): The operator $A = \Pi_{odd}\Pi_{even}$, when restricted to the invariant subspace $\mathcal{H}'$, is norm bounded above as follows:

$$\|A|_{\mathcal{H}'}\| \leq \frac{1}{(\epsilon/2 + 1)^{1/3}} \ .$$

Note that the above bound holds also when the ground state is not unique.

Proof: The following is a simple but crucial fact for the proof, upper bounding the product of a projection Y, followed by another projection X, followed by the complement of Y, $\mathbb{1} - Y$:
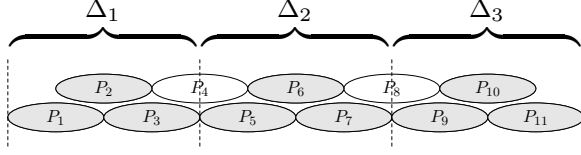
Figure 3. A set of pyramids (gray ovals) in the 1D case. The white ovals are the projections that make up the operator R. The pyramid-like structure guarantees that $A = (P_2 \cdot P_4 \cdot P_6 \cdots) \cdot (P_1 \cdot P_3 \cdot P_5 \cdots) = \Delta_1 \Delta_2 \ldots \Delta_m R$.

Lemma 3.2: Given arbitrary projections X, Y and $|v\rangle$ of norm 1, if $\|XYv\|^2 = 1 - \epsilon$ then

$$\|(\mathbb{1} - Y)XYv\|^2 \leq \epsilon(1 - \epsilon) . \qquad (4)$$

The proof of the above lemma is a few lines of simple algebra.

Let us now continue with the proof of the DL. Consider $|\psi\rangle \in \mathcal{H}'$, a norm 1 state that is orthogonal to the ground space (and therefore has energy at least the spectral gap ϵ) and define $|\phi\rangle \overset{\text{def}}{=} A|\psi\rangle$. We want to show that

$$\|\phi\| \leq \frac{1}{(\epsilon/2 + 1)^{1/3}} . \qquad (5)$$

We will show upper and lower bounds for the energy $\langle\phi|H|\phi\rangle$. The lower bound is easy since $|\phi\rangle \in \mathcal{H}'$, as A preserves the ground space, and so

$$\langle\phi|H|\phi\rangle \geq \epsilon\|\phi\|^2 . \qquad (6)$$

The main effort is to upper bound $\langle\phi|H|\phi\rangle$. We will show that $\langle\phi|H|\phi\rangle \leq 2(1 - \|\phi\|^2)/\|\phi\|$, which implies the inequality

$$\epsilon\|\phi\|^2 \leq \langle\phi|H|\phi\rangle \leq 2\frac{1 - \|\phi\|^2}{\|\phi\|} . \qquad (7)$$

This implies $\epsilon\|\phi\|^3 \leq 2(1 - \|\phi\|^3)$ from which Eq. (5) immediately follows.

We begin by noting that $\langle\phi|H|\phi\rangle$ is a sum over the contributions of the individual terms $\langle\phi|Q_i|\phi\rangle$. These terms equal to 0 for i odd since $A = \Pi_{odd}\Pi_{even}$ and $Q_i\Pi_{odd} = 0$ for any odd i (recall that Π_{even}, Π_{odd} are products of the projections $P_i = \mathbb{1} - Q_i$). We now want to bound the contributions coming from the even terms.

For this purpose we present A in a convenient form, by reordering its terms. We call the triplet product of projections $(P_1P_3P_2), (P_5P_7P_6), \ldots$ *Pyramids*, and denote them by Δ_i; the remaining operators are combined to the operator $R \overset{\text{def}}{=} P_4P_8 \ldots$ (see Fig. 3). We can write:

$$A = \Delta_1 \Delta_2 \ldots \Delta_m R ,$$

where m is the number of pyramids (which is approximately $n/4$). This holds since the terms in R commute with the terms with which they were swapped.

We will use this reordering to bound the energy contribution of the terms $Q_2, Q_6, \ldots$; a symmetric argument will bound the remaining even terms, namely Q_4, Q_8 etc.

The key point in proving this bound is this. We view the transformation of $|\psi\rangle \rightarrow A|\psi\rangle = |\phi\rangle$ as a series of steps given by the application of the pyramids Δ_i. Let $|v_i\rangle \overset{\text{def}}{=} \Delta_i\Delta_{i+1}\cdots\Delta_m R|\psi\rangle$, with $|v_1\rangle = |\phi\rangle$. The square of the norm of the first state, $R|\psi\rangle$, (which we denote by v_{i+1}) is $a_m \overset{\text{def}}{=} \|R\psi\|^2$. Let $a_i \overset{\text{def}}{=} \|v_i\|^2/\|v_{i+1}\|^2$ be the "shrinkage" resulting from the application of the ith pyramid, for $1 \leq i < m$. We are interested in bounding $\|v_1\|^2 = \|\phi\|^2$; it is exactly the product of the a_i's.

It turns out that the shrinkage a_i is related to the energy of the operator Q at the top of the relevant pyramid Δ_i. Indeed, since this Q (whose index is Q_{4i-2}) commutes with all the pyramids up to Δ_i, we have:

$$\begin{aligned}\langle\phi|Q_{4i-2}|\phi\rangle &= \|(\mathbb{1} - P_{4i-2})A\psi\|^2 \\ &= \|\Delta_1\cdots\Delta_{i-1}(\mathbb{1} - P_{4i-2})\Delta_i v_{i+1}\|^2 \\ &\leq \|(\mathbb{1} - P_{4i-2})\Delta_i v_{i+1}\|^2 .\end{aligned}$$

Now recall that $\Delta_i = P_{4i-3}P_{4i-1}P_{4i-2}$, and so we can apply Lemma 3.2 to $(\mathbb{1} - P_{4i-2})\Delta_i\frac{v_{i+1}}{\|v_{i+1}\|}$ (with $Y = P_{4i-2}$ and $X = P_{4i-3}P_{4i-1}$). We conclude that $\|(\mathbb{1} - P_{4i-2})\Delta_i v_{i+1}\|^2 \leq (1 - \frac{\|\Delta_i v_{i+1}\|}{\|v_{i+1}\|})(1 - a_i)\|v_{i+1}\|^2 = (1 - a_i)\|v_{i+1}\|^2 \leq (1 - a_i)$. Consequently

$$\langle\phi|Q_{4i-2}|\phi\rangle \leq 1 - a_i .$$

This upper bound gives an upper bound for the energy contribution for Q_i, $i \in \{2, 6, 10, \ldots\}$:

$$\langle\phi|(Q_2 + Q_6 + \ldots)|\phi\rangle \leq \sum_{i=2,6,\ldots} (1 - a_i) ,$$

with the constraint $\prod a_i = \|\phi\|^2$. The term $\sum_i(1 - a_i)$ is maximized under this constraint when all a_i are equal, i.e., when $a_i = \|\phi\|^{\frac{2}{m}}$, and therefore we are left with the following upper bound on the energy of $Q_2 + Q_6 + \ldots$:

$$\langle\phi|(Q_2 + Q_6 + \ldots)|\phi\rangle \leq m\left[1 - \|\phi\|^{2/m}\right] \leq \frac{1 - \|\phi\|^2}{\|\phi\|} .$$

The last inequality follows from the fact[2] that for every $x \in [0, 1]$, we have $m\left[1 - x^{1/m}\right] \leq \frac{1-x}{\sqrt{x}}$.

For the energy of $Q_4 + Q_8 + \ldots$, a similar argument can be applied, and therefore the total energy is upperbounded by $2(1 - \|\phi\|^2)/\|\phi\|$.

∎

---

[2]This inequality can be easily verified by noticing that $f_m(x) \overset{\text{def}}{=} m\sqrt{x}\left[1 - x^{1/m}\right] + x$ is equal to 1 for $x = 1$ and has a non-negative derivative for $x \in [0, 1]$.
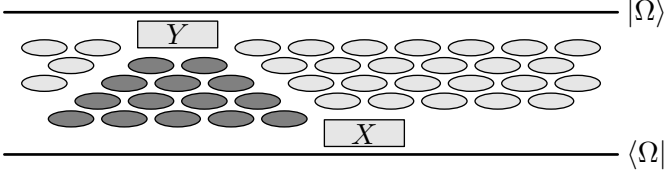
Figure 4. An illustration of the statement $\langle\Omega|XY|\Omega\rangle = \langle\Omega|XP_{in}P_{out}Y|\Omega\rangle$. The operator P_{out} is drawn in light gray and P_{in} is in dark gray. Note that the number of projection layers is proportional to the distance between X and Y.

4. EXPONENTIAL DECAY OF CORRELATIONS

Consider now a k-local Hamiltonian $H = \sum_i Q_i$ defined on a D-dimensional grid. Once again, we assume that Q_i are projections, and that H is frustration-free with a unique ground state $|\Omega\rangle$ (i.e., $Q_i|\Omega\rangle = 0$), and a spectral gap $\epsilon > 0$. Let X, Y be two local observables whose distance on the grid between each other is m. Denote $\bar{X} \stackrel{\text{def}}{=} \langle\Omega|X|\Omega\rangle$, $\bar{Y} \stackrel{\text{def}}{=} \langle\Omega|Y|\Omega\rangle$.

We wish to show:

Lemma 4.1 (Exponential decay of correlations): The correlation of the X, Y observables in the ground state obeys

$$|\langle\Omega|(X - \bar{X})(Y - \bar{Y})|\Omega\rangle| = |\langle\Omega|XY|\Omega\rangle - \bar{X}\bar{Y}| \quad (8)$$
$$\leq \|X\| \cdot \|Y\| \cdot e^{-\mathcal{O}(m)} .$$

Proof: Let us now consider two operators: P_{in}, P_{out}: P_{in} is defined by applying the DL ℓ times to Y and discarding all projections outside the causality cone of Y. ℓ is chosen such that the resulting cone will not overlap with X (see Fig. 4). Therefore $\ell \propto m$, with the proportionality constant that is a geometrical factor. P_{out} is the complement of P_{in}, i.e., it is the product of layers that one get by applying the DL ℓ times, but with a "hole" where the causality cone of Y is. The fact the non-overlapping projections commute implies $P_{in} \cdot P_{out} = A^\ell$ – See Fig. 4 for an illustration in 1D.

Both P_{in}, P_{out} leave the ground-state invariant. In addition, they commute with X and Y respectively, hence

$$\langle\Omega|X = \langle\Omega|XP_{in} , \qquad Y|\Omega\rangle = P_{out}Y|\Omega\rangle ,$$

and therefore

$$\langle\Omega|XY|\Omega\rangle = \langle\Omega|XP_{in}P_{out}Y|\Omega\rangle = \langle\Omega|XA^\ell Y|\Omega\rangle .$$

But by the DL 3.1, $\|A^\ell - \Pi_{gs}\| \leq e^{-\mathcal{O}(\ell)} = e^{-\mathcal{O}(m)}$ where Π_{gs} is the projection to the ground state (see Eq. (2)). Therefore

$$\left|\langle\Omega|XY|\Omega\rangle - \langle\Omega|X\Pi_{gs}Y|\Omega\rangle\right| \leq \|X\| \cdot \|Y\| \cdot e^{-\mathcal{O}(m)} .$$

Assuming that the ground state is unique, $\Pi_{gs} = |\Omega\rangle\langle\Omega|$, and therefore $\langle\Omega|X\Pi_{gs}Y|\Omega\rangle = \bar{X}\bar{Y}$, hence

$$\left|\langle\Omega|XY|\Omega\rangle - \bar{X}\bar{Y}\right| \leq \|X\| \cdot \|Y\| \cdot e^{-\mathcal{O}(m)} .$$

∎

5. IMPROVED 1D AREA LAW

Consider a $1D$ frustration-free, nearest-neighbor Hamiltonian over n particles of dimension d, given by $H = \sum_{i=1}^{n-1} Q_i$. Assume that H has a unique ground state $|\Omega\rangle$ and a spectral gap $\epsilon > 0$. Then $|\Omega\rangle$ satisfies the following area law:

Theorem 5.1: Along any cut in the chain, the von-Neumann entanglement entropy of $|\Omega\rangle$ bounded by

$$S \leq X^{\mathcal{O}(\log X)} \quad (9)$$

for $X \stackrel{\text{def}}{=} \frac{1}{\epsilon}\log d$.

Proof: As explained in the introduction, we consider the Schmidt decomposition of the ground state $|\Omega\rangle = \sum_i \lambda_i|L_i\rangle \otimes |R_i\rangle$ for a given cut. We want to approximate the ground state by starting from the product state $|\phi\rangle \stackrel{\text{def}}{=} |L_1\rangle \otimes |R_1\rangle = \lambda_1|\Omega\rangle + |\Omega^\perp\rangle$, which by Fact 2.1 is the product state with the largest overlap with the ground state. Applying the DL ℓ times, we obtain a state $|\phi_\ell\rangle \stackrel{\text{def}}{=} A^\ell|L_1\rangle \otimes |R_1\rangle$. In terms of the ground state, and its perpendicular space, $|\phi_\ell\rangle$ can be written as

$$|\phi_\ell\rangle = \lambda_1|\Omega\rangle + |\Omega_\ell^\perp\rangle , \quad (10)$$

where $|\Omega_\ell^\perp\rangle \in \mathcal{H}'$. By the DL, $\|\Omega_\ell^\perp\| \leq \Delta_0^\ell$, where $\Delta_0 \stackrel{\text{def}}{=} (1 + \epsilon/2)^{-1/3}$ is the shrinkage factor provided by one application of the DL (see Lemma 3.1). Moreover, by the same argument as for the commuting case, it is easy to see that every application of A increased the SR along the cut by at most $D_0 = d^2$, and so $\text{SR}(\phi_\ell) \leq D_0^\ell$.

Now, there is a tradeoff between how fast the component orthogonal to the ground state shrinks and the increase in the SR. This tradeoff is captured by the product $\Delta_0 D_0$. Consider for a moment the case of $\Delta_0 D_0 < 1$. Applying the DL for $\ell = \frac{-1}{\log(\Delta_0 D_0)}$ would give us a state $|\phi_\ell\rangle$ such that $\text{SR}(\phi_\ell) \cdot \|\phi_\ell\| = (D_0\Delta_0)^\ell < 1/2$. The following is a simple observation is that if such a state exists, then λ_1 is large, and consequently, the ground state has a large projection on the product state $|L_1\rangle \otimes |R_1\rangle$:

Lemma 5.2: If λ_1 is the largest Schmidt coefficient of $|\Omega\rangle$ and there is a state $|\psi\rangle$ such that $|\psi\rangle = \lambda_1|\Omega\rangle + |\Omega^\perp\rangle$ and $\text{SR}(\psi) \cdot \|\Omega^\perp\| \leq 1/2$, then

$$\lambda_1^2 \geq \frac{1}{2\,\text{SR}(\psi)} . \quad (11)$$

The proof of this lemma is simple and is given in the Appendix.

It is well known that if the ground state has a constant projection on a product state, the area law follows; indeed,

a similar argument is used in Hastings' 1D area law for bounding the entropy once a constant overlap with a product state was found. Here we can replace this argument once again with the DL. Using the same sequence of states $|\phi_\ell\rangle$, together with the fact that λ_1 is constant, enables us to achieve good approximations of $|\Omega\rangle$ after only a constant number of steps ℓ. This provides a sequence of states which on one hand we know have a small SR, and on the other approximate the ground states well; by the Eckart-Young theorem (Fact 2.1) this induces a series of constraints on the Schmidt coefficients $\{\lambda_i\}$ of $|\Omega\rangle$ that exponentially bound the amount of mass in the high coefficients. This bound can then be exploited to bound the entanglement entropy $S = -\sum_i \lambda_i^2 \log \lambda_i^2$. We arrive at the following lemma, whose exact derivation can be found in Appendix A.

Lemma 5.3 (Large overlap with implies area law): If there exists a product state $|\phi_1\rangle \otimes |\phi_2\rangle$ such that $|\langle\phi_1 \otimes \phi_2|\Omega\rangle| = \mu$, the entanglement entropy of $|\Omega\rangle$ is upper bounded by

$$S(\Omega) \leq \mathcal{O}(X) \cdot \left[\log \frac{1}{\mu^2(1-\Delta_0^2)} + 2 \right] , \qquad (12)$$

with $X \stackrel{\text{def}}{=} \frac{1}{\epsilon} \log d$.

Now, the problem is that of course, there is no reason to assume that $D_0\Delta_0 < 1$; for most values of ϵ and d this will not be the case.

The solution, which is the main technical innovation of this proof, is to improve the operators A^ℓ of the DL by a process of *diluting*. Instead of applying A^ℓ, we apply a different operator, which "applies" fewer projection terms than A^ℓ, and therefore generates $|\phi_\ell\rangle$ with a much lower SR, while maintaining a small $\|\Omega_\ell^\perp\|$. This is achieved by the following lemma:

Lemma 5.4 (The diluting lemma): For every $\epsilon > 0$ and d, there exists a series of operators $\hat{A}_\ell$ such that $|\phi_\ell\rangle \stackrel{\text{def}}{=} \hat{A}_\ell|L_1\rangle \otimes |R_1\rangle = \lambda_1|\Omega\rangle + |\Omega_\ell^\perp\rangle$ has the following properties: $\text{SR}(\phi_\ell) \leq D_I D^\ell$ and $\|\Omega_\ell^\perp\| \leq \Delta^\ell$, with $\Delta \cdot D < 1/2$ and $\log D = \mathcal{O}(X)$, $\log D_I = X^{\mathcal{O}(\log X)}$.

Using this lemma together with Lemma 5.2 gives us the lower bound $\lambda_1^2 \geq \frac{1}{(DD_I)D^{\log D_I}}$. Inserting this bound into Lemma 5.3 then proves the theorem.

We now proceed to sketch the proof of Lemma 5.4.

5.1. The proof of the diluting lemma (a sketch)

Assume without loss of generality that the projection in A that straddles the given cut belongs to the even layer in A. We will focus our attention on this layer, since the other layer does not increase SR.

Consider a segment I_1 (of $2m$ qudits) around the cut, which contains m projections $P^{(1)}, \ldots, P^{(m)}$ from the even layer (we assume that the cut intersects with $P^{(m/2)}$). We denote the projections by $P^{(i)}$ rather than P_i, for reasons to become clear later. The projection $P_1 \stackrel{\text{def}}{=} P^{(1)} \cdots P^{(m)}$

projects into the common ground state of the region I_1 in the even layer. We will now define the operator $\hat{P}_1$, which closely approximates P_1, but which creates significantly less entanglement. We call $\hat{P}_1$ the "dilution" of P_1.

Sectors: We first observe that the projections $P^{(i)}$ commute among themselves. This means that the "number of violations" of those projections has a meaning; we can decompose the Hilbert space in terms of the mutual eigenspaces of those operators (there are 2^m such subspaces), and those eigenspaces will be denoted by m bits strings $s = (s_1, s_2, \ldots, s_m)$ corresponding to whether the state is inside the null space of the projection or orthogonal to it. A subspace with $s_i = 1$ would be the subspace that is annihilated by $P^{(i)}$. We refer to these subspaces as "violations", since they are orthogonal to the ground state. For a given sector s, we will denote its total number of violations by $|s|$. Note that P_1 annihilates all but the zero violations sector.

Defining the diluted operator $\hat{P}_1$: $\hat{P}_1$ is given by the following formula:

$$\hat{P}_1 \stackrel{\text{def}}{=} \mathbb{B}_1 \cdots \mathbb{B}_j \cdot P^{(i_1)} \cdots P^{(i_{rm})} , \qquad (13)$$

where we use the following definitions. First, the $P^{(i_j)}$ projections are randomly chosen rm projections out of the possible m projections in I_1. Then the $\mathbb{B}_k$ operators are defined as follows: let $\mathbb{N} \stackrel{\text{def}}{=} \sum_{i=1}^m (\mathbb{1} - P^{(i)})$. $\mathbb{N}$ counts the number of violations in each sector: $\mathbb{N}|s\rangle = |s| \cdot |s\rangle$. Then we set $\mathbb{B}_k \stackrel{\text{def}}{=} \mathbb{1} - \frac{1}{k}\mathbb{N}$. It is easy to see that $\mathbb{B}_k|s\rangle = (1 - \frac{|s|}{k})|s\rangle$, and so $\mathbb{B}_k$ "kills" the k sector, while shrinking sectors with $|s| < k$ and amplifying those with $|s| > k$.

We now want to claim two things about this operator: a) that the operator approximates the projection P_1 very well, and b) that it creates (on average, in a sense to be clarified soon) significantly less entanglement along the cut. To this aim, we define two important factors: 1) The shrinking factor Δ_1, and 2) The SR factor D_1. These factors are defined by replacing $P_1 \mapsto \hat{P}_1$ in Π_{even}, and applying this modified DL operator for ℓ times on $|L_1\rangle \otimes |R_1\rangle$. Then Δ_1 is the factor that bounds the norm of the part perpendicular to the ground state, and D_1 is asymptotically the average SR that is created by each line.

Let us first argue regarding the quality of the approximation. We claim that

Claim 5.5: The shrinking factor of $\hat{P}_1$ is $\Delta_1 = \Delta_0 + 2^{-j}$.

It is easy to see that if $|s\rangle$ is in the s sector with $k = |s|$ violations, then the probability that it would be annihilated by our random product is

$$\Pr = \frac{\binom{m-k}{rm}}{\binom{m}{rm}} \leq (1-r)^k .$$

Using the probabilistic method, it is then straightforward to deduce that for a general state $|\psi\rangle = \sum_s c_s |s\rangle$, one can always find rm projections that would yield $\|P^{(i_1)} \cdots P^{(i_{rm})}|\psi\rangle\|^2 \leq \sum_s |c_s|^2 (1-r)^{|s|}$. We see that the random projections can erode most of the mass of the high violations sectors. The sectors with low violations are taken care of by the $\mathbb{B}_k$ operators: they shrink the mass in the $|s| < k$ sectors, annihilate the the $|s| = k$ sectors and inflate the mass in the $|s| > k$ sectors. The idea is that we use these operators to kill the first j sectors, and the rm random projections to kill much of the mass of the high sectors.

It is easy to verify that $\hat{P}_1$ leaves the zero violation sector invariant and annihilates all the sectors with $k \leq j$ violations. For sectors with $k > j$ violations, it shrinks the mass by a factor at most $\left[\binom{k-1}{j-1}\right]^2 (1-r)^k \leq 4^k (1-r)^k$, and therefore by taking $r = 9/10$, we obtain a shrinking by a factor smaller than 2^{-k}. Since $k \geq j$, the overall error is at most 2^{-j}. While the shrinkage factor per application of the DL, associated with the operator P_1 is Δ_0, if we apply the operator $\hat{P}_1$ in Π_{even} it would yield an operator whose effective shrinkage factor per application of an even layer is given by the factor $\Delta_1 = \Delta_0 + 2^{-j}$. Taking j large enough we can maintain a very close shrinkage factor to the original one.

We would now like to claim that the operator $\hat{P}_1$ is much more efficient in terms of the SR it generates than P_1. We claim:

Claim 5.6: The SR factor of $\hat{P}_1$ is $D_1 = (20 j^{3/2} J^{1/2}) D_0^{j \log m / m} D_0^{9/10}$, where $J \overset{\text{def}}{=} j^{\log j}$.

To argue this, we will argue separately for the contribution of the $P^{(i)}$ to the SR, and of the $\mathbb{B}_k$.

Roughly speaking, applying $\hat{P}_1$ for ℓ times, we expect the columns in the segment I_1 to contain $r\ell$ projections instead of ℓ projections. Therefore, the expected SR (per application of the even layer) should grow like D_0^r (with $r = 9/10$) instead of D_0. This intuition is correct, but one has to argue why the $\mathbb{B}_k$ operators do not introduce too much entanglement.

To understand how to bound the SR that is generated by the $\mathbb{B}_1 \cdots \mathbb{B}_j$ operators, let us start by considering just the SR generated by one $\mathbb{B}_k$. Consider the operator $\mathbb{P} \overset{\text{def}}{=} \sum_{i=1}^m P^{(i)}$, which appears in every $\mathbb{B}_k$ since we can write $\mathbb{B}_k = \left(1 - \frac{m}{k}\right) \mathbb{1} + \frac{1}{k} \mathbb{P}$. Let us understand how to upper bound the SR of the state $(\mathbb{P}\Pi_{odd})^\ell |L_1\rangle \otimes |R_1\rangle$ along the cut in the middle. This is a simpler problem, which, nevertheless, contains all the necessary ingredients needed for the general case; the analysis of applying several $\mathbb{B}_k$'s is more involved but relies on similar ideas, and it is not too difficult to see that the operators $P^{(i)}$, which we currently ignore, do not affect this argument.

Naively, an application of $\mathbb{P}$ would increase the SR by the factor $D_0 + 1$ (since it is the sum of the entangling $P^{(i)}$ and the other, non-entangling projections). We need to provide a better upper bound. The idea is to find one column, not too far from the cut, in which the number of entangling projectors is significantly smaller. If such a column exists within distance m from the middle cut, then the SR generated at the cut can be at most the SR generated in that column times d^m, because d^m is the dimension of the Hilbert space of all the qudits in the segment.

To argue the existence of such a sparse column, write $\mathbb{P} = \mathbb{P}_L + \mathbb{P}_R$ where $\mathbb{P}_L$ is the sum of all the projections left to the cut, and $\mathbb{P}_R$ is the sum of the rest. Then $(\mathbb{P}\Pi_{odd})^\ell = [(\mathbb{P}_L + \mathbb{P}_R)\Pi_{odd}]^\ell$ breaks into 2^ℓ terms. We will upper bound the SR of each term individually, and so the overall SR would be bounded by the sum of the different SRs. Considering one such term, either a $\mathbb{P}_L$ or a $\mathbb{P}_R$ at every even layer. Therefore there must be a side in which there are at most $\ell/2$ operators, and assume WLOG that this is the left side. Then we choose to estimate the SR in the middle of that side. In the worst case, the SR contribution in the middle of the left side is due to $\ell/2$ $\mathbb{P}_L$. To bound it, we bisect the left segment and write $\mathbb{P}_L = \mathbb{P}_{LL} + \mathbb{P}_{LR}$ — similarly to what was done before. Now the product of $\ell/2$ $\mathbb{P}_L$ terms breaks into the sum of $2^{\ell/2}$ terms, and again we bound the SR of each term individually, by looking at the side that has less than $\ell/4$ entangling operators.

Proceeding this way for $\log m$ times, we end up with at most $2^{\ell + \ell/2 + \ell/4 + \cdots} \leq 4^\ell$ terms, each of which has a column i that contains at most ℓ/m entangling operators $P^{(i)}$. The SR of each such term at that column is $D_0^{\ell/m}$, and so its SR at the middle cut is at most $D_0^{\ell/m} d^m$. Summing up the SR of all terms, we find $SR \leq (4 D_0^{1/m})^\ell d^m$. Except from an overall factor of d^m (which we will handle separately), we see that the effective SR factor of the $\mathbb{P}$ operators is $4 D_0^{1/m}$, which can be significantly smaller than D_0 for large D_0.

In the general case, where we have $\mathbb{B}_1 \cdots \mathbb{B}_j$, a similar derivation implies that we end up with $(20 j^{3/2} J^{1/2})^\ell$ terms, where $J \overset{\text{def}}{=} j^{\log j}$, and each term would have a column i with only $\frac{j \log m}{m} \ell$ entangling operator. Together with the SR contribution of the random operators, we can upper bound the SR of ℓ such layers by $\left[(20 j^{3/2} J^{1/2}) D_0^{j \log m / m} D_0^{9/10}\right]^\ell d^m$.

Concatenation analysis: We conclude that diluting the I_1 segment, namely replacing the projection P_1 on its ground state by $\hat{P}_1$, resulted in new effective factors

$$\Delta_0 \mapsto \Delta_1 = \Delta_0 + 2^{-j} \ , \tag{14}$$

$$D_0 \mapsto (20 j^{3/2} J^{1/2}) D_0^{j \log m / m} D_0^{9/10} \ . \tag{15}$$

The final step of the proof is to use the diluting procedure iteratively, much like the concatenation of error correction codes. In the next level, we consider the segment I_2 that contains m^2 projections, arranged as m segments of type I_1. Using the same recipe, we dilute its corresponding projection $P_2 = P_1^{(1)} \cdots P_1^{(m)}$, where here each of the projections $P_1^{(i)}$

is itself a product of m projections. We then replace the $P_1^{(i)}$ projections by their diluted version $\hat{P}_1^{(i)}$. Continuing this way, it is possible to prove, that as long as $m < 40 \cdot 2^j$, the error does not accumulate, and consequently, after n iterations:

$$\Delta_n = \Delta_0 + 2^{-j} , \tag{16}$$

$$D_n = (20 j^{3/2} J^{1/2}) D_0^{j \log m / m} D_{n-1}^{9/10} \tag{17}$$

$$= \left[(20 j^{3/2} J^{1/2}) D_0^{j \log m / m} \right]^{10} D_0^{(9/10)^n} . \tag{18}$$

The segment over which we performed the n iterations consists of m^n projections, which are defined on $2m^n$ qudit. The distance between the middle cut and the any column in the segment is therefore at most m^n qubits, and so the estimate the SR after ℓ layers of the diluted operator is $D^\ell D_I$, where $D_I = d^{m^n}$.

It is now easy to pick n, m, j as functions of ϵ, d that would give us the promised values for Δ, D, D_I.
∎

ACKNOWLEDGMENT

APPENDIX

Proof of Lemma 5.2:

We write the (un-normalized) Schmidt decomposition of $|\psi\rangle$ as

$$|\psi\rangle = \sum_{i=1}^R \mu_i |A_i\rangle \otimes |B_i\rangle = \lambda_1 |\Omega\rangle + |\Omega^\perp\rangle . \tag{19}$$

(where $\mu_1 \geq \mu_2 \geq \ldots$, $R \overset{\text{def}}{=} \mathrm{SR}(\psi)$, and $\|\psi\|^2 = \sum_i \mu_i^2$). Using Fact 2.1, the overlap of every product state $|X\rangle \otimes |Y\rangle$ with $|\Omega\rangle$ is at most λ_1, its largest Schmidt coefficient. Therefore for every i, $|\langle\Omega|A_i\rangle|B_i\rangle| \leq \lambda_1$, and so multiplying Eq. (19) by $\langle\Omega|$ and taking the absolute value, we get

$$\lambda_1 = |\langle\Omega|\psi\rangle| = |\sum_{i=1}^R \mu_i \langle\Omega|A_i\rangle|B_i\rangle| \leq \sum_{i=1}^R \mu_i \lambda_1 .$$

Therefore μ_1, the largest Schmidt coefficient of $|\psi\rangle$, must satisfy $\mu_1 \geq \frac{1}{R}$. On the other hand, multiplying Eq. (19) by $\langle A_1|\langle B_1|$, and using once again the fact that $|\langle A_1|\langle A_1|\Omega\rangle| \leq \lambda_1$ gives us $\mu_1 \leq \lambda_1^2 + \|\Omega^\perp\|$. All together, we get

$$\frac{1}{R} \leq \mu_1 \leq \lambda_1^2 + \|\Omega^\perp\| , \tag{20}$$

from which we get

$$\lambda_1^2 \geq \frac{1}{R} - \|\Omega^\perp\| = \frac{1}{R}(1 - R \cdot \|\Omega^\perp\|) \geq \frac{1}{2R} .$$
∎

Proof of Lemma 5.3:

Let

$$|\Omega\rangle = \sum_{i \geq 1} \lambda_i |L_1\rangle |R_i\rangle \tag{21}$$

be the Schmidt decomposition of $|\Omega\rangle$ with respect to its bi-partitioning, with, use the DL on asserted product state $|\phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle$ to generate a sequence of converging states $|\phi_\ell\rangle \overset{\text{def}}{=} A^\ell |\phi\rangle$. Then $\mathrm{SR}(\phi_\ell) \leq D_0^\ell$ and $|\phi_\ell\rangle = \mu|\Omega\rangle + |\Omega_\ell^\perp\rangle$ with $\|\Omega_\ell^\perp\| \leq \Delta_0^\ell$.

Defining $|v_\ell\rangle$ to be the normalization of $|\phi_\ell\rangle$, we get

$$|\langle v_\ell|\Omega\rangle| \geq \frac{\mu}{\sqrt{\mu^2 + \Delta_0^{2\ell}}} .$$

Since $\mathrm{SR}(v_\ell) \leq D_0^\ell$, it follows from Fact 2.1 (Eckart-Young theorem) that

$$\sum_{i > D_0^\ell} \lambda_i^2 \leq 1 - \frac{\mu^2}{\mu^2 + \Delta_0^{2\ell}} \leq \frac{1}{\mu^2} \Delta_0^{2\ell} .$$

We found that the distribution of Schmidt coefficients of $|\Omega\rangle$ must satisfy a series of inequalities, which essentially limit the amount of mass found in the high Schmidt coefficients. It is easy to see that in such case the entanglement entropy is bounded by the entanglement of the steps-like distribution of weights that is implied by the above inequalities. This bound is formally given by the following lemma, which is proved in the end of this section.

Lemma A.1: If a probability distribution obeys

$$\sum_{i > AD^\ell} p_i \leq K\theta^\ell , \tag{22}$$

with $A, K, D > 1$ and $0 < \theta < 1$, then

$$S \leq 2 \log A + 3 \left(\frac{\log \frac{K}{1-\theta} + 1}{\log(1/\theta)} + 2 \right) \log D . \tag{23}$$

Substituting $A = 1, D = D_0, K = \mu^{-2}$, and $\theta = \Delta_0^2$, then gives us

$$S \leq 3 \left(\frac{\log \frac{1}{\mu^2(1-\Delta_0^2)} + 1}{2 \log(1/\Delta_0)} + 2 \right) \log D_0 \tag{24}$$

$$= \frac{3}{2} \frac{\log D_0}{\log(1/\Delta_0)} \left[\log \frac{1}{\mu^2(1-\Delta_0^2)} + 1 + 4 \log(1/\Delta_0) \right] \tag{25}$$

$$\leq \frac{3}{2} X \left[\log \frac{1}{\mu^2(1-\Delta_0^2)} + 2 \right] . \tag{26}$$

Where in the last inequality, we used the fact that $X = \frac{\log D_0}{\log(1/\Delta_0)}$ (see text above Lemma 5.4) and the assumption that $\epsilon \leq 1$, which implies $4 \log(1/\Delta_0) \leq 1$.

We end the proof of the lemma by proving Lemma A.1

Proof:

Call the set of weights $\{p_j\}$ for $AD^\ell + 1 \leq j \leq AD^{\ell+1}$ the ℓ'th block. Then the constraints in Eq. (22) imply that for every block $\ell \geq 1$,

$$\sum_{j=AD^\ell+1}^{AD^{\ell+1}} p_j \leq K\theta^\ell \ . \tag{27}$$

Obviously, by reshuffling the mass within a block we maintain the constraints in Eq. (22). Moreover, it is straight forward to see that the entropy contribution of every block is maximized when all the weights in it are equal. The maximal distribution is therefore a steps function, which satisfies:

$$\text{in block } \ell, \quad p_j \leq \frac{K\theta^\ell}{AD^{\ell+1} - AD^\ell} = \frac{K}{A(D-1)}(\theta/D)^\ell \ . \tag{28}$$

We now define ℓ_0 to be the first block for which $K\theta^\ell \leq \frac{1}{2}(1-\theta)\theta$:

$$\frac{\log \frac{2K}{\theta(1-\theta)}}{\log(1/\theta)} \leq \ell_0 \leq \frac{\log \frac{2K}{\theta(1-\theta)}}{\log(1/\theta)} + 1 = \frac{\log \frac{K}{1-\theta} + 1}{\log(1/\theta)} + 2 \ . \tag{29}$$

We will bound the maximal entropy by bounding the entropy contribution of blocks up to (and including) $\ell_0 - 1$ and blocks from ℓ_0 onwards. The first is easy, as there are AD^{ℓ_0} weights in the low blocks:

$$S_I \leq \log(AD^{\ell_0}) = \ell_0 \log D + \log A \ . \tag{30}$$

In the high blocks, $p_j \leq \frac{1}{2}(1 - \theta) \leq 1/2$, so we can use the monotonicity of the function $-p\log p$ in the range $(0 : 1/2]$ to bound the entropy by

$$
\begin{aligned}
S_{II} &\leq -\sum_{\ell \geq \ell_0} K\theta^\ell \log\left[\frac{K}{A(D-1)}(\theta/D)^\ell\right] \\
&\leq \sum_{\ell \geq \ell_0} K\theta^\ell \log(AD^{\ell+1}) \\
&= \frac{K\theta^{\ell_0}}{1-\theta}\left[\log A + \left(\ell_0 + \frac{1}{1-\theta}\right)\log D\right] \\
&\leq \log A + \left(\ell_0 + \frac{1}{1-\theta}\right)\log D \ ,
\end{aligned}
$$

where the second equality follows from standard geometric sums identities, and the last inequality follows from the definition of ℓ_0. Next, looking at the lower bound of ℓ_0 in Eq. (29), it takes standard calculus to verify that $\ell_0 \geq \frac{1+\log\frac{1}{1-\theta}}{\log(1/\theta)} + 1 \geq \frac{1}{1-\theta}$, and so $S_{II} \leq \log A + 2\ell_0 \log D$, and

$$S = S_I + S_{II} \leq 2\log A + 3\ell_0 \log D \ .$$

Plugging the upper bound of ℓ_0 from Eq. (29), we get Eq. (23). ∎
∎

REFERENCES

[1] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani, "The detectability lemma and quantum gap amplification," in *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, `arXiv:0811.3412`. New York, NY, USA: ACM, 2009, pp. 417–426.

[2] P. Calabrese and J. Cardy, "Entanglement entropy and quantum field theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2004, no. 06, p. P06002, 2004. [Online]. Available: http://stacks.iop.org/1742-5468/2004/i=06/a=P06002

[3] H. Casini and M. Huerta, "Entanglement and alpha entropies for a massive scalar field in two dimensions," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 12, p. P12012, 2005. [Online]. Available: http://stacks.iop.org/1742-5468/2005/i=12/a=P12012

[4] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.

[5] D. Gottesman and M. B. Hastings, "Entanglement versus gap for one-dimensional spin systems," *New Journal of Physics*, vol. 12, no. 2, p. 025002, 2010. [Online]. Available: http://stacks.iop.org/1367-2630/12/i=2/a=025002

[6] M. B. Hastings, "Lieb-schultz-mattis in higher dimensions," *Phys. Rev. B*, vol. 69, no. 10, p. 104431, Mar 2004.

[7] M. B. Hastings, "Entropy and entanglement in quantum ground states," *Phys. Rev. B*, vol. 76, no. 3, p. 035114, 2007.

[8] M. Hastings, "An Area Law for One Dimensional Quantum Systems," *JSTAT, P*, vol. 8024, 2007.

[9] S. Irani, "Ground state entanglement in one-dimensional translationally invariant quantum systems," *Journal of Mathematical Physics*, vol. 51, no. 2, p. 022101, 2010. [Online]. Available: http://link.aip.org/link/?JMP/51/022101/1

[10] ——, Sandy Irani, private communication, 2011.

[11] G. Vidal, "Efficient simulation of one-dimensional quantum many-body systems," *Phys. Rev. Lett.*, vol. 93, no. 4, p. 040502, Jul 2004.