High-Temperature Gibbs States are Unentangled and Efficiently Preparable

Ainesh Bakshi	Allen Liu	Ankur Moitra	Ewin Tang
ainesh@mit.edu	cliu568@mit.edu	moitra@mit.edu	ewin@berkeley.edu
MIT	MIT	MIT	UC Berkeley

Abstract

We show that thermal states of local Hamiltonians are separable above a constant temperature. Specifically, for a local Hamiltonian H on a graph with degree $\mathfrak{d}$, its Gibbs state at inverse temperature β, denoted by $\rho = e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$, is a classical distribution over product states for all $\beta < 1/(c\mathfrak{d})$, where c is a constant. This *sudden death of thermal entanglement* upends conventional wisdom about the presence of short-range quantum correlations in Gibbs states.

Moreover, we show that we can efficiently sample from the distribution over product states. In particular, for any $\beta < 1/(c\mathfrak{d}^3)$, we can prepare a state ε-close to ρ in trace distance with a depth-one quantum circuit and $\operatorname{poly}(n)\log(1/\varepsilon)$ classical overhead.[1] A priori the task of preparing a Gibbs state is a natural candidate for achieving super-polynomial quantum speedups, but our results rule out this possibility above a fixed constant temperature.

[1] In independent and concurrent work, Rouzé, França, and Alhambra [RFA24] obtain an efficient quantum algorithm for preparing high-temperature Gibbs states via a dissipative evolution.

Contents

1 Introduction

A central motivation behind the study of quantum many-body systems is to understand the behavior of entanglement, i.e. non-classical correlations. Spurred by experimental breakthroughs in quantum simulation and quantum phases of matter, a rich body of work has sprung around understanding entanglement in Gibbs states, which model quantum systems at thermal equilibrium [Alh23]. The goal of this literature is to quantify the allowable entanglement of these states as dictated by the locality structure of their underlying Hamiltonians.

Prior rigorous studies on entanglement in Gibbs states, including thermal area laws [WVHC08; KAA21], bounds on conditional mutual information [KB19], local indistinguishability [KGKRE14], efficient state preparation [BK18; BCGLPR23], and efficient learning algorithms [AAKS21; BLMT23], proceeds by bounding the entanglement through proxy correlation measures that combine both classical and quantum correlations. Consequently these results only provide meaningful bounds on entanglement at *long range*, when classical correlations are sufficiently small. This remains true even for results which assume the Gibbs state is above a critical temperature [BK18; HMS20; HKT22]. Taken together, this body of work sows the conventional wisdom that *short-range* quantum correlations, like short-range classical correlations, exist at any constant temperature.

We upend this conventional wisdom by showing that above some constant temperature, the Gibbs state of any local Hamiltonian exhibits *zero* entanglement, even at short range.

Theorem 1.1 (Informal version of Theorem 1.5)**.** *Let $H = \sum_{a=1}^{m} \lambda_a G_a$ be a Hamiltonian where each term G_a acts on a constant number of qubits and each qubit is acted on by at most $\mathfrak{d}$ terms, and $|\lambda_a| \leqslant 1$ for all $a \in [m]$. Then there is a constant γ such that, for any non-negative $\beta < 1/(\gamma\mathfrak{d})$, the Gibbs state at inverse temperature β, $\rho = e^{-\beta H} / \operatorname{tr}(e^{-\beta H})$, can be written as a distribution over tensor products of stabilizer states[2], i.e.*

$$\rho = \sum_{|\psi\rangle \in \mathcal{S}} p_\psi \, |\psi\rangle \langle\psi| \,,$$

where $\mathcal{S} = \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |\mathrm{i}\rangle, |-\mathrm{i}\rangle\}^{\otimes n}$ and the p_ψ's are non-negative and sum to 1.

With this result, we witness a "sudden death of thermal entanglement": there is a constant critical temperature, above which correlations are purely classical. Alternatively formulated, this suggests that an entangled state, if weakly coupled to a bath at high temperature and allowed to thermalize, will become fully disentangled at a *finite* time. To the best of our knowledge, this result is the first to show strong bounds on short-range entanglement at constant temperature.

Our structural result for high-temperature Gibbs states has implications for the computational task of preparing quantum Gibbs states. This task, known as quantum Gibbs sampling, has been studied extensively [CKBG23, Table 1], dating back to the work of Temme, Osborne, Vollbrecht, Poulin, and Verstraete [TOVPV11]. However, remarkably little is known about when Gibbs states can be prepared efficiently. Despite a wealth of approaches and proposals, efficient algorithms have only been rigorously established in fairly restricted settings, such as for Hamiltonians with constant operator norm [GSLW19], commuting Hamiltonians [KB16], 1D Hamiltonians [BK18; BCGLPR23], or under strong assumptions, like the eigenstate thermalization hypothesis [CB21]. On the other hand, sampling from the Gibbs distribution at low

[2]For our purposes, it makes sense to think of these as the set of eigenvectors of tensored Pauli matrices. Throughout, we write the density matrices of single-qubit stabilizer states in terms of Pauli matrices, $\frac{1}{2}(I \pm \sigma_z)$, $\frac{1}{2}(I \pm \sigma_x)$, and $\frac{1}{2}(I \pm \sigma_y)$.

temperature is known to be NP-hard, even in the classical setting [Sly10; SS14; GŠV16]. Thus there is a natural target: Are all high-temperature Gibbs states efficiently preparable?

We resolve this question as well. Our second main result is:

Theorem 1.2 (Informal version of Theorem 1.6)**.** *Let $H = \sum_{a=1}^{m} \lambda_a G_a$ be a Hamiltonian where each term G_a acts on a constant number of qubits and each qubit is acted on by at most $\mathfrak{d}$ terms, and $|\lambda_a| \leqslant 1$ for all $a \in [m]$. Let $\beta < 1/(\gamma \mathfrak{d}^3)$, for a fixed constant γ and let $\rho = e^{-\beta H} / \operatorname{tr}(e^{-\beta H})$ be the corresponding Gibbs state. Given $0 < \varepsilon < 1$, there exists an algorithm that outputs a state $\hat{\rho}$ such that $\|\rho - \mathbb{E}[\hat{\rho}]\|_1 \leqslant \varepsilon$, and requires a depth-1 quantum circuit with $\mathrm{poly}(n) \cdot \log(1/\varepsilon)$ classical overhead.*

The task of preparing a Gibbs state is a natural place to look for quantum speedups. However, our result shows that this task offers no super-polynomial quantum speedup for temperature larger than a fixed constant. On the other hand, assuming that NP-hard problems cannot be solved in BQP, we know that for preparing a Gibbs state for temperature smaller than a different fixed constant will not work either. Going forward, finer-grained control of the separability and computational thresholds across different models appears crucial to understanding Gibbs sampling, both as a testbed for quantum thermodynamics and as a candidate for quantum advantage.

1.1 Results

We now formally state our results. Throughout, we consider systems on n qubits, at inverse temperatures $\beta > 0$. Let $\mathcal{P}$ denote the set of n-fold tensor products of the four 2×2 Pauli matrices, $\mathcal{P} = \{I, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$ (see Definition 3.1). We consider a class of Hamiltonians which are local with respect to an underlying graph, which we call *low-intersection* Hamiltonians.

Definition 1.3 (Hamiltonian)**.** A *Hamiltonian* on n qubits is an operator $H \in \mathbb{C}^{2^n \times 2^n}$ that we consider as a linear combination of local *terms* $G_a \in \mathcal{P}$ with associated *coefficients* λ_a, $H = H(\lambda) = \sum_{a=1}^{m} \lambda_a G_a$. We also refer to these qubits as *sites*.

For normalization, we assume that the G_a's are distinct, non-identity, and that $|\lambda_a| \leqslant 1$. We say this Hamiltonian is $\mathfrak{K}$-*local* if every term G_a is supported on at most $\mathfrak{K}$ qubits: $|\mathrm{supp}(G_a)| \leqslant \mathfrak{K}$.

Assuming that the terms are Pauli can be done without loss of generality: any term supported on $\mathfrak{K}$ qubits can be expanded into the Pauli basis, inflating the number of terms by at most a factor of $4^{\mathfrak{K}}$.

Definition 1.4 ($(\mathfrak{d}, \mathfrak{K})$-low-intersection Hamiltonian)**.** For an n-qubit Hamiltonian $H = \sum_{a=1}^{m} \lambda_a G_a$, we define its *underlying graph* $\mathfrak{G}$ to be the hypergraph on n vertices whose edges are given by the sets $\mathrm{supp}(G_a)$ for $a \in [m]$. We say H has degree $\mathfrak{d}$ if the degree of every vertex in the graph is at most $\mathfrak{d}$.

We call a Hamiltonian H a $(\mathfrak{d}, \mathfrak{K})$-*low-intersection Hamiltonian* if H has locality $\mathfrak{K}$ and degree $\mathfrak{d}$.

Low-intersection Hamiltonians generalize geometrically local Hamiltonians in low-dimensional spaces, which is the type of Hamiltonian often considered in physically motivated settings. For example, a $\mathfrak{K}$-local Hamiltonian which is geometrically local with respect to a d-dimensional lattice is a $((2d)^{\mathfrak{K}-1}, \mathfrak{K})$-low intersection Hamiltonian. We now formally state our structural result about high-temperature Gibbs states.

Theorem 1.5 (High-temperature Gibbs states are separable)**.** *Given a $(\mathfrak{d}, \mathfrak{K})$-low-intersection Hamiltonian (see Definition 1.4) and $\beta < 1/(\gamma \cdot \mathfrak{d} \cdot \mathfrak{K}^2)$ for a fixed universal constant γ, the corresponding*

Gibbs state $\rho = e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$ can be expressed as a positive linear combination of $A_1 \otimes A_2 \otimes \ldots \otimes A_n$, such that for each $j \in [n]$,

$$A_j \in \left\{ I/2, (I \pm \sigma_x)/2, (I \pm \sigma_y)/2, (I \pm \sigma_z)/2 \right\}.$$

Prior work on thermal area laws shows that the mutual information between two subsystems L and R of a Gibbs state is bounded by a linear function in the size of the surface area of L, $\beta|\partial L|$ [WVHC08; KAA21]. Here, mutual information serves as a proxy for entanglement, mirroring area laws for entanglement entropy in ground states [Has07; AKLV13; AAG22]. In sharp contrast, our results treat entanglement directly and demonstrate that, for any temperature higher than a fixed constant, it is identically zero.

Next, we state our result on efficiently preparing Gibbs states.

Theorem 1.6 (High-temperature Gibbs states are efficiently preparable)**.** *Given $0 < \varepsilon, \delta < 1$ and a $(\mathfrak{d}, \mathfrak{K})$-low-intersection Hamiltonian (see Definition 1.4) and $\beta < \beta_c = 1/(\gamma \mathfrak{d} \mathfrak{K})^2$, for a fixed universal constant γ, let $\rho = e^{-\beta H}/\operatorname{tr}\left(e^{-\beta H}\right)$. Then, there exists a classical randomized algorithm that runs in time*

$$\widetilde{\mathcal{O}}\left(n^{7+\frac{\log(\mathfrak{d})}{\log(\beta_c/\beta)}} \cdot \log^2(n/\varepsilon) \cdot \log(1/\delta) \cdot \operatorname{poly}(\mathfrak{K}, \mathfrak{d})\right).$$

It satisfies the following properties:

1. *With probability at least $1 - \delta$, the algorithm outputs a classical description of a product state $\hat{\rho} = A_1 \otimes \cdots \otimes A_n$ where every A_j is either an eigenvector of a Pauli matrix or maximally mixed,*

$$A_j \in \left\{ \frac{I}{2}, \frac{I \pm \sigma_x}{2}, \frac{I \pm \sigma_y}{2}, \frac{I \pm \sigma_z}{2} \right\};$$

   otherwise, the algorithm outputs $\perp$.

2. *Conditioned on successfully outputting a state $\hat{\rho}$, the mixture over $\hat{\rho}$ is close to ρ in trace distance,*

$$\|\rho - \mathbb{E}[\hat{\rho}]\|_1 \leqslant \varepsilon,$$

   where the expectation is only over the randomness of the algorithm.

Theorem 1.2 follows from this theorem by considering $\beta < \beta_c/\mathfrak{d}$, so that the exponent on n, $7 + \frac{\log(\mathfrak{d})}{\log(\beta_c/\beta)} < 8$, is a constant.

With this algorithm, one can prepare a copy of ρ by running our randomized algorithm, taking the classical description of $\hat{\rho}$, and preparing it with a depth-one quantum circuit. Note that preparing ρ in expectation is equivalent to preparing ρ, since one can just "forget the algorithm's steps" to get a copy of ρ without classical side correlations. Even stronger, if one performs the algorithm coherently, it outputs a purification of ρ.

On temperature. Our bound on the critical temperature for the algorithm cannot be significantly improved. Work by Sly and Sun [SS14] shows that approximately sampling from the anti-ferromagnetic Ising model on a d-regular graph is NP-hard at the "uniqueness threshold" for the d-regular tree, which is at $\beta = \Theta(1/d) = \Theta(1/\mathfrak{d})$ [SST14]. This hardness statement for classical Gibbs sampling implies hardness for the more general problem of quantum Gibbs sampling.

Note that the threshold for β in Theorem 1.5 of $1/(\gamma \mathfrak{d} \mathfrak{K}^2)$ is larger than the threshold in Theorem 1.6 of $1/(\gamma \mathfrak{d}^2 \mathfrak{K}^2)$. This is because all of the structural properties that we need hold up

to $1/(\gamma \mathfrak{d} \mathfrak{K}^2)$ but algorithmically, we need $\beta < 1/(\gamma \mathfrak{d}^2 \mathfrak{K}^2)$ because we are approximating the partition function using the subroutine in [HKT22].

1.2 Related work

Locality in high-temperature Gibbs states. Several works in the quantum information literature focus on high-temperature Gibbs states. These bound correlation measures which do not distinguish quantum and classical, including covariance of observables [HMS20] and local indistinguishability [KGKRE14; BK18]. Our work, which shows a lack of quantum correlation even in the presence of classical correlations, is a significant departure from this prior work.

Sudden death of entanglement. Sudden death of entanglement is the phenomenon that two entangled qubits, when subject to environmental noise, does not exhibit exponentially decaying entanglement with time, as classical correlations do, but rather become entirely disentangled after a finite amount of time [YE09]. The body of literature on ESD (entanglement sudden death) focuses on analyzing two-qubit systems under various noise models [YE04; FMB05; AJ08] and experimental demonstrations of ESD [Alm+07]. Its study as a phenomenon of many-body systems is more limited, likely because even defining a measure of entanglement for mixed states is non-trivial [HHHH09], and separability is difficult to detect for large systems.

Existing work studies the sudden death of *entanglement negativity*, an entanglement monotone which can be computed efficiently [VW02], for specific spin systems, either through heuristic arguments or numerical calculations [ABV01; AMD15; SDHS16; HC18]. States with zero entanglement negativity need not be separable [HHH98]. So, our work is vastly more general, and proves separability of Gibbs states at high temperature, a much stronger result than lack of entanglement negativity.

Classical Gibbs sampling. Classical Gibbs sampling is, comparatively, well-understood and researchers have characterized sharp phase transitions wherein there is some critical temperature, above which sampling the Gibbs state is computationally efficient and below which it is computationally hard [Sly10; SS14; GŠV16]. The fact that there are such wide gaps in our understanding of quantum Gibbs sampling is especially surprising given the diverse range of techniques we have for classical spin systems, such as path coupling [BD97], canonical paths [JSV04], correlation decay [Wei06], abstract polymer models [KP86], zero-free regions [Bar16], spectral independence [ALG21], and stochastic localization [CE22]. Our results can be thought of as a sampling-to-counting reduction for quantum systems. Thus, we give an algorithmic alternative to directly working with Lindbladians of dissipative evolutions.

Concurrent work on high-temperature Gibbs sampling. In concurrent and independent work, Rouzé, França, and Alhambra [RFA24] prove that the dissipative evolution studied by Chen, Kastoryano, and Gilyén [CKG23] has a constant spectral gap at high temperature, showing that this evolution is an efficient Gibbs sampling algorithm at high temperature. The techniques are significantly different than ours, controlling the evolution by viewing it as a perturbation of an infinite-temperature dissipation. We do not analyze such an evolution.

Cluster expansion and abstract polymer models. The foundational tool of our approach is cluster expansion, which allows quantities like the log-partition function and marginals of high-temperature Gibbs states to be expressed as exponentially decaying Taylor series. This tool has

been used to show a variety of efficient algorithms around Gibbs states and real-time evolution, including the computation of partition functions [MH21], learning of high-temperature Gibbs states [HKT22], and sampling from the measurement distribution of a high-temperature Gibbs state [YL23].

Among these, the latter sampling result of Yin and Lucas bears the most similarity to our result, using a sampling-to-counting reduction with cluster expansion to give a classical algorithm to sample from the measurement probabilities of a Gibbs state in, say, the computational basis. Our work also implies an efficient algorithm for this task, and achieves a stronger result by tackling the additional challenge of performing these arguments "without measuring".

2 Technical overview

In this section, we describe the key technical ingredients that we need to obtain Theorems 1.5 and 1.6. We refer the reader to Section 3 for notation and background.

2.1 Gibbs states are unentangled

First, we argue that high-temperature Gibbs states are *unentangled*. For intuition, we first present a simple argument which shows that Gibbs states are separable for a temperature dependent on system size. Consider the maximally mixed state, $I/2^n$. This state is the Gibbs state at infinite temperature, and is in the interior of the convex hull of product states. So, as β tends to zero, ρ will eventually enter the interior of this convex hull, making it separable. This happens at a finite β which depends on system size. Our proof will proceed by extracting sites one by one and performing a similar argument as described locally; because we only ever do it for small subsystems, we can show separability at constant temperature.

Let $H = \sum_{a\in[m]} \lambda_a H_a$ be a $(\mathfrak{d}, \mathfrak{K})$-low-intersection Hamiltonian, and let $\beta < 1/(100\mathfrak{d}\cdot\mathfrak{K}^2)$. Then, we can define the Hamiltonian restricted to a single site $j \in [n]$, $H_{(j)}$, to be the terms that contain j in their support (see Definition 3.4). With the goal of isolating the Gibbs state at site j, we consider the Taylor expansion of $e^{-\beta H}\cdot e^{\beta(H-H_{(j)})}$. We show that for our choice of β, this series converges exponentially fast and we can sample a single term from it in an unbiased manner.

Sampling a term. Expanding the Taylor series, we show that we can write

$$e^{-\beta H}\cdot e^{\beta(H-H_{(j)})} = \sum_{t=0}^{\infty} p_t\big(H, H_{(j)}\big), \tag{1}$$

where $p_t(H, H_{(j)})$ is a matrix-valued, degree-t polynomial satisfying the following recurrence relation:

$$p_{t+1}(H, H_{(j)}) = \frac{\beta}{t+1}\Big(-\Big[H, p_t(H, H_{(j)})\Big] - p_t(H, H_{(j)})H_{(j)}\Big), \tag{2}$$

This allows us to conclude that the terms are $t\mathfrak{K}$-local and exponentially decaying with t. More precisely, we have that $p_t(H, H_{(j)}) = \sum_{F_b\in\mathcal{P}_{(j),t}} c_{b,t}F_b$, where $\mathcal{P}_{(j),t} \subset \{\pm 1, \pm i\}\cdot\mathcal{P}$ and $\{j\}\cup\mathrm{supp}(F_b)$ is contained in a connected component of size $t\mathfrak{K}$ for every F_b. Further, the size of p_t is exponentially decaying with t: $\sum_{F_b\in\mathcal{P}_{(j),t}}|c_{b,t}| \leqslant 1/4^t$. This is proven in Theorem 4.1.

Since the weights are geometrically decaying, they induce a distribution over all the terms in series expansion in Eq. (1), which can be rewritten as

$$e^{-\beta H} \cdot e^{\beta(H-H_{(j)})} = I + \sum_{t=1}^{\infty} \sum_{F_b \in \mathcal{P}_{(j),t}} c_{b,t} F_b \,.$$

We will always keep I and then sample a term from the remaining sum in an unbiased manner. We do this by picking some term F_b with probability proportional to $|c_{b,t}|$ and rescaling it appropriately. Let the resulting term be denoted by $I + cE$, where $c \in \mathbb{R}$ and $E \in \{\pm 1, \pm i\} \cdot \mathcal{P}$. We have that $\mathbb{E}\left[I + cE\right] = e^{-\beta H} \cdot e^{\beta(H-H_{(j)})}$.

Pinning the first site. Next, we show that we can pin the Gibbs state at a site, say site 1, which means we fix the 2×2 density matrix on this site. We begin by decomposing the Gibbs state as follows:

$$e^{-\beta H} = e^{-\beta(H-H_{(1)})/2} \underbrace{\left(e^{\beta(H-H_{(1)})/2} \cdot e^{-\beta H/2}\right)}_{(3).(1)} \cdot \underbrace{\left(e^{-\beta H/2} \cdot e^{\beta(H-H_{(1)})/2}\right)}_{(3).(2)} e^{-\beta(H-H_{(1)})/2}. \tag{3}$$

The two terms (3).(1) and (3).(2) are operators that appear in Eq. (1), with β replaced by $\beta/2$. Therefore, we can invoke the aforementioned sampling primitive, independently, to obtain unbiased samples $I + c_1 E_1$ and $I + c_2 E_2$ such that, by linearity of expectation,

$$\begin{aligned}
&\mathbb{E}\left[e^{-\beta(H-H_{(1)})/2}(I + c_1 E_1)^\dagger \cdot (I + c_2 E_2) e^{-\beta(H-H_{(1)})/2}\right] \\
&= e^{-\beta(H-H_{(1)})/2}\, \mathbb{E}\left[(I + c_1 E_1)\right]^\dagger \cdot \mathbb{E}\left[(I + c_2 E_2)\right] e^{-\beta(H-H_{(1)})/2} = e^{-\beta H},
\end{aligned}$$

where the expectation is over the random choice of picking a particular term. First, we make the above Hermitian by averaging with its Hermitian conjugate. We want to factorize the state prepared thus far into the part that acts on site 1 and the part that acts on the remaining Hilbert space. To this end, consider the following expression:

$$\begin{aligned}
&\frac{1}{2} e^{-\beta(H-H_{(1)})/2} \left((I + c_1 E_1)^\dagger (I + c_2 E_2) + (I + c_2 E_2)^\dagger (I + c_1 E_1)\right) e^{-\beta(H-H_{(1)})/2} \\
&= e^{-\beta(H-H_{(1)})/2} \left(I + \underbrace{\left(\frac{c_1 E_1 + c_1 E_1^\dagger}{2}\right) + \left(\frac{c_2 E_2 + c_2 E_2^\dagger}{2}\right) + \left(\frac{c_1 c_2 (E_1^\dagger E_2 + E_2^\dagger E_1)}{2}\right)}_{} \right) e^{-\beta(H-H_{(1)})/2},
\end{aligned} \tag{4}$$

and sample one of the three terms uniformly at random and multiply the coefficient by 3. Again, in expectation, the resulting expression is precisely $e^{-\beta H}$. We focus on the case where we sample the first term. Since E_1 is a product of Paulis up to a factor of $\pm 1, \pm i$, the expression $c_1(E_1 + E_1^\dagger)/2$ is equal to $c_1 E_1$ when E_1 is Hermitian and otherwise is equal to 0 when E_1 is anti-Hermitian. Consider the case where it is equal to $c_1 E_1$ as this is the nontrivial case. Let $E_1 = A \otimes B$, where A acts on site 1, and $A \in \{ I, \sigma_x, \sigma_y, \sigma_z \}$. Assuming for sake of exposition $A = \sigma_x$, observe

$$I + 3c_1 E_1 = I + 3c_1(\sigma_x \otimes B) = \frac{1}{2}((I + \sigma_x) \otimes (I + 3c_1 B)) + \frac{1}{2}((I - \sigma_x) \otimes (I - 3c_1 B))$$

and yet again we sample one of the terms uniformly at random and scale it up by a factor of 2. Let the sample be $(I+\sigma_x)\otimes(I+3c_1B)$, then

$$e^{-\beta(H-H_{(1)})/2}((I+\sigma_x)\otimes(I+3c_1B))e^{-\beta(H-H_{(1)})/2}$$
$$=\underbrace{2\cdot\frac{(I+\sigma_x)}{2}}_{(5).(1)}\otimes\underbrace{\left(e^{-\beta(H-H_{(1)})/2}(I+3c_1B)e^{-\beta(H-H_{(1)})/2}\right)}_{(5).(2)}, \tag{5}$$

and thus we have managed to factorize the matrix into the part that acts on site 1 (term (5).(1)) and the part that acts on the remaining $n-1$ sites (term (5).(2))[3]. We treat the expression in (5).(1) as the "state" corresponding to site 1, and while strictly speaking this is not a state, we can think of the constant in front as adjusting the probabilities of outputting the various product states at the end of the process. Finally, note that by linearity of expectation, we have

$$\mathbb{E}\left[2\cdot\frac{(I+\sigma_x)}{2}\otimes\left(e^{-\beta(H-H_{(1)})/2}(I+3c_1B)e^{-\beta(H-H_{(1)})/2}\right)\right]=e^{-\beta H}. \tag{6}$$

Recursing on the remaining sites. Given the operator from term (5).(2), we show that we can continue to peel off another site by considering the following expression:

$$\left(e^{-\frac{\beta(H-H_{(1)})}{2}}(I+3c_1B)e^{-\frac{\beta(H-H_{(1)})}{2}}\right)$$
$$=e^{-\frac{\beta(H-H_{(1)}-H_{(2)})}{2}}\underbrace{\left(e^{\frac{\beta(H-H_{(1)}-H_{(2)})}{2}}\cdot e^{\frac{-\beta(H-H_{(1)})}{2}}\right)}_{(7).(1)}(I+3c_1B)\underbrace{\left(e^{\frac{-\beta(H-H_{(1)})}{2}}\cdot e^{\frac{\beta(H-H_{(1)}-H_{(2)})}{2}}\right)}_{(7).(2)}e^{-\frac{\beta(H-H_{(1)}-H_{(2)})}{2}}. \tag{7}$$

Recall, that we can obtain independent, unbiased samples of (7).(1) and (7).(2) using the expansion from Eq. (1). Let $I+d_1F_1$ and $I+d_2F_2$ be the resulting samples, and thus

$$\frac{1}{2}e^{-\frac{\beta(H-H_{(1)}-H_{(2)})}{2}}\Big((I+d_1F_1)^\dagger(I+3c_1B)(I+d_2F_2)+(I+d_2F_2)^\dagger(I+3c_1B)(I+d_1F_1)\Big)e^{-\frac{\beta(H-H_{(1)}-H_{(2)})}{2}} \tag{8}$$

is an unbiased estimator of $\left(e^{-\frac{\beta(H-H_{(1)})}{2}}(I+3c_1B)e^{-\frac{\beta(H-H_{(1)})}{2}}\right)$. As before, we then expand each of the terms above and show that we can uniformly sample one of them to pin the second site and so on.

Gibbs states are *separable*. To finish the proof, we crucially need to argue that our procedure produces a valid *positive* linear combination over product states. To do this, it suffices to argue that the matrix in the "middle" e.g. the $I+3c_1B$ in (6), is PSD throughout the process. Note, the perturbation to the Identity is changing as we recurse, as evidenced by expanding out the middle expression in Eq. (8). This is challenging since this perturbation is not monotonically decreasing. To control the perturbation, we show that the coefficient c_1 is bounded. We do this by introducing a carefully chosen potential function that tracks an upper bound on the size of the coefficient and inductively argue the following holds

$$|c_1|\leqslant\frac{1}{2}\left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(B)|}$$

[3]Here we overload notation and use $e^{-\beta(H-H_1)/2}$ to denote the $2^n\times 2^n$ operator that is I on the first site, and the $2^{n-1}\times 2^{n-1}$ operator with the first site factored out.

throughout this process (see Lemma 6.7). In other words, we show that in each iteration of the recursion, if the support of B grows, then the coefficient in front gets exponentially smaller, whereas if the support of B shrinks, the coefficient can grow but not too drastically. Carrying out this argument requires carefully adjusting our sampling probabilities e.g. for the different cases in (8) instead of sampling uniformly at random—see Algorithm 6.3 for more details.

2.2 Gibbs states are efficiently preparable

We now switch gears and describe how to efficiently implement each of the aforementioned steps to prepare a Gibbs state.

Efficiently sampling a single term. We begin by observing that the Taylor expansion in Eq. (1) can be truncated at $T = \mathcal{O}(\log(n/\varepsilon))$, and since the coefficients decay exponentially, we can conclude that

$$e^{-\beta H} \cdot e^{\beta\left(H - H_{(1)}\right)} = \underbrace{I + \sum_{t=1}^{T} p_t\left(H, H_{(1)}\right)}_{(9).(1)} + \Phi, \tag{9}$$

where $\|\Phi\|_{\mathrm{op}} \leqslant \mathrm{poly}(\varepsilon/n)$. Since the process terminates after $\mathrm{poly}(n)$ steps, it suffices to construct an unbiased estimator for the expression in (9).(1). As before, we always keep the identity matrix. Note, that we could in $\mathrm{poly}(n/\varepsilon)$ time, explicitly compute each term $\sum_{t=1}^{T} p_t(H, H_{(1)})$ and sample it proportional to its coefficients. However, we can actually design a faster sampling algorithm, in particular getting $\log(1/\varepsilon)$, by exploiting the recursive structure of p_t. We do as follows: let $E_0 = I$ and suppose for some $t \in [T]$, the current sample is E_t. Then, using Eq. (2),

$$p_{t+1}(H, H_{(1)}) = \frac{\beta}{t+1}\, \mathbb{E}\left[\left(-[H, E_t] - E_t \cdot H_{(1)}\right)\right],$$

and thus with probability $t/(t+1)$ we sample a single Pauli term from $[H, E_t]$, and with the remaining probability sample a single Pauli term from $E_t \cdot H_{(1)}$. We can verify that the resulting sample is unbiased and runs in time $\mathcal{O}(\log(n/\varepsilon)\,\mathrm{poly}(\mathfrak{d}, \mathfrak{K}))$ (see Lemma 4.5).

Efficiently pinning sites. Suppose invoking the aforementioned sampling primitive twice, independently, with unbiased terms $I + c_1 E_1$ and $I + c_2 E_2$. As before, we have an expression of the form:

$$e^{-\beta(H - H_{(1)})/2}\left(I + \underbrace{\left(\frac{c_1 E_1 + c_1 E_1^\dagger}{2}\right)}_{(10).(1)} + \underbrace{\left(\frac{c_2 E_2 + c_2 E_2^\dagger}{2}\right)}_{(10).(2)} + \underbrace{\left(\frac{c_1 c_2 (E_1^\dagger E_2 + E_2^\dagger E_1)}{2}\right)}_{(10).(3)}\right) e^{-\beta(H - H_{(1)})/2}, \tag{10}$$

such that in expectation, this operator is close to $e^{-\beta H}$. Note, in the structural result, we could get away with uniform sampling, since it suffices to explore the entire state space, and we do not have to adjust the probabilities carefully. Whereas, for preparing a state close to the Gibbs state, we require a density matrix, thus need to adjust the sampling weights. In particular we

need to sample each term with the following probabilities:

$$\begin{aligned}
\mathbf{Pr}[(10).(1)] &\propto \mathrm{tr}\left(e^{-\beta(H-H_{(1)})/2}\left(I+\left(\frac{c_1E_1+c_1E_1^\dagger}{2}\right)\right)e^{-\beta(H-H_{(1)})/2}\right)\\
\mathbf{Pr}[(10).(2)] &\propto \mathrm{tr}\left(e^{-\beta(H-H_{(1)})/2}\left(I+\left(\frac{c_2E_2+c_2E_2^\dagger}{2}\right)\right)e^{-\beta(H-H_{(1)})/2}\right), \qquad (11)\\
\mathbf{Pr}[(10).(3)] &\propto \mathrm{tr}\left(e^{-\beta(H-H_{(1)})/2}\left(I+\left(\frac{c_1c_2(E_1^\dagger E_2+E_2^\dagger E_1)}{2}\right)\right)e^{-\beta(H-H_{(1)})/2}\right).
\end{aligned}$$

Approximately sampling from this distribution requires estimating the expectation of marginals (in (11)). Whenever $\beta < 1/(\gamma\mathfrak{d}^2\mathfrak{K}^2)$, for a fixed constant γ, we can leverage cluster expansion analysis from [HKT22] to estimate these quantities (see Theorem 3.5). However, we note that running time for estimating the marginals grows exponentially in $\mathfrak{d}$ times the support size of the Hamiltonian. Since the support size can be as large as $\Theta(\log(n/\varepsilon))$, this would result in an algorithm with running time $(n/\varepsilon)^{\Omega(\mathfrak{d})}$.

Fast Gibbs state preparation. We show that it suffices for us to have an $\mathcal{O}(1)$-approximate estimator for $\mathrm{tr}\left(e^{-\beta H}\right)$ and the partition function for any residual Gibbs state e.g. $\mathrm{tr}\left(e^{-\beta(H-H_{(1)})}\right)$. Our reduction from ε-approximate sampling to this "weak" constant-factor approximate counting is reminiscent of the one by Jerrum and Sinclair [SJ89] and works as follows: we imagine setting up a tree over the space of possible choices that our sampling algorithm makes. These choices correspond to picking a site, expanding a series similar to Eq. (9), and sampling one term in the expansion. The tree T has depth n and

- The root note is labeled by $e^{-\beta H}$
- A node at depth k is indexed by a product state over k sites and a "Gibbs-like" state over $n-k$ sites e.g. an expression of the form in (5).(2)
- The possible children of a node are obtained by pinning one additional site in the "Gibbs-like" state and adding this site to the product state part via the procedure described earlier

Ideally, we want to sample a product state by simply walking straight down this tree. However, this requires exactly computing the probabilities in Eq. (11). Since we can only do approximate counting, we don't know the exact probabilities we should go down each of the branches in the tree. Instead, we set up a random walk on this tree that goes both up and down (i.e. we could pin some site and then later unpin it and resample it) but has the desired stationary distribution on the leaves. In particular, with respect to this distribution, the average of the product states at the leaves is close to the Gibbs state. We show that this random walk mixes quickly – it turns out that the approximation factor in our counting oracle only shows up in the mixing time. Thus we can simply run the walk until it hits a leaf and output the product state corresponding to that leaf. We remark that this tree random walk is inspired by the tree random walk used in the reduction from weak approximate counting to sampling for self-reducible problems in [SJ89].

3 Background

3.1 Linear algebra

We work in the Hilbert space $\mathbb{C}^N$ corresponding to a system of n qubits, $\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$, so that $N = 2^n$. For a matrix A, we use $A^\dagger$ to denote its conjugate transpose, $\|A\|_{\text{op}}$ to denote its operator norm, and $\|A\|_1$ to denote its trace norm; for a vector v, we use $\|v\|$ to denote its Euclidean norm. We will work with this Hilbert space, often considering it in the basis of (tensor products of) Pauli matrices.

Definition 3.1 (Pauli matrices)**.** The Pauli matrices are the following 2×2 Hermitian matrices.

$$\sigma_I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These matrices are unitary and (consequently) involutory. Further, $\sigma_x\sigma_y = \mathrm{i}\sigma_z$, $\sigma_y\sigma_z = \mathrm{i}\sigma_x$, and $\sigma_z\sigma_x = \mathrm{i}\sigma_y$, so the product of Pauli matrices is a Pauli matrix, possibly up to a factor of $\{\mathrm{i}, -1, -\mathrm{i}\}$. The non-identity Pauli matrices are traceless. We also consider tensor products of Pauli matrices, $P_1 \otimes \cdots \otimes P_n$ where $P_i \in \{\sigma_I, \sigma_x, \sigma_y, \sigma_z\}$ for all $i \in [n]$. The set of such products of Pauli matrices, which we denote $\mathcal{P}$, form an orthogonal basis for the vector space of $2^n \times 2^n$ (complex) Hermitian matrices under the trace inner product. The product of two elements of $\mathcal{P}$ is an element of $\mathcal{P}$, possibly up to a factor of $\{\mathrm{i}, -1, -\mathrm{i}\}$.

Definition 3.2 (Support of an operator)**.** For an operator $P \in \mathbb{C}^{N \times N}$ on a system of n qubits, its *support*, $\text{supp}(P) \subset [n]$ is the subset of qubits that P acts non-trivially on. That is, $\text{supp}(P)$ is the minimal set of qubits such that P can be written as $P = O_{\text{supp}(P)} \otimes I_{[n]\setminus\text{supp}(P)}$ for some operator O.

So, for example, the support of a tensor product of Paulis, $P_1 \otimes \cdots \otimes P_n$ are the set of $i \in [n]$ such that $P_i \neq \sigma_I$.

Fact 3.3. *For any square matrix B, $e^{I \otimes B} = I \otimes e^B$.*

3.2 Hamiltonians of interacting systems

Definition 3.4 (Restricted Hamiltonian)**.** For a Hamiltonian $H = \sum_{a=1}^m \lambda_a G_a$ on n sites and a subset $S \subseteq [n]$, we define

$$H^{(S)} = \sum_{a:\text{supp}(G_a)\subseteq S} \lambda_a G_a \,.$$

For an element $j \in [n]$, we define

$$H_{(j)} = \sum_{a:j\in\text{supp}(G_a)} \lambda_a G_a \,.$$

3.3 Approximating the partition function

At high temperature, we can efficiently estimate the log-partition function. We recap this result here, following the analysis in [HKT22].

Theorem 3.5 (Estimating the log-partition function)**.** *Let $H = H(\lambda) = \sum_{a=1}^m \lambda_a G_a$ be a Hamiltonian with degree $\mathfrak{d}$ and locality $\mathfrak{K}$. Let $0 \leqslant \beta \leqslant \beta_c = 1/(4e^2(\mathfrak{d}+1)^2)$. Given any $0 < \eta < 1$ we can compute an estimate $\hat{z}$ such that*

$$\log(\text{tr}(e^{-\beta H})) - \eta \leqslant \hat{z} \leqslant \log(\text{tr}(e^{-\beta H})) + \eta,$$

in time

$$n \cdot (n/\eta)^{\frac{4+\log(\mathfrak{d})}{\log(\beta_c/\beta)}} \cdot \mathfrak{K} \cdot \mathrm{polylog}(n/\eta).$$

This statement follows from cluster expansion. Specifically, we can conclude the following about the log-partition function.

Lemma 3.6 (Version of [HKT22, Theorem 3.1] for the log-partition function)**.** *Let $H = H(\lambda) = \sum_{a=1}^{m} \lambda_a G_a$ be a Hamiltonian with degree $\mathfrak{d}$ and locality $\mathfrak{K}$. Let $0 \leqslant \beta \leqslant \beta_c = 1/(4e^2(\mathfrak{d}+1)^2)$. Then the log-partition function of H can be expressed as a power series in β,*

$$\mathcal{L} := \log(\mathrm{tr}(e^{-\beta H})) = \sum_{\ell \geqslant 0} \beta^\ell p_\ell(\lambda),$$

where p_ℓ is a degree-ℓ homogeneous polynomial in λ with the following properties:

1. *p_ℓ consists of at most $ne\mathfrak{d}(1+e(\mathfrak{d}-1))^{\ell-1}$ monomials.*
2. *The coefficient in front of any monomial in p_ℓ is at most $(2e(\mathfrak{d}+1))^\ell$ in magnitude.*

Further, after $\mathcal{O}(\mathfrak{K}m\mathfrak{d}\log\mathfrak{d})$ pre-processing time, we have the following form of access to p_ℓ:

A. *The list of monomials that appear in p_ℓ can be enumerated in time $\mathcal{O}(\ell\mathfrak{d}\mu)$, where μ is the number of monomials.*

B. *The coefficient of any monomial in p_ℓ can be computed exactly in $\mathcal{O}(\mathfrak{K}\ell^3 + 8^\ell\ell^5\log^2\ell) = (8^\ell + \mathfrak{K})\,\mathrm{poly}(\ell)$ time.*

Proof. Everything needed for the proof of this is contained in Section 3 of [HKT22]. We direct the reader there for further details. Here, we only identify where our statements come from.

First, we observe that the $\mathcal{L} = \log(\mathrm{tr}(e^{-\beta H}))$ has a formal multivariate Taylor series expansion around $\lambda = (0,\dots,0)$ [HKT22, Eqs. 24 and 25],

$$\mathcal{L} = \sum_{\ell \geqslant 0} \underbrace{\sum_{\boldsymbol{V}:|\boldsymbol{V}|=\ell} \frac{\lambda^{\boldsymbol{V}}}{\boldsymbol{V}!}\mathcal{D}_{\boldsymbol{V}}\mathcal{L}}_{\beta^\ell p_\ell(\lambda)},$$

where $\boldsymbol{V}$ denotes a multiset over terms $[m]$; $\lambda^{\boldsymbol{V}} = \prod_{a\in\boldsymbol{V}} \lambda^a$ is the product of all coefficients associated to the terms in $\boldsymbol{V}$ with multiplicity; and $\mathcal{D}_{\boldsymbol{V}}\mathcal{L} = \prod_{a\in\boldsymbol{V}} \frac{\partial}{\partial\lambda_a}\mathcal{L}|_{\lambda=(0,\dots,0)}$ is the log-partition function, with derivatives taken for every λ_a with $a \in \boldsymbol{V}$ with multiplicity, evaluated at $\lambda = (0,\dots,0)$. Note that $\mathcal{D}_{\boldsymbol{V}}\mathcal{L}$ is a constant in λ. This formal expression becomes a true equality whenever the right-hand side series converges.

The coefficient, $\mathcal{D}_{\boldsymbol{V}}\mathcal{L}$, is only non-zero when $\boldsymbol{V}$ is connected [HKT22, Proposition 3.5]. Because of the degree bound $\mathfrak{d}$, the number of such "clusters" (connected $\boldsymbol{V}$) of size ℓ is merely exponential, bounded by $ne\mathfrak{d}(1+e(\mathfrak{d}-1))^{\ell-1}$ [HKT22, Proposition 3.6]. This gives the monomial bound. When it is connected, $|\mathcal{D}_{\boldsymbol{V}}\mathcal{L}| \leqslant (2e(\mathfrak{d}+1)\beta)^\ell(\boldsymbol{V}!)$ [HKT22, Proposition 3.8]. This gives the coefficient bound.

As for the running time statements, enumerating monomials amounts to enumerating clusters, which is done in [HKT22, Section 3.4]. Computing a coefficient amounts to computing $\mathcal{D}_{\boldsymbol{V}}\mathcal{L}$, which is done in [HKT22, Proposition 3.13]. □

Proof of Theorem 3.5. We use Lemma 3.6 to write $\mathcal{L} = \log(\mathrm{tr}(e^{-\beta H}))$ as a power series

$$\mathcal{L} = \sum_{\ell \geqslant 0} \beta^\ell p_\ell(\lambda),$$

and we estimate it by truncating at some level d,

$$\widetilde{\mathcal{L}} = \sum_{\ell=0}^{d} \beta^\ell p_\ell(\lambda).$$

Then, using Parts 1. and 2. of the lemma,

$$\begin{aligned}
|\mathcal{L} - \widetilde{\mathcal{L}}| &\leqslant \sum_{\ell \geqslant d+1} \beta^\ell |p_\ell(\lambda)| \\
&\leqslant \sum_{\ell \geqslant d+1} \beta^\ell \underbrace{ne\mathfrak{d}(1 + e(\mathfrak{d}-1))^{\ell-1}}_{\text{\# of monomials}} \underbrace{(2e(\mathfrak{d}+1))^\ell}_{\text{coefficient size bound}} \\
&\leqslant n \sum_{\ell \geqslant d+1} (2e^2(\mathfrak{d}+1)^2\beta)^\ell \\
&= \frac{n(2e^2(\mathfrak{d}+1)^2\beta)^{d+1}}{1 - 2e^2(\mathfrak{d}+1)^2\beta},
\end{aligned}$$

where the last step follows for $\beta < \beta_c = 1/(2e^2(\mathfrak{d}+1)^2)$, since then this series converges. This difference can be made to be η by taking

$$d = \left\lceil \frac{\log(n/((1-\beta/\beta_c)\eta)}{\log(\beta_c/\beta)} \right\rceil. \tag{12}$$

To compute $\widetilde{\mathcal{L}}$, we enumerate the list of monomials to order d and then compute all corresponding coefficients. The running time of this is dominated by the task at order d, which is bounded by

$$\underbrace{n(e\mathfrak{d})^d}_{\text{\# of clusters}} \cdot \underbrace{(8^d + \mathfrak{K})\,\mathrm{poly}(d)}_{\text{time to compute a coefficient}} \tag{13}$$

$$= \widetilde{\mathcal{O}}\Big(n\big(\frac{n}{(1-\beta/\beta_c)\eta}\big)^{\frac{\log(8e\mathfrak{d})}{\log(\beta_c/\beta)}}\mathfrak{K}\Big) \tag{14}$$

This gives the desired bound. Since we are not optimizing the constant in β_c, we take $\beta_c \leftarrow \beta_c/2$, to avoid writing the $(1-\beta/\beta_c)$ term. □

4 Low-degree polynomial approximation to a restricted Gibbs state

In this section, we decompose the matrix expression $e^{-\beta H}e^{\beta(H-H_{(j)})}$, where $H_{(j)}$ is the Hamiltonian restricted to a site j (Definition 3.4), into an infinite, exponentially decaying series. This implies that this operator is quasi-local, and that the truncation of this series, a low-degree polynomial p in the terms of H, is a good approximation to it. This allows us to extract the dependence of site j from the Gibbs state,

$$e^{-\beta H} \approx p \cdot e^{-\beta(H-H_{(j)})}.$$

Theorem 4.1 (Restricted Gibbs state series)**.** *Let $H = \sum_{a\in[m]} \lambda_a G_a$ be a $(\mathfrak{d}, \mathfrak{K})$-low-intersection Hamiltonian and let $j \in [n]$. If $\beta < \frac{1}{2C\mathfrak{d}\mathfrak{K}}$ for some universal constant $C > 1$, then we can write*

$$e^{-\beta H} \cdot e^{\beta(H-H_{(j)})} = \sum_{t=0}^{\infty} p_t(H, H_{(j)})$$

where $p_0(H, H_{(j)}) = I$ and p_t satisfies the recurrence

$$p_{t+1}(H, H_{(j)}) = \frac{\beta}{t+1}\left(-[H, p_t(H, H_{(j)})] - p_t(H, H_{(j)})H_{(j)}\right). \tag{15}$$

Furthermore, for each $t > 0$, p_t can be written as $\sum_{F_b \in \mathcal{P}_{(j),t}} c_{b,t} F_b$ where $\sum_{F_b \in \mathcal{P}_{(j),t}} |c_{b,t}| \leqslant \frac{1}{C^t}$, and

$$\mathcal{P}_{(j),t} = \Big\{ P \in \{\pm 1, \pm i\} \cdot \mathcal{P} \;\Big|\; \{j\} \cup \mathrm{supp}(P) \subset S \text{ for some } S \subset [n] \text{ which is connected in } \mathfrak{G} \text{ and satisfies } |S| \leqslant t\mathfrak{K} \Big\}.$$

Proof. We can write

$$\begin{aligned} e^{-\beta H} e^{\beta(H - H_{(j)})} &= \Big(\sum_{k=0}^{\infty} \frac{\beta^k (-H)^k}{k!}\Big)\Big(\sum_{\ell=0}^{\infty} \frac{\beta^\ell (H - H_{(j)})^\ell}{\ell!}\Big) \\ &= \sum_{t=0}^{\infty} \frac{\beta^t}{t!} \sum_{k=0}^{t} \frac{(-H)^k (H - H_{(j)})^{t-k} t!}{k!(t-k)!} \\ &= \sum_{t=0}^{\infty} \frac{\beta^t}{t!} \underbrace{\sum_{k=0}^{t} \binom{t}{k} (-H)^k (H - H_{(j)})^{t-k}}_{f_t(H, H_{(j)})}. \end{aligned} \tag{16}$$

Now observe that $f_0(H, H_{(j)}) = I$ and $f_t(H, H_{(j)})$ satisfies the recurrence

$$\begin{aligned} f_t(H, H_{(j)}) &= -H f_{t-1}(H, H_{(j)}) + f_{t-1}(H, H_{(j)})(H - H_{(j)}) \\ &= -[H, f_{t-1}(H, H_{(j)})] - f_{t-1}(H, H_{(j)}) H_{(j)}. \end{aligned} \tag{17}$$

Now let $p_t(H, H_{(j)}) = \frac{\beta^t}{t!} f_t(H, H_{(j)})$. From the above, we immediately get (15). Now we will prove the desired properties by induction. The base case is clear. Now for the inductive step, assume that we have proven the desired properties up to some t. Then we can write

$$p_t(H, H_{(j)}) = \sum_{F_b \in \mathcal{P}_{(j),t}} c_{b,t} F_b.$$

The above recurrence implies

$$p_{t+1}(H, H_{(j)}) = \frac{\beta}{t+1} \sum_{F_b \in \mathcal{P}_{(j),t}} c_{b,t} \left(-[H, F_b] - F_b H_{(j)}\right).$$

We now consider the terms in two parts. First, recall $H = \sum_a \lambda_a G_a$ where the G_a form a bounded degree graph on the sites. We can write

$$[H, F_b] = \sum_{a : \mathrm{supp}(G_a) \cap \mathrm{supp}(F_b) \neq \emptyset} \lambda_a [G_a, F_b] \tag{18}$$

since the commutator $[G_a, F_b]$ is zero when the supports of G_a and F_b don't intersect. Now each commutator $[G_a, F_b]$ is equal to $2P$ for some $P \in \{\pm 1, \pm i\} \cdot \mathcal{P}$. Furthermore, by the inductive hypothesis, $\{j\} \cup \mathrm{supp}(P) \subseteq (\{j\} \cup \mathrm{supp}(F_b)) \cup \mathrm{supp}(G_a)$ must be contained in some connected component of size at most $(t+1)\mathfrak{K}$. Also, in (18), there are at most $\mathfrak{d}|\mathrm{supp}(F_b)| \leqslant \mathfrak{d} t \mathfrak{K}$ nonzero terms by the inductive hypothesis. Next, note that

$$F_b H_{(j)} = \sum_{a : j \in \mathrm{supp}(G_a)} \lambda_a F_b G_a \tag{19}$$

and $F_b G_a \in \{\pm 1, \pm i\} \cdot \mathcal{P}$ and $\mathrm{supp}(F_b G_a) \subseteq \mathrm{supp}(G_a) \cup \mathrm{supp}(F_b)$ which again, by the inductive hypothesis, must be contained in some connected component of size at most $(t+1)\mathfrak{K}$ that also contains $\{j\}$. Also, by assumption, the above sum has at most $\mathfrak{d}$ terms. Combining the two parts (18) and (19), we can write

$$p_{t+1}(H, H_{(j)}) = \sum_{F_b \in \mathcal{P}_{(j),t+1}} c_{b,t+1} F_b$$

where all elements $F_b \in \mathcal{P}_{(j),t+1}$ have $\mathrm{supp}(F_b)$ contained in some connected component of size at most $(t+1)\mathfrak{K}$ that also contains $\{j\}$. Finally, by the inductive hypothesis,

$$\sum_{F_b \in \mathcal{P}_{(j),t+1}} |c_{b,t+1}| \leqslant \frac{\beta}{t+1} \sum_{F_b \in S_t} |c_b|(\mathfrak{d} + 2t\mathfrak{d}\mathfrak{K}) \leqslant \frac{1}{C^{t+1}},$$

with our setting of β, which completes the proof. □

Algorithm 4.2 (Recursively sampling a term)**.**

Input: A $(\mathfrak{d}, \mathfrak{K})$-low-intersection Hamiltonian $H = \sum_a \lambda_a G_a$, parameters β, site $j \in [n]$

Input: Integer $k \geqslant 0$.

Operations:

1. Initialize $c_0 = 1, E_0 = I$
2. For $t = \{0, 1, \ldots, k-1\}$
   (a) Sample $g \in \{0, 1, \ldots, t\}$
   (b) If $g > 0$
      i. Sample a random element $j' \in \mathrm{supp}(E_t)$
      ii. Let S be the set $\{(\lambda_a, G_a)\}_{a \in [m], j' \in \mathrm{supp}(G_a)}$
      iii. Sample each element of S independently with probability $1/\mathfrak{d}$ and otherwise sample the pair $(0, I)$
      iv. If we sample $(0, I)$, set $c_{t+1} = 0, E_{t+1} = I$
      v. Otherwise let (λ_a, G_a) be the pair we sampled and let $k = |\mathrm{supp}(G_a) \cap \mathrm{supp}(E_t)|$
      vi. Set
      $$c_{t+1} = \frac{2\beta\mathfrak{d} c_t \lambda_a |\mathrm{supp}(E_t)|}{tk}, \quad E_{t+1} = \frac{-[G_a, E_t]}{2}.$$
   (c) If $g = 0$
      i. Let S be the set $\{(\lambda_a, G_a)\}_{a \in [m], j \in \mathrm{supp}(G_a)}$
      ii. Sample each element of the set S with probability $1/\mathfrak{d}$ and otherwise sample the pair $(0, I)$
      iii. If we sample $(0, I)$, set $c_{t+1} = 0, E_{t+1} = I$

iv. Otherwise let (λ_a, G_a) be the pair we sampled and set

$$c_{t+1} = \beta \mathfrak{d} c_t \lambda_a \ , \ E_{t+1} = -E_t G_a$$

Output: c_k, E_k

In light of Theorem 4.1, we make the following definition:

Definition 4.3 (Truncating the polynomial series)**.** For any integer $k \geqslant 0$ and parameter β, we define $T_{k,\beta}(H, H_{(j)}) = \sum_{t=0}^{k} p_t(H, H_{(j)})$ where p_t is as constructed in Theorem 4.1.

We now give a fast algorithm for sampling a single Pauli term from one of the polynomials in Theorem 4.1. This subroutine will be useful in our sampling algorithm later on. We will assume that we are given the following form of access to the terms of H. For any $j \in [n]$, we can enumerate the collection $\{(\lambda_a, G_a)\}_{a \in [m], j \in \mathrm{supp}(G_a)}$ in $O(\mathfrak{d})$ time.

Remark 4.4. Note that in both cases $|S| \leqslant \mathfrak{d}$ by the assumption on the degree of H so the distributions specified are valid.

Lemma 4.5 (Sample access to terms of $p_{t,j}$)**.** *In the same setting as Theorem 4.1, for any integer $t \geqslant 0$, if we run Algorithm 4.2 with $k \leftarrow t$, then it runs in $(t+1) \cdot \mathrm{poly}(\mathfrak{K}, \mathfrak{d})$ time and outputs a $c \in \mathbb{R}$ with $|c| \leqslant \frac{1}{C^t}$ and $E \in \{\pm 1, \pm i\} \cdot \mathcal{P}$ such that*

$$\mathbb{E}[cE] = p_t(H, H_{(j)})$$

and $\mathrm{supp}(E)$ *is contained in some connected component of size at most $t\mathfrak{K}$ that also contains $\{j\}$ in the underlying graph.*

Proof. We prove the lemma by induction. For the base case of $t = 0$ we simply output $c = 1, E = I$. Now we show how to go from t to $t+1$. By the inductive hypothesis, we can assume that after t iterations in the algorithm c_t, E_t satisfies $\mathbb{E}[c_t E_t] = p_t(H, H_{(j)})$ and $|c_t| \leqslant \frac{1}{C^t}, E_t \in \{\pm 1, \pm i\} \cdot \mathcal{P}$. Now recall that

$$p_{t+1}(H, H_{(j)}) = \frac{\beta}{t+1} \left(-[H, p_t(H, H_{(j)})] - p_t(H, H_{(j)}) H_{(j)} \right) .$$

Now in Algorithm 4.2, there are two cases for g which we call case 1 and case 2. We show that in case 1, which occurs with probability $\frac{t}{t+1}$, we are sampling a single Pauli term from $-[H, E_t]$ and in case 2, which occurs with probability $\frac{1}{t+1}$, we are sampling a single Pauli term from $-E_t H_{(j)}$. Now we analyze the two cases.

Case 1: In this case, recall that either $c_{t+1} = 0, E_{t+1} = I$ or

$$c_{t+1} = \frac{2\beta \mathfrak{d} c_t \lambda_a |\mathrm{supp}(E_t)|}{tk} \ , \ E_{t+1} = \frac{-[G_a, E_t]}{2} .$$

Also recall $k = |\mathrm{supp}(G_a) \cap \mathrm{supp}(E_t)|$. Note that by the inductive hypothesis, $\mathrm{supp}(E_{t+1})$ is contained in some connected component of size at most $(t+1)\mathfrak{K}$ that also contains $\{j\}$ and

$$|c_{t+1}| \leqslant 2\beta \mathfrak{d} \mathfrak{K} |c_t| \leqslant \frac{1}{C^{t+1}} .$$

Also, we can compute the expectation of $c_{t+1} E_{t+1}$ in this case. For a fixed term G_a, the probability of sampling it is exactly equal to $\frac{|\mathrm{supp}(G_a) \cap \mathrm{supp}(E_t)|}{\mathfrak{d} |\mathrm{supp}(E_t)|}$. Note that for all terms G_a where $|\mathrm{supp}(G_a) \cap$

$\mathrm{supp}(E_t)| = 0$, we also have $[G_a, E_t] = 0$. Summing over all choices of G_a, by linearity of expectation, we conclude

$$\mathbb{E}[cE| \text{ case 1}] = -\frac{\beta c_t}{t}[\sum_a \lambda_a G_a, E_t] = -\frac{\beta c_t}{t}[H, E_t]\,. \tag{20}$$

Case 2: In this case, recall that either $c_{t+1} = 0, E_{t+1} = I$ or

$$c_{t+1} = \beta \mathfrak{d} c_t \lambda_a\,,\ E_{t+1} = -E_t G_a\,.$$

It is clear that this setting of parameters satisfies the desired inductive statement. It remains to compute the expectation. By linearity, we have

$$\mathbb{E}[c_{t+1}E_{t+1}| \text{ case 2}] = -\beta c_t E_t \left(\sum_{a, j \in \mathrm{supp}(G_a)} \lambda_a G_a\right) = -\beta c_t E_t H_{(j)}\,. \tag{21}$$

Putting (20) and (21) together, we have

$$\mathbb{E}[c_{t+1}E_{t+1}] = \frac{\beta}{t+1}\left(-[H, c_t E_t] - c_t E_t H_{(j)}\right)\,.$$

Thus, if $c_t E_t$ was drawn from a distribution such that $\mathbb{E}[c_t E_t] = p_t(H, H_{(j)})$, then $\mathbb{E}[c_{t+1}E_{t+1}] = p_{t+1}(H, H_{(j)})$ as desired. Each iterative step of the sampling can be implemented in time $\mathrm{poly}(\mathfrak{K}, \mathfrak{d})$ so we get the desired runtime bound and this completes the proof. □

We now have a basic primitive for sampling a single term of the form $I_{2^n} + cE$, for $c \in \mathbb{R}, E \in \{\pm 1, \pm i\} \cdot \mathcal{P}$ whose expectation is $T_{t_{\max},\beta}(H, H_{(j)})$ (which is an approximation to $e^{-\beta H} \cdot e^{\beta(H - H_{(j)})}$, recall Definition 4.3) for some parameters $t_{\max}, \beta$.

Algorithm 4.6 (Sampling a monomial)**.**

Input: Hamiltonian $H = \sum_a \lambda_a G_a$, parameters $\beta, \mathfrak{K}, \mathfrak{d}$, site $j \in [n]$, threshold $t_{\max}$

Operations:

1. Sample $t \sim \{0, 1, 2, \ldots, t_{\max}\}$ with probabilities $\{1 - s, \frac{1}{3}, \left(\frac{1}{3}\right)^2, \left(\frac{1}{3}\right)^3, \ldots, \left(\frac{1}{3}\right)^{t_{\max}}\}$ (where s is such that the probabilities sum to 1)
2. If $t = 0$, set $c = 0, E = I$
3. Otherwise, run Algorithm 4.2 on H, j with parameter $k \leftarrow t$ to obtain c, E

Output: $t, I + 3^t cE$

Lemma 4.7 (Unbiased sample)**.** *The output of Algorithm 4.6 satisfies*

$$\mathbb{E}[I + 3^t cE] = T_{t_{\max},\beta}(H, H_{(j)})\,.$$

Proof. By linearity and the guarantees of Lemma 4.5,

$$\mathbb{E}[I + 3^t cE] = I + \sum_{t=1}^{t_{\max}} p_t(H, H_{(j)}) = T_{t_{\max},\beta}(H, H_{(j)})$$

where the last step follows from Theorem 4.1. □

4.1 Additional structural properties of restricted Gibbs states

We prove a few additional structural properties about Gibbs states. We begin with a few basic facts about the Gibbs state corresponding to the residual Hamiltonian, $H - H_{(j)}$, where $H_{(j)}$ is the Hamiltonian restricted to site j,

Fact 4.8 (Factoring out a site). *Let $H = \sum_{a\in[m]} \lambda_a G_a$ be a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$. Let $j \in [n]$. Then $e^{-\beta(H-H_{(j)})} = I \otimes e^{-\beta H^{([n]\setminus j)}}$.*

Proof. Recall that by definition $H - H_{(j)}$ is a Hamiltonian on n sites that acts trivially on site j and is equal to $I \otimes H^{([n]\setminus j)}$. Then we are done by Fact 3.3. □

Now we show that the Gibbs state corresponding to $H - H_{(j)}$ has roughly the same spectrum as the Gibbs state corresponding to H.

Lemma 4.9 (Spectrum of residual Hamiltonians). *Let $H = \sum_{a\in[m]} \lambda_a G_a$ be a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$. Let $j \in [n]$. If $\beta < \frac{1}{2C\mathfrak{d}\mathfrak{K}}$ for some constant $C > 1$ then*

$$\left(1 - \frac{3}{2C-1}\right) e^{-\beta H} \preceq e^{-\beta(H-H_{(j)})} \preceq \left(1 + \frac{3}{2C-1}\right) e^{-\beta H}$$

where we view $H - H_{(j)}$ as a $2^n \times 2^n$ matrix that acts trivially on site j.

Proof. By Theorem 4.1, there is a matrix X with $\|X\|_{\text{op}} \leqslant \frac{1}{2C-1} \leqslant 1$ such that

$$e^{-\beta(H-H_{(j)})/2} = e^{-\beta H/2}(I + X)\,.$$

Taking the Hermitian conjugate of both sides, we also have

$$e^{-\beta(H-H_{(j)})/2} = (I + X^\dagger)e^{-\beta H/2}\,.$$

Now we have

$$\left\|(I + X)(I + X^\dagger)\right\|_{\text{op}} \leqslant 1 + 3\,\|X\|_{\text{op}} \leqslant 1 + \frac{3}{2C-1}$$

and thus,

$$(I + X)(I + X^\dagger) \preceq \left(1 + \frac{3}{2C-1}\right) I$$

which implies

$$e^{\beta H/2} \cdot e^{-\beta(H-H_{(j)})} \cdot e^{\beta H/2} = (I + X)(I + X^\dagger) \preceq \left(1 + \frac{3}{2C-1}\right) I$$

proving the upper bound. The lower bound is proven similarly. □

Next, we prove a sharper statement. Recalling the construction of $T_{t,\beta/2}(H, H_{(j)})$ in Definition 4.3, we show that left and right multiplying a matrix by $e^{-\beta H/2}$ is very close to the same thing as left and right multiplying by $e^{-\beta(H-H_{(j)})/2} T_{t,\beta/2}(H, H_{(j)})^\dagger$ and its Hermitian conjugate.

Lemma 4.10 (Peeling the residual Gibbs state). *Let P be a $2^n \times 2^n$ Hermitian matrix such that $0.5 I_{2^n} \preceq P \preceq 2 I_{2^n}$ and let $t \geqslant 0$ be an integer. Let $H = \sum_{a\in[m]} \lambda_a G_a$ be a $\mathfrak{K}$-local Hamiltonian with*

degree $\mathfrak{d}$ *and let* $H_{(j)}$ *be the restriction to site* $j \in [n]$*. Given any* $\beta < \frac{1}{2C\mathfrak{d}\mathfrak{K}}$*, let* $Z_{H,\beta} = e^{-\beta H/2}$*. for some constant* $C > 1$*, we have*

$$\left(1 - \frac{40}{C^t}\right) Z_{H,\beta} \cdot P \cdot Z_{H,\beta}$$
$$\preceq Z_{H-H_{(j)},\beta} \cdot T_{t,\beta/2}(H, H_{(j)})^\dagger \cdot P \cdot T_{t,\beta/2}(H, H_{(j)}) \cdot Z_{H-H_{(j)},\beta} \preceq \left(1 + \frac{40}{C^t}\right) Z_{H,\beta} \cdot P \cdot Z_{H,\beta}\,,$$

where $T_{t,\beta/2}$ *is the truncation defined in Definition 4.3.*

Proof. It follows from Theorem 4.1 that

$$e^{-\beta H/2} \cdot e^{\beta(H-H_{(j)})/2} = T_{t,\beta/2}(H, H_{(j)}) + E$$

for some $E \in \mathbb{C}^{2^n \times 2^n}$ such that $\|E\|_{\mathrm{op}} \leqslant \frac{1}{(2C)^t}$. Now we can rewrite the LHS as

$$Z_{H,\beta} P Z_{H,\beta} - Z_{H-H_{(j)},\beta} \cdot T_{t,\beta/2}(H, H_{(j)})^\dagger \cdot P \cdot T_{t,\beta/2}(H, H_{(j)}) \cdot Z_{H-H_{(j)},\beta}$$
$$= Z_{H-H_{(j)},\beta}\Big(T_{t,\beta/2}(H, H_{(j)})^\dagger \cdot P \cdot E + E^\dagger \cdot P \cdot T_{t,\beta/2}\Big) Z_{H-H_{(j)},\beta} + Z_{H-H_{(j)},\beta} E^\dagger \cdot P \cdot E Z_{H-H_{(j)},\beta}\,.$$

Now consider multiplying the above by $e^{\beta H/2} = Z_{H,\beta}^{-1}$ on both sides. Note that Lemma 4.9 implies that

$$\left\| Z_{H,\beta}^{-1} Z_{H-H_{(j)},\beta} \right\|_{\mathrm{op}} \leqslant 2$$

and also $\left\| T_{t,\beta/2}(H, H_{(j)})^\dagger \right\|_{\mathrm{op}} \leqslant 2$. Thus,

$$\left\| Z_{H,\beta}^{-1} \cdot Z_{H-H_{(j)},\beta} \cdot \left(T_{t,\beta/2}(H, H_{(j)})^\dagger \cdot P \cdot E + E^\dagger \cdot P \cdot T_{t,\beta/2}(H, H_{(j)})\right) \cdot Z_{H-H_{(j)},\beta} \cdot Z_{H,\beta}^{-1} \right\|_{\mathrm{op}} \leqslant \frac{16\,\|P\|_{\mathrm{op}}}{(2C)^t}$$

and also

$$\left\| Z_{H,\beta}^{-1} \cdot Z_{H-H_{(j)},\beta} \cdot (E^\dagger P E) \cdot Z_{H-H_{(j)},\beta} \cdot Z_{H,\beta}^{-1} \right\|_{\mathrm{op}} \leqslant \frac{4\,\|P\|_{\mathrm{op}}}{(2C)^{2t}}\,.$$

Thus,

$$\frac{-20}{(2C)^t} I_{2^n} \preceq P - Z_{H,\beta}^{-1} \cdot Z_{H-H_{(j)},\beta} \cdot T_{t,\beta/2}(H, H_{(j)})^\dagger \cdot P \cdot T_{t,\beta/2}(H, H_{(j)}) \cdot Z_{H-H_{(j)},\beta} \cdot Z_{H,\beta}^{-1} \preceq \frac{20}{(2C)^t} I_{2^n}\,.$$

Finally, recalling the assumption about P and left and right multiplying the above by $e^{-\beta H/2}$ gives the desired relations. □

5 Random walks on trees

As a subroutine of our main sampling algorithm, we will design a (classical) random walk on a tree. In this section, we present some general machinery for analyzing the mixing times of random walks on trees. We begin with the definition of a *weighted tree*.

Definition 5.1 (Weighted tree)**.** Let T be a tree of depth n with a unique root such that all root-to-leaf paths have length exactly n. A weighted tree (T, w) of depth n is obtained by assigning some non-negative weight w_v to each leaf v of the tree. Further, the weight of each interior node v is equal to the sum of the weights of the leaves in the sub-tree rooted at v.

Next, we assume access to the following sampling sub-routine:

Definition 5.2 (Sampling a sub-tree)**.** For a weighted tree, we define a sample query as querying a node v and sampling one of its children with probabilities proportional to their weights.

Since the weights on each leaf are non-negative, they induce a probability distribution over the leaves as follows:

Definition 5.3 (Leaf distribution)**.** For a weighted tree, we say the leaf-distribution is the distribution over leaves where each leaf is sampled proportional to its weight.

We then recall the definition of a Markov chain and the corresponding stationary distribution.

Definition 5.4 (Transition matrix and stationary distribution)**.** A Markov chain on N states is given by a transition matrix P with entries P_{ij} for $i, j \in [N]$ given by the probability of moving from state i to state j. We define the stationary distribution, denoted by $\pi = (\pi_1, \ldots, \pi_N)$, such that $P\pi = \pi$.

Definition 5.5 (Ergodic and time-reversible Markov chain)**.** A Markov chain P is *ergodic* there exists a positive integer z such that P^z is entry-wise positive. It is *time-reversible* if its stationary distribution π satisfies

$$P_{ij}\pi_i = P_{ji}\pi_j$$

for all $i, j \in [N]$.

Next, we define the notion of conductance of a Markov chain.

Definition 5.6 (Conductance)**.** Given a Markov chain on N states with transition matrix P and stationary distribution π, for any subset $S \subset [N]$, the conductance Φ_S is defined by

$$\Phi_S = \frac{\sum_{i \in S, j \notin S} P_{ij}\pi_i}{\sum_{i \in S} \pi_i}.$$

The global conductance of the chain is defined by $\Phi = \min_{C_S \leqslant 1/2} \Phi_S$, where $C_S = \sum_{i \in S} \pi_i$.

A classical result of Jerrum and Sinclair [SJ89] bounds the spectral gap of an ergodic, time-reversible Markov chain as a function of the conductance.

Lemma 5.7 (Spectral gap of a Markov chain [SJ89])**.** *For an ergodic time-reversible Markov chain P, if we order the eigenvalues of P as $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_N$ where $\lambda_1 = 1$, then*

$$\lambda_1 \leqslant 1 - \frac{\Phi^2}{2}.$$

The main result in this section, stated in Theorem 5.8 below, is about relating two different weighted trees on the same vertex set, which we denote by (T, w) and (T, w'). For a vertex v, the distortion between the two weight functions is just w_v / w'_v. Note that scaling a weight function by a constant factor doesn't affect any of the resulting distributions. Given any edge (u, v) on this tree, we assume that the distortion between w and w' along this edge is bounded i.e. $0.1 \leqslant (w_u / w_v) \cdot (w'_u / w'_v) \leqslant 10$. With this assumption, we show that given sample access to w' and exact access to w_v / w'_v at the leaves, but only a constant approximate oracle for w_v / w'_v at interior nodes, we can actually efficiently sample from the leaf distribution of w via a Markov chain that mixes quickly. This is closely related to the reduction from weak approximate counting to sampling for self-reducible problems in [SJ89].

Theorem 5.8 (Sampling a leaf via a Markov chain)**.** *Let $(T, w), (T, w')$ be weighted trees of depth n on the same vertex set such that the arity of the tree is k, i.e. each vertex has at most k children. Assume*

that for any adjacent nodes $u, v \in T$, $0.1 \leqslant (w_u w'_v)/(w_v w'_u) \leqslant 10$. Further, assume we are given an oracle that responds to the following types of queries

- *For any internal node v, compute an estimate $\widehat{r}_v$ such that $0.1(w_v/w'_v) \leqslant \widehat{r}_v \leqslant 10(w_v/w'_v)$.*
- *For any leaf node v, exactly compute $\widehat{r}_v = w_v/w'_v$.*
- *Responds to sample queries for (T, w') (see Definition 5.2).*

Then, for any $0 < \varepsilon, \delta < 1$, there exists an algorithm that uses $\mathcal{O}(n^4 \log(nk/\varepsilon) \log(1/\delta))$ queries to the aforementioned oracle, and

- *Outputs a leaf with probability at least $1 - \delta$.*
- *When the algorithm outputs a leaf, its distribution is ε-close to the leaf distribution of (T, w) in TV distance*
- *Otherwise the algorithm outputs $\perp$.*

Proof. Consider the following random walk on (T, w): if we are at a vertex v, let u be its parent and then:

- With probability $0.01 \cdot \widehat{r}_u/\widehat{r}_v$, go to u
- With probability 0.01, sample query (T, w') for a child of v and go to that child
- Otherwise remain at v.

Note that for all vertices, we query the oracle once for $\widehat{r}_v$ and always use the same estimate throughout the random walk i.e. we never query the same vertex again for a new estimate. Note that by assumption, $\widehat{r}_u/\widehat{r}_v \leqslant 10$ so this walk is well-defined.

First, we prove that the stationary distribution of this walk has probability mass on each vertex proportional to $\widehat{r}_v w'_v$. In particular, since $\widehat{r}_v = \frac{w_v}{w'_v}$ on any leaf v, this is exactly the distribution proportional to w on the leaves.

We do this by verifying reversibility (see Definition 5.5). Consider any two vertices u, v such that u is a parent of v. Then we have

$$P_{uv}\pi_v = 0.01\frac{\widehat{r}_u}{\widehat{r}_v}\widehat{r}_v w'_v = 0.01\widehat{r}_u w'_v = 0.01\frac{w'_v}{w'_u}\widehat{r}_u w'_u = P_{vu}\pi_u$$

as desired. Thus we can conclude that the Markov chain is ergodic and time-reversible.

Next, we will lower bound the spectral gap of this walk by bounding the conductance and applying Lemma 5.7. To lower bound the conductance we show that it suffices to consider when the subset S is a subtree rooted at some vertex v. To see this, consider any cut (T_1, T_2) such that T_2 contains the root. Let $\mathcal{M} = \{ v_i \}_{i \in [r]}$ be the maximal elements in T_1, i.e. for each v_i, the parent of v_i , denoted by u_i, is in T_2. Then,

$$\Phi_{T_1} \geqslant \frac{\sum_{v_i \in \mathcal{M}} P_{v_i u_i}\pi_{v_i}}{\sum_{v_i \in T_1} \pi_{v_i}} \geqslant \frac{\sum_{v_i \in \mathcal{M}} P_{v_i u_i}\pi_{v_i}}{\sum_{v_i \in \mathcal{M}} \sum_{v_j \in \text{ sub-tree } (v_i)} \pi_{v_j}} \geqslant \frac{P_{v_i u_i}\pi_{v_i}}{\sum_{v_j \in \text{ sub-tree } (v_i)} \pi_{v_j}}$$

Let S be sub-tree rooted at v and let u be the parent. Then

$$\Phi_S \geqslant \frac{0.1\widehat{r}_u w'_v}{\sum_{v' \preceq v} \widehat{r}_v w'_v} \geqslant \frac{w_u w'_v}{40 w'_u (\sum_{v' \preceq v} w_v)} \geqslant \frac{w_u w'_v}{40 n w'_u w_v} \geqslant \frac{1}{80n}$$

where in the above, we used that $\sum_{v' \preceq v} w_v \leqslant n w_v$ from the definition of a weighted tree of depth n. Thus, by Lemma 5.7, the spectral gap of the Markov chain is $\Omega(1/n^2)$. Finally, the number of nodes in the tree is at most $(k+1)^n$ so after $O(n^3 \log(k/\varepsilon))$ steps of the Markov chain, the distribution will be ε-close to the stationary distribution (see for instance [LP17]). Note that the stationary distribution matches the leaf distribution of (T, w) on the leaves and also the probability of being at a leaf in the stationary distribution is at least

$$\frac{\sum_{v \text{ leaf}} \widehat{r}_v w'_v}{\sum_v \widehat{r}_v w'_v} \leqslant \frac{1}{4} \frac{\sum_{v \text{ leaf}} w_v}{\sum_v w_v} \geqslant \frac{1}{4n}.$$

Thus, we can first run $O(n^3 \log(nk/\varepsilon))$ steps of the Markov chain to get $\varepsilon/(100n)$-close to the stationary distribution and then run epochs of $O(n^3 \log(nk/\varepsilon))$ steps until we reach a leaf. We output as soon as we hit a leaf. With probability $1 - \delta$, we will hit a leaf within $O(n \log(1/\delta))$ epochs and this gives the desired output. □

6 Fast state preparation and analysis

Next, we describe our full sampling algorithm. We construct a tree of depth n. Each node is labeled with a tuple (S, α, B, M) consisting of

- A subset $S \subseteq [n]$
- A multiplier $\alpha \in \mathbb{R}$
- A matrix B that is a product state over sites in $[n] \backslash S$.
- A matrix $M = (I_{2^{|S|}} + cE)$ where $c \in \mathbb{R}$ and $E \in \mathcal{P}$ is a Hermitian matrix acting on the sites in S

The root is labeled with $S = [n], \alpha = 1, B = 1, M = I_{2^n}$. We now show in Algorithm 6.3 how to construct the children of a given node. At a high-level, we can interpret the children of each node as obtained by adding one additional qubit to the product state B and marginalizing it out from the matrix M.

Remark 6.1. In the above, we use $B \otimes_j A$ to denote a product state with A in the j-th qubit and B on the remaining qubits.

Note that each time we run the sampling procedure in Algorithm 6.3 on say (S, α, B, M), the child node has a set $S' = S \backslash j$ for some $j \in S$, matrix B' that is a product state over B and one additional qubit, and matrix M acting on one fewer qubit. The sampling procedure gives us a way to construct a tree of depth n.

Definition 6.2 (Sample tree of a Hamiltonian)**.** Given a Hamiltonian $H = \sum_a \lambda_a E_a$ over n sites and parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that H is a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$, we construct its sample tree as follows:

- The root node is labeled with $(S, \alpha, B, M) = ([n], 1, 1, I_{2^n})$
- For each node with a nonempty subset S, its children are labeled with the possible outcomes of running Algorithm 6.3 on that node.

Our overall strategy is as follows. We will first assign a matrix to each of the nodes of the sample tree. Specifically, for a node v indexed by $(S_v, \alpha_v, B_v, M_v)$, we assign it the matrix

$$Q_v = \alpha_v B_v \otimes e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}.$$

To interpret this matrix, note that it is a $2^n \times 2^n$ matrix. The matrix B_v is over the qubits in $[n]\backslash S_v$ and the latter part is over the qubits in S_v. Roughly at this node, we think of having "pinned" all qubits in $[n]\backslash S$ to be a product state B_v and are left with a "Gibbs-like" state on the qubits in S_v (note if M_v were identity, then it would actually be a multiple of the Gibbs state on S_v).

Algorithm 6.3 (Pinning a single site)**.**

Input: Hamiltonian $H = \sum_a \lambda_a E_a$ over set of sites $[n]$, parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that $\beta \leqslant 1/(100\mathfrak{d}\mathfrak{K}^2)$

Input: Accuracy parameter ε

Input: Set $S \subseteq [n]$ of unpinned sites, multiplier α, matrices $B, I_{2^{|S|}} + cE$

1. Let $H^{(S)}$ be the effective Hamiltonian on the unpinned sites. Choose an unpinned site $j \in \mathrm{supp}(E)$ or arbitrarily if $\mathrm{supp}(E) = \emptyset$.
2. Run Algorithm 4.6 with inputs $H^{(S)}, j, \beta/2, \mathfrak{K}, \mathfrak{d}, t_{\max} = 10\log(n/\varepsilon)$ twice (independently) to obtain $t_1, I_{2^{|S|}} + c_1 E_1$ and $t_2, I_{2^{|S|}} + c_2 E_2$.
3. Sample $b \in \{0,1,2,3,4,5,6\}$ with probabilities $\{1 - \frac{1}{\mathfrak{K}}, \frac{1}{6\mathfrak{K}}, \frac{1}{6\mathfrak{K}}, \frac{1}{6\mathfrak{K}}, \frac{1}{6\mathfrak{K}}, \frac{1}{6\mathfrak{K}}, \frac{1}{6\mathfrak{K}}\}$. Re-weight the sample as follows:
   - If $b = 0$, set $c' = \frac{c}{1-1/\mathfrak{K}}, E' = E$
   - If $b = 1$ set $c' = 6\mathfrak{K}c_1, E' = (E_1^\dagger + E_1)/2$
   - If $b = 2$ set $c' = 6\mathfrak{K}c_2, E' = (E_2^\dagger + E_2)/2$
   - If $b = 3$ set $c' = 6\mathfrak{K}cc_1, E' = (E_1^\dagger E + EE_1)/2$
   - If $b = 4$ set $c' = 6\mathfrak{K}cc_2, E' = (E_2^\dagger E + EE_2)/2$
   - If $b = 5$ set $c' = 6\mathfrak{K}c_1c_2, E' = (E_2^\dagger E_1 + E_1^\dagger E_2)/2$
   - If $b = 6$ set $c' = 6\mathfrak{K}cc_1c_2, E' = (E_2^\dagger EE_1 + E_1^\dagger EE_2)/2$
4. Write $E' = E'_j \otimes E'_{S\backslash j}$ where $E'_j \in \{I, \sigma_x, \sigma_y, \sigma_z\}$ is the jth Pauli matrix in E' and $E'_{S\backslash j} \in \mathcal{P}$ is the product of the Paulis over the remaining $|S| - 1$ sites
5. Round c' down to the nearest integer multiple of $\varepsilon/(100n)$.
6. If $E'_j = I$ then set $A = I/2, c'' = c'$.
7. If $E'_j \in \{\sigma_x, \sigma_y, \sigma_z\}$ set $A = (I + E'_j)/2, c'' = c'$ or $A = (I - E'_j)/2, c'' = -c'$ with probability $1/2$ each.
8. If $E'_{S\backslash j} = \pm I_{2^{|S|-1}}$, set $\alpha \leftarrow \alpha(1 \pm c'')$, $c'' \leftarrow 0$
9. Set $\alpha \leftarrow 2\alpha$, to account for uniform sampling at step 7.

Output: Set $S\backslash j$, multiplier α, matrices $B \otimes_j A, I_{2^{|S|-1}} + c''E'_{S\backslash j}$.

Now going from a node to its children involves pinning one more site. The key lemma, Lemma 6.13, shows that the average of $Q_{v'}$ over the children v' of a node v, with respect to the distribution given by the sampling algorithm Algorithm 6.3, is close to the matrix Q_v. Then we will be able to iterate this lemma to show that the average over the leaves, which are all just

product states weighted by α_v, is close to the matrix at the root which is just $e^{-\beta H}$.

Thus to sample the Gibbs state $e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$, it suffices to sample a leaf from a certain distribution, which we define in Definition 6.11. However, this distribution is not the same as the distribution naturally induced by running Algorithm 6.3, which is defined in Definition 6.9. We will show how to approximate the ratio between these two distributions and then use Theorem 5.8 to sample.

6.1 Basic properties of the sample tree

First we need to prove that various quantities that appear when constructing the sample tree are positive and well-defined. We have the following claim which ensures that the α are real and the matrices M at all of the nodes in the sample tree are indeed of the form $I_{2^{|S|}} + cE$ for $c \in \mathbb{R}$ and $E \in \mathcal{P}$ a $2^{|S|} \times 2^{|S|}$ Hermitian matrix.

Lemma 6.4 (Sample tree has Hermitian nodes)**.** *Assuming that the input $I_{2^{|S|}} + cE$ to Algorithm 6.3 has $c \in \mathbb{R}$ and E is Hermitian with $E \in \mathcal{P}$, then the output also has $c'' \in \mathbb{R}$ and $E'_{[n]\setminus j}$ Hermitian with $E'_{[n]\setminus j} \in \mathcal{P}$.*

Proof. Clearly c'' is real if c is real. Also, if $E \in \mathcal{P}$ and E is Hermitian, then as constructed in the algorithm, $E' \in \mathcal{P}$ and E' is Hermitian. Thus, it is a product of 2×2 Pauli matrices up to a sign of either ± 1. Thus, both E'_j and $E'_{S\setminus j}$ are also just products of Paulis up to a sign of either ± 1 and this means they are both Hermitian. □

Next, we record a few basic observations about the structure of the sample tree.

Fact 6.5. *The sample tree has the following properties. For a node $v = (S, \alpha, B, M)$ at depth d:*

1. $|S| = n - d$
2. *B is a product state over sites in $[n]\setminus S$ sites of the form $\bigotimes_{j\in[n]\setminus S} A_j$ where $A_j \in \{I/2, (I \pm \sigma_x)/2, (I \pm \sigma_y)/2, (I \pm \sigma_z)/2\}$*
3. *M is a $2^{n-d} \times 2^{n-d}$ matrix on sites in S*

Proof. These are immediate from the definition in Algorithm 6.3. □

Lemma 6.6 (Sample tree arity)**.** *In the sample tree T, each node has at most $\mathcal{O}(n^2 4^n/\varepsilon)$ children.*

Proof. This is immediate from the definition in Algorithm 6.3 since there are at most n choices for j, $O(n/\varepsilon)$ choices for c by Lemma 6.7 and at most 4^n possible choices for A and E'. □

In order for the sample tree to be useful, we need to ensure that the weights α are positive and the matrices $I_{2^{|S|}} + cE$ are PSD i.e. $|c| \leqslant 1$. This is nontrivial to show, but we are able to prove it via a carefully chosen potential function. We note that this lemma is crucial to showing that we can keep pinning sites iteratively, and implies the structural result for Gibbs states.

Lemma 6.7 (Coefficients remain small)**.** *Assuming that the input $I_{2^{|S|}} + cE$ to Algorithm 6.3 satisfies either of the following two conditions*

- $|\operatorname{supp}(E)| > 0$ *and* $|c| \leqslant \frac{1}{2}\left(1 - \frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E)|}$
- $c = 0$

Then the output satisfies $|c''| \leqslant \frac{1}{2}\left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E'_{S\setminus j})|}$.

Proof. We start by considering the first condition when $|\operatorname{supp}(E)| > 0$. We consider each of the cases in the algorithm. If $b = 0$ then $E' = E$ then $|\operatorname{supp}(E'_{S\setminus j})| = |\operatorname{supp}(E)| - 1$ and $|c''| = \frac{c}{1-1/\mathfrak{K}}$ so the desired statement follows from the inductive hypothesis. If $b = 1$ then by Lemma 4.5 and the definition in Algorithm 4.6,

$$|c''| = 6\mathfrak{K}|c_1| \leqslant 6\mathfrak{K} \cdot \left(\frac{3}{100\mathfrak{K}}\right)^{\lceil |\operatorname{supp}(E_1)|/\mathfrak{K}\rceil} \leqslant \frac{1}{2}\left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E_1)|} \leqslant \frac{1}{2}\left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E'_{S\setminus j})|}.$$

The argument for the $b = 2$ case and $b = 5$ case are similar. If $b = 3$ then

$$|c''| = 6\mathfrak{K}|c||c_1| \leqslant 6\mathfrak{K} \cdot \left(\frac{3}{100\mathfrak{K}}\right)^{\lceil |\operatorname{supp}(E_1)|/\mathfrak{K}\rceil} |c| \leqslant \left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E_1)|} |c| \leqslant \frac{1}{2}\left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E'_{S\setminus j})|}$$

where the last step used the inductive hypothesis and the fact that $|\operatorname{supp}(E'_{S\setminus j})| \leqslant |\operatorname{supp}(E_1)| + |\operatorname{supp}(E)|$. The argument for the remaining cases $b = 4$ and $b = 6$ are similar.

Next, it remains to consider when $c = 0$. In this case we only need to deal with when $b = 1$ or $b = 2$ (as $c'' = 0$ in all other cases). The argument for these cases is the same as above. □

Corollary 6.8 (Sample tree is well-defined)**.** *Given a Hamiltonian* $H = \sum_a \lambda_a E_a$ *over* n *sites and parameters* $\beta, \mathfrak{K}, \mathfrak{d}$ *such that* H *is a* $\mathfrak{K}$*-local Hamiltonian with degree* $\mathfrak{d}$ *and* $\beta \leqslant 1/(100\mathfrak{d}\mathfrak{K}^2)$*, all nodes in the sample tree, say* (S, α, B, M)*, have*

- M *is a PSD matrix with* $M = I_{2^{|S|}} + cE$ *where* $E \in \mathcal{P}$, $\operatorname{tr}(cE) = 0$ *and* $|c| \leqslant \frac{1}{2}\left(1-\frac{1}{\mathfrak{K}}\right)^{|\operatorname{supp}(E)|}$
- $\alpha > 0$

Proof. First observe that whenever we call Algorithm 6.3 in the tree, either $\operatorname{supp}(E) \neq \emptyset$ or $c = 0$. This holds for the root and for all descendants because of Step 8 in the algorithm. Thus, we can apply Lemma 6.7 and induct first statement. The second statement also follows from Lemma 6.7 now, since we must have $|c''| \leqslant 1/2$ whenever we execute Algorithm 6.3. □

6.2 Weight functions

Now we define two different weighted trees on the sample tree. The first is the weight derived from the sampling process.

Definition 6.9 (Natural weight)**.** Given a Hamiltonian $H = \sum_a \lambda_a E_a$ over n sites and parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that H is a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$, let T be its sample tree. We define the natural weight function ω to have for each node $v \in T$, $\omega(v)$ is the probability of reaching v from the root by running the sampling process in Algorithm 6.3 at each intermediate node.

Remark 6.10. It is clear from the definition that ω indeed defines a valid weighted tree i.e. the weight at any intermediate node is equal to the sum of the weights of the leaves of its subtree.

The second weight function, the true weight, is defined by adjusting the natural weight at each leaf by α_v. We will show (see Corollary 6.14) that to sample from the Gibbs state, it suffices to sample a leaf according to this true weight distribution.

Definition 6.11 (True weight)**.** Given a Hamiltonian $H = \sum_a \lambda_a E_a$ over n sites and parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that H is a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$, let T be its sample tree. We define the true weight function κ as follows: for each leaf node $v = (\emptyset, \alpha, B, 1)$, we set $\kappa(v) = \alpha\omega(v)$ and for each intermediate node v, $\kappa(v)$ is the sum of the weights of the leaves in its subtree.

Remark 6.12. Recall from Corollary 6.8 that all of the α are positive and thus this is a valid weighted tree.

6.3 Analyzing the weight functions

First, we prove the key lemma, which shows that at each node v, the matrix

$$Q_v = \alpha_v B_v \otimes e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}$$

is close to the average of $Q_{v'}$ over its children v' (according to the distribution induced by running Algorithm 6.3 on v). Since these matrices are not trace-normalized, it will be important to ensure that our error bounds are "at the right scale". We do this by bounding our errors multiplicatively in PSD ordering.

Lemma 6.13. *Given a Hamiltonian $H = \sum_a \lambda_a E_a$ over n sites and parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that H is a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$ and $\beta \leqslant 1/(100\mathfrak{d}\mathfrak{K}^2)$, let T be its sample tree. Let v be a node indexed by $(S_v, \alpha_v, B_v, M_v)$. Then we have*

$$\left(1 - \frac{\varepsilon}{10n}\right) \alpha_v B_v \otimes e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}$$
$$\preceq \sum_{v' \text{ child of } v} \frac{\omega(v')}{\omega(v)} \alpha_{v'} B_{v'} \otimes e^{-\beta H^{(S_{v'})}/2} M_{v'} e^{-\beta H^{(S_{v'})}/2} \preceq \left(1 + \frac{\varepsilon}{10n}\right) \alpha_v B_v \otimes e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}$$

where recall $H^{(S_v)}$ is the Hamiltonian restricted to the set S_v.

Proof. Consider the execution of Algorithm 6.3 with input $(S_v, \alpha_v, B_v, M_v)$ where $M_v = I_{2^{|S_v|}} + cE$. We first note that in the execution,

$$\left\| \mathbb{E}[\alpha A \otimes (I_{2^{|S_v|-1}} + c'' E'_{S_v \setminus j})] - \alpha_v \mathbb{E}[I_{2^{|S_v|}} + c' E'] \right\|_{\text{op}} \leqslant \frac{\alpha_v \varepsilon}{100n}, \tag{22}$$

where α, A is the output of Algorithm 6.3 and c', E' are the intermediate parameters appearing in the execution of Algorithm 6.3. This follows from observing that the cases all result in a state that in expectation is $\alpha_v \mathbb{E}[I_{2^{|S_v|}} + c' E']$ and that the only difference between the two sides is due to the rounding of c''. Next, we compute

$$\alpha_v \mathbb{E}[I_{2^{|S_v|}} + c' E']$$
$$= \alpha_v \mathbb{E}\left[\frac{1}{2}\left((I_{2^{|S_v|}} + c_1 E_1^\dagger)(I_{2^{|S_v|}} + cE)(I_{2^{|S_v|}} + c_1 E_1) + (I_{2^{|S_v|}} + c_2 E_2^\dagger)(I_{2^{|S_v|}} + cE)(I_{2^{|S_v|}} + c_1 E_1)\right)\right]$$
$$= \alpha_v T_{t_{\max}, \beta/2}(H^{(S_v)}, H^{(S_v)}_{(j)})^\dagger (I_{2^{|S_v|}} + cE) T_{t_{\max}, \beta/2}(H^{(S_v)}, H^{(S_v)}_{(j)}),$$

where the last step follows from Lemma 4.7. Recall that $M_v = I_{2^{|S_v|}} + cE$. Left and right multiplying the above by $e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2}$ and applying Lemma 4.10, we get

$$\begin{aligned} &\alpha_v e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2} - \alpha_v e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} \mathbb{E}[I_{2^{|S_v|}} + c'E'] e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} \\ &\preceq \frac{\alpha_v \varepsilon}{100n} \cdot e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}. \end{aligned} \tag{23}$$

Next, we can left and right multiply both sides of (22) by $e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2}$ to get

$$\begin{aligned} &e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} \left(\mathbb{E}[\alpha A \otimes (I_{2^{|S_v|-1}} + c''E'_{S_v \setminus j})] - \alpha_v \mathbb{E}[I_{2^{|S_v|}} + c'E'] \right) e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} \\ &\preceq \frac{\alpha_v \varepsilon}{100n} \cdot e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})} \end{aligned} \tag{24}$$

Adding (23) and (24), we get

$$\begin{aligned} &\alpha_v e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2} \\ &\quad - \mathbb{E}\left[\alpha e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} \left(A \otimes (I_{2^{|S_v|-1}} + c''E'_{S_v \setminus j}) \right) e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} \right] \\ &\preceq \frac{\varepsilon}{100n} \cdot \left(\alpha_v e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2} + \alpha_v e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})} \right) \\ &\preceq \frac{\varepsilon}{10n} \cdot \alpha_v e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2} \end{aligned}$$

where in the last step, we applied Lemma 4.9 and used that $0.5 I_{2^{|S_v|}} \preceq M_v$ from Corollary 6.8. Similarly, we can prove a lower bound of

$$-\frac{\varepsilon}{10n} \cdot \alpha_v e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}.$$

Now we immediately get the desired inequality since $H^{(S_v)} - H^{(S_v)}_{(j)} = I \otimes H^{(S_v \setminus j)}$ (where the identity matrix is on the jth site) and thus

$$e^{-\beta(H^{(S_v)} - H^{(S_v)}_{(j)})/2} = I \otimes e^{-\beta H^{(S_v \setminus j)}/2}.$$

Also recall that averaging over the children of v with weights $\omega(v')/\omega(v)$ is exactly the same as taking the expectation over the execution of Algorithm 6.3. □

By iterating Lemma 6.13, we can relate the average at the leaves to the Gibbs state. Specifically, since the true weight distribution on the leaves is exactly defined to be equal to the natural weight distribution arising from the sampling process and then distorted by a factor of α_v at each leaf v, we get that the average of the product states B_v at the leaves according to the true weight distribution is close to the Gibbs state $e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$.

Corollary 6.14 (Average of the leaves is close to the Gibbs state). *Given a Hamiltonian $H = \sum_a \lambda_a E_a$ over n sites and parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that H is a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$ and $\beta \leqslant 1/(100\mathfrak{d}\mathfrak{K}^2)$, let T be its sample tree . Then*

$$\left\| \frac{e^{-\beta H}}{\operatorname{tr}(e^{-\beta H})} - \frac{\sum_{v' \text{ leaf of } T} \kappa(v') B_{v'}}{\sum_{v' \text{ leaf of } T} \kappa(v')} \right\|_1 \leqslant \frac{\varepsilon}{2}.$$

Proof. By repeatedly applying Lemma 6.13 starting from the root, we have

$$\left(1-\frac{\varepsilon}{10}\right)e^{-\beta H} \preceq \sum_{v' \text{ leaf of } T} \kappa(v')B_{v'} \preceq \left(1+\frac{\varepsilon}{10}\right)e^{-\beta H},$$

which follows from recalling that $\kappa(v) = \alpha(v)\omega(w)$. Thus

$$\left\| \frac{e^{-\beta H}}{\operatorname{tr}(e^{-\beta H})} - \frac{\sum_{v' \text{ leaf of } T} \kappa(v')B_{v'}}{\sum_{v' \text{ leaf of } T} \kappa(v')} \right\|_1 \leqslant 2 \left\| \frac{e^{-\beta H}}{\operatorname{tr}(e^{-\beta H})} - \frac{\sum_{v' \text{ leaf of } T} \kappa(v')B_{v'}}{\operatorname{tr}(e^{-\beta H})} \right\|_1 \leqslant \frac{\varepsilon}{2}$$

as desired. □

To sample from the distribution induced by κ on the leaves, we will need to approximate the ratio $\kappa(v)/\omega(v)$ in order to then apply Theorem 5.8 (since we can run Algorithm 6.3 to sample from the weighted tree ω). We do this below.

Corollary 6.15 (Bounded weight ratio)**.** *Given a Hamiltonian $H = \sum_a \lambda_a E_a$ over n sites and parameters $\beta, \mathfrak{K}, \mathfrak{d}$ such that H is a $\mathfrak{K}$-local Hamiltonian with degree $\mathfrak{d}$ and $\beta \leqslant 1/(100\mathfrak{d}\mathfrak{K}^2)$, let T be its sample tree. Let v be a node indexed by $(S_v, \alpha_v, B_v, M_v)$. Then we have*

$$0.4\alpha_v \operatorname{tr}(e^{-\beta H^{(S_v)}}) \leqslant \frac{\kappa(v)}{\omega(v)} \leqslant 1.6\alpha_v \operatorname{tr}(e^{-\beta H^{(S_v)}}).$$

Proof. By repeatedly applying Lemma 6.13, we have

$$\begin{aligned}\left(1-\frac{\varepsilon}{10n}\right)^{|S_v|} \alpha_v B_v \otimes e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2} &\preceq \sum_{\substack{v' \text{ leaf} \\ v' \preceq v}} \frac{\omega(v')}{\omega(v)} \alpha_{v'} B_{v'} \\ &\preceq \left(1+\frac{\varepsilon}{10n}\right)^{|S_v|} \alpha_v B_v \otimes e^{-\beta H^{(S_v)}/2} M_v e^{-\beta H^{(S_v)}/2}\end{aligned}.$$

Now take the trace of the above. Recall that by Corollary 6.8, $0.5I \preceq M_v \preceq 1.5I$ and $\operatorname{tr}(B_{v'}) = 1$ for all leaves v', so we have

$$0.4\alpha_v \operatorname{tr}(e^{-\beta H^{(S_v)}}) \leqslant \frac{\kappa(v)}{\omega(v)} \leqslant 1.6\alpha_v \operatorname{tr}(e^{-\beta H^{(S_v)}})$$

which is exactly what we set out to prove. □

We will also need a bound on the running time of Algorithm 6.3.

Lemma 6.16 (Running time)**.** *Algorithm 6.3 can be implemented to run in time $O(\log(n/\varepsilon)\operatorname{poly}(\mathfrak{K}, \mathfrak{d}))$.*

Proof. By Lemma 4.5, Algorithm 4.6 runs in $\log(n/\varepsilon)\operatorname{poly}(\mathfrak{K}, \mathfrak{d})$ time. The remaining operations can be implemented in time $O(|\operatorname{supp}(E_1)| + |\operatorname{supp}(E_2)|)$ since we can multiply the Pauli matrices qubit-wise. This is $O(\log(n/\varepsilon)\mathfrak{K})$ by the guarantees of Lemma 4.5. Thus, the total runtime is $\log(n/\varepsilon)\operatorname{poly}(\mathfrak{K}, \mathfrak{d})$ as claimed. □

Now we can put everything together to prove our our main theorem, Theorem 1.6.

Proof of Theorem 1.6. We will apply Theorem 5.8 on the sample-tree with $w' \leftarrow \omega$ and $w \leftarrow \kappa$. First, we verify the hypotheses of Theorem 5.8. By Lemma 6.6, the number of children of each node is at most $k \leqslant n^2 4^n/\varepsilon$. Consider two adjacent nodes $u = (S_u, \alpha_u, B_u, M_u)$ and

$v = (S_v, \alpha_v, B_v, M_v)$ where u is the parent of v. Note this means $S_v = S_u \backslash j$ for some element $j \in S_u$. By Corollary 6.15,

$$\frac{1}{4} \cdot \frac{\alpha_u \operatorname{tr}(e^{-\beta H^{(S_u)}})}{\alpha_v \operatorname{tr}(e^{-\beta H^{(S_v)}})} \leqslant \frac{w_u w_v'}{w_v w_u'} \leqslant 4 \cdot \frac{\alpha_u \operatorname{tr}(e^{-\beta H^{(S_u)}})}{\alpha_v \operatorname{tr}(e^{-\beta H^{(S_v)}})}.$$

Note that $|c''| \leqslant 1/2$ by Corollary 6.8 so

$$\frac{1}{4} \leqslant \frac{\alpha_u}{\alpha_v} \leqslant \frac{3}{4}.$$

Also recall that

$$e^{-\beta(H^{(S_u)} - H^{(S_u)}_{(j)})} = I \otimes e^{-\beta H^{(S_u \backslash j)}} = I \otimes e^{-\beta H^{(S_v)}}$$

so by Lemma 4.9,

$$1.9 \leqslant \frac{\operatorname{tr}(e^{-\beta H^{(S_u)}})}{\operatorname{tr}(e^{-\beta H^{(S_v)}})} \leqslant 2.1.$$

Thus, we deduce

$$0.1 \leqslant \frac{w_u w_v'}{w_v w_u'} \leqslant 10.$$

Next, we show how to implement the types of queries required in Theorem 5.8. For the first type of query, we can apply Corollary 6.15 and Theorem 3.5 with $\eta \leftarrow 0.01$. The runtime of answering this query is

$$O\left(n \cdot (100n)^{\frac{4+\log(\mathfrak{d})}{\log(\beta_c/\beta)}} \cdot \mathfrak{K} \cdot \operatorname{polylog}(n)\right)$$

For the third type of queries, we simply run Algorithm 6.3. By Lemma 6.16, the runtime is $O(\log(n/\varepsilon)\operatorname{poly}(\mathfrak{K}, \mathfrak{d}))$. Note that when running the Markov chain in Theorem 5.8, we can store the of all of the labels (S, α, B, M) of all of the nodes that we have visited so far. When visiting a new node, computing (S, α, B, M) involves just one execution of Algorithm 6.3 on its parent, which we must have already visited. Thus, whenever we visit a leaf we already have an exact value for $\alpha = \kappa(v)/\omega(v)$ so this allows us to answer the second type of query whenever we need to (which is only when the Markov chain visits a leaf). Thus, we have verified all of the hypotheses of Theorem 5.8. Putting everything together, we get that with probability $1 - \delta$, we get a sample from a distribution that is $\varepsilon/2$-close in TV to the leaf distribution of κ in time

$$\widetilde{\mathcal{O}}\left(n^{6+\frac{\cdot(4+\log(\mathfrak{d}))}{\log(\beta_c/\beta)}} \log^2(1/\varepsilon) \log(1/\delta) \operatorname{poly}(\mathfrak{K}, \mathfrak{d})\right).$$

Now let this sample be indexed by a product state $B_v = A_1 \otimes \cdots \otimes A_n$. By Corollary 6.8, this is a product state over $\{I/2, (I \pm \sigma_x)/2, (I \pm \sigma_y)/2, (I \pm \sigma_z)/2\}$. Finally

$$\begin{aligned}\left\| \frac{e^{-\beta H}}{\operatorname{tr}(e^{-\beta H})} - \mathbb{E}[B_v] \right\|_1 &\leqslant \left\| \frac{e^{-\beta H}}{\operatorname{tr}(e^{-\beta H})} - \frac{\sum_{v' \text{ leaf of } T} \kappa(v') B_{v'}}{\sum_{v' \text{ leaf of } T} \kappa(v')} \right\|_1 + \left\| \frac{\sum_{v' \text{ leaf of } T} \kappa(v') B_{v'}}{\sum_{v' \text{ leaf of } T} \kappa(v')} - \mathbb{E}[B_v] \right\| \\ &\leqslant \varepsilon.\end{aligned}$$

where the expectation is over the execution of the Markov chain in Theorem 5.8 (conditioned on a valid output). Note that in the last step above we used Corollary 6.14 and the fact that the distribution of our output is $\varepsilon/2$ close to the distribution induced by $\kappa(\cdot)$ on the leaves. Thus, we can simply output $\widehat{\rho} = B_v$ and this completes the proof. □

Finally we prove the structural property Theorem 1.5. It follows directly from Corollary 6.14.

Proof of Theorem 1.5. Note that the set of all convex combinations of $Q_1 \otimes Q_2 \otimes \ldots \otimes Q_n$ where

$$Q_i \in \left\{ (I+\sigma_x)/2, (I-\sigma_x)/2, (I+\sigma_y)/2, (I-\sigma_y)/2, (I+\sigma_z)/2, (I-\sigma_z)/2 \right\},$$

for each $i \in [n]$, forms a closed set. By Corollary 6.14, for any $\varepsilon > 0$, the Gibbs state $e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$ is ε-close to this set (in trace distance). Taking the limit as $\varepsilon \to 0$, we get that $e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$ must actually be in this set of convex combinations, as desired. □

Acknowledgments

The authors thank Nikhil Srivastava, Álvaro Alhambra, Cambyse Rouzé, Daniel Stilck França, and Chi-Fang Chen for enlightening discussions. In particular, we thank Álvaro for highlighting the structural implications of our work and for drawing a connection to ESD.

AB is supported by Ankur Moitra's ONR grant and the NSF TRIPODS program (award DMS-2022448). AL is supported in part by an NSF GRFP and a Hertz Fellowship. AM is supported in part by a Microsoft Trustworthy AI Grant, an ONR grant and a David and Lucile Packard Fellowship. ET is supported by the Miller Institute for Basic Research in Science, University of California Berkeley. This work was done in part while the authors were at the Simons Institute for the Theory of Computing.

References

[AAG22] Anurag Anshu, Itai Arad, and David Gosset. "An area law for 2d frustration-free spin systems". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC '22. ACM, June 2022. DOI: 10.1145/3519935.3519962. arXiv: 2103.02492 [quant-ph] (page 3).

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. "Sample-efficient learning of interacting quantum systems". In: *Nature Physics* (May 2021). Preliminary version in FOCS 2020. DOI: 10.1038/s41567-021-01232-0. arXiv: 2004.07266 [quant-ph] (page 1).

[ABV01] M. C. Arnesen, S. Bose, and V. Vedral. "Natural thermal and magnetic entanglement in the 1D Heisenberg model". In: *Physical Review Letters* 87.1 (June 2001), p. 017901. ISSN: 1079-7114. DOI: 10.1103/physrevlett.87.017901. arXiv: quant-ph/0009060 (page 4).

[AJ08] Asma Al-Qasimi and Daniel F. V. James. "Sudden death of entanglement at finite temperature". In: *Physical Review A* 77.1 (Jan. 2008), p. 012117. ISSN: 1094-1622. DOI: 10.1103/physreva.77.012117 (page 4).

[AKLV13] Itai Arad, Alexei Kitaev, Zeph Landau, and Umesh Vazirani. "An area law and sub-exponential algorithm for 1d systems". 2013. DOI: 10.48550/ARXIV.1301.1162. arXiv: 1301.1162 [quant-ph] (page 3).

[ALG21] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral independence in high-dimensional expanders and applications to the hardcore model". In: *SIAM Journal on Computing* (July 2021). ISSN: 1095-7111. DOI: 10.1137/20m1367696. arXiv: 2001.00303 [cs.DS] (page 4).

[Alh23] Álvaro M. Alhambra. "Quantum many-body systems in thermal equilibrium". In: *PRX Quantum* 4.4 (Nov. 2023), p. 040201. ISSN: 2691-3399. DOI: 10.1103/prxquantum.4.040201. arXiv: 2204.08349 [quant-ph] (page 1).

[Alm+07] M. P. Almeida, F. de Melo, M. Hor-Meyll, A. Salles, S. P. Walborn, P. H. Souto Ribeiro, and L. Davidovich. "Environment-induced sudden death of entanglement". In: *Science* 316.5824 (Apr. 2007), pp. 579–582. ISSN: 1095-9203. DOI: 10.1126/science.1139892. arXiv: quant-ph/0701184 (page 4).

[AMD15] Leandro Aolita, Fernando de Melo, and Luiz Davidovich. "Open-system dynamics of entanglement: A key issues review". In: *Reports on Progress in Physics* 78.4 (Mar. 2015), p. 042001. ISSN: 1361-6633. DOI: 10.1088/0034-4885/78/4/042001. arXiv: 1402.3713 [quant-ph] (page 4).

[Bar16] Alexander Barvinok. *Combinatorics and complexity of partition functions*. Springer International Publishing, 2016. ISBN: 9783319518299. DOI: 10.1007/978-3-319-51829-9 (page 4).

[BCGLPR23] Ivan Bardet, Ángela Capel, Li Gao, Angelo Lucia, David Pérez-García, and Cambyse Rouzé. "Rapid thermalization of spin chain commuting Hamiltonians". In: *Physical Review Letters* 130.6 (Feb. 2023), p. 060401. ISSN: 1079-7114. DOI: 10.1103/physrevlett.130.060401. arXiv: 2112.00593 [quant-ph] (page 1).

[BD97] R. Bubley and M. Dyer. "Path coupling: A technique for proving rapid mixing in Markov chains". In: *Proceedings 38th Annual Symposium on Foundations of Computer Science*. SFCS-97. IEEE Comput. Soc, 1997. DOI: 10.1109/sfcs.1997.646111 (page 4).

[BK18] Fernando G. S. L. Brandão and Michael J. Kastoryano. "Finite correlation length implies efficient preparation of quantum thermal states". In: *Communications in Mathematical Physics* 365.1 (May 2018), pp. 1–16. ISSN: 1432-0916. DOI: 10.1007/s00220-018-3150-8. arXiv: 1609.07877 [quant-ph] (pages 1, 4).

[BLMT23] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. "Learning quantum hamiltonians at any temperature in polynomial time". Oct. 3, 2023. arXiv: 2310.02243 [quant-ph] (page 1).

[CB21] Chi-Fang Chen and Fernando G. S. L. Brandão. "Fast thermalization from the eigenstate thermalization hypothesis". Dec. 14, 2021. DOI: 10.48550/ARXIV.2112.07646. arXiv: 2112.07646 [quant-ph] (page 1).

[CE22] Yuansi Chen and Ronen Eldan. "Localization schemes: A framework for proving mixing bounds for Markov chains". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, Oct. 2022. DOI: 10.1109/focs54457.2022.00018. arXiv: 2203.04163 [math.PR] (page 4).

[CKBG23] Chi-Fang Chen, Michael J. Kastoryano, Fernando G. S. L. Brandão, and András Gilyén. "Quantum thermal state preparation". Mar. 31, 2023. DOI: 10.48550/ARXIV.2303.18224. arXiv: 2303.18224 [quant-ph] (page 1).

[CKG23] Chi-Fang Chen, Michael J. Kastoryano, and András Gilyén. "An efficient and exact noncommutative quantum Gibbs sampler". Nov. 15, 2023. arXiv: 2311.09207 [quant-ph] (page 4).

[FMB05] B. V. Fine, F. Mintert, and A. Buchleitner. "Equilibrium entanglement vanishes at finite temperature". In: *Physical Review B* 71.15 (Apr. 2005), p. 153105. ISSN: 1550-235X. DOI: 10.1103/physrevb.71.153105. arXiv: cond-mat/0505739 (page 4).

[GSLW19] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. "Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics". In: *Proceedings of the 51st ACM Symposium on the Theory of Computing (STOC)*. ACM, June 2019, pp. 193–204. DOI: 10.1145/3313276.3316366. arXiv: 1806.01838 (page 1).

[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models". In: *Combinatorics, Probability and Computing* 25.4 (Feb. 2016), pp. 500–559. ISSN: 1469-2163. DOI: 10.1017/s0963548315000401. arXiv: 1203.2226 [cs.DM] (pages 2, 4).

[Has07] M B Hastings. "An area law for one-dimensional quantum systems". In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.08 (Aug. 2007), P08024–P08024. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2007/08/p08024. arXiv: 0705.2024 [quant-ph] (page 3).

[HC18] O. Hart and C. Castelnovo. "Entanglement negativity and sudden death in the toric code at finite temperature". In: *Physical Review B* 97.14 (Apr. 2018), p. 144410. ISSN: 2469-9969. DOI: 10.1103/physrevb.97.144410. arXiv: 1710.11139 [cond-mat.str-el] (page 4).

[HHH98] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. "Mixed-state entanglement and distillation: is there a "bound" entanglement in nature?" In: *Physical Review Letters* 80.24 (June 1998), pp. 5239–5242. ISSN: 1079-7114. DOI: 10.1103/physrevlett.80.5239. arXiv: quant-ph/9801069 (page 4).

[HHHH09] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. "Quantum entanglement". In: *Reviews of Modern Physics* 81.2 (June 2009), pp. 865–942. ISSN: 1539-0756. DOI: 10.1103/revmodphys.81.865. arXiv: quant-ph/0702225 (page 4).

[HKT22] Jeongwan Haah, Robin Kothari, and Ewin Tang. "Optimal learning of quantum Hamiltonians from high-temperature Gibbs states". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, Oct. 2022. DOI: 10.1109/focs54457.2022.00020. arXiv: 2108.04842 [quant-ph] (pages 1, 4, 5, 9–11).

[HMS20] Aram W. Harrow, Saeed Mehraban, and Mehdi Soleimanifar. "Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. STOC '20. ACM, June 2020. DOI: 10.1145/3357713.3384322. arXiv: 1910.09071 [quant-ph] (pages 1, 4).

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *Journal of the ACM* 51.4 (July 2004), pp. 671–697. ISSN: 1557-735X. DOI: 10.1145/1008731.1008738 (page 4).

[KAA21] Tomotaka Kuwahara, Álvaro M. Alhambra, and Anurag Anshu. "Improved thermal area law and quasilinear time algorithm for quantum Gibbs states". In: *Physical Review X* 11.1 (Mar. 2021), p. 011047. DOI: 10.1103/physrevx.11.011047. arXiv: 2007.11174 [quant-ph] (pages 1, 3).

[KB16] Michael J. Kastoryano and Fernando G. S. L. Brandão. "Quantum Gibbs samplers: The commuting case". In: *Communications in Mathematical Physics* 344.3 (May 2016), pp. 915–957. ISSN: 1432-0916. DOI: 10.1007/s00220-016-2641-8. arXiv: 1409.3435 [quant-ph] (page 1).

[KB19] Kohtaro Kato and Fernando G. S. L. Brandão. "Quantum approximate Markov chains are thermal". In: *Communications in Mathematical Physics* 370.1 (June 2019), pp. 117–149. DOI: 10.1007/s00220-019-03485-6 (page 1).

[KGKRE14] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert. "Locality of temperature". In: *Physical Review X* 4.3 (July 2014), p. 031019. ISSN: 2160-3308. DOI: 10.1103/physrevx.4.031019. arXiv: 1309.0816 [quant-ph] (pages 1, 4).

[KP86] R. Kotecký and D. Preiss. "Cluster expansion for abstract polymer models". In: *Communications in Mathematical Physics* 103.3 (Sept. 1986), pp. 491–498. ISSN: 1432-0916. DOI: 10.1007/bf01211762 (page 4).

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. Second. American Mathematical Society, Providence, RI, 2017, pp. xvi+447. ISBN: 978-1-4704-2962-1. DOI: 10.1090/mbk/107 (page 21).

[MH21] Ryan L. Mann and Tyler Helmuth. "Efficient algorithms for approximating quantum partition functions". In: *Journal of Mathematical Physics* 62.2 (Feb. 2021). ISSN: 1089-7658. DOI: 10.1063/5.0013689. arXiv: 2004.11568 [quant-ph] (page 5).

[RFA24] Cambyse Rouzé, Daniel Stilck França, and Álvaro M. Alhambra. "Efficient thermalization and universal quantum computing with quantum Gibbs samplers". Mar. 19, 2024. arXiv: 2403.12691 [quant-ph] (pages 1, 4).

[SDHS16] Nicholas E. Sherman, Trithep Devakul, Matthew B. Hastings, and Rajiv R. P. Singh. "Nonzero-temperature entanglement negativity of quantum spin models: area law, linked cluster expansions, and sudden death". In: *Physical Review E* 93.2 (Feb. 2016), p. 022128. ISSN: 2470-0053. DOI: 10.1103/physreve.93.022128. arXiv: 1510.08005 [cond-mat.stat-mech] (page 4).

[SJ89] Alistair Sinclair and Mark Jerrum. "Approximate counting, uniform generation and rapidly mixing Markov chains". In: *Information and Computation* 82.1 (July 1989), pp. 93–133. ISSN: 0890-5401. DOI: 10.1016/0890-5401(89)90067-9 (pages 9, 19).

[Sly10] Allan Sly. "Computational transition at the uniqueness threshold". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, Oct. 2010. DOI: 10.1109/focs.2010.34. arXiv: 1005.5584 [cs.CC] (pages 2, 4).

[SS14] Allan Sly and Nike Sun. "Counting in two-spin models on d-regular graphs". In: *The Annals of Probability* 42.6 (Nov. 2014). ISSN: 0091-1798. DOI: 10.1214/13-aop888. arXiv: 1203.2602 [math.PR] (pages 2–4).

[SST14] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. "Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs". In: *Journal of Statistical Physics* 155.4 (Mar. 2014), pp. 666–686. ISSN: 1572-9613. DOI: 10.1007/s10955-014-0947-5. arXiv: 1107.2368 [cs.DM] (page 3).

[TOVPV11] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete. "Quantum Metropolis sampling". In: *Nature* 471.7336 (Mar. 2011), pp. 87–90. ISSN: 1476-4687. DOI: 10.1038/nature09770. arXiv: 0911.3635 [quant-ph] (page 1).

[VW02] G. Vidal and R. F. Werner. "Computable measure of entanglement". In: *Physical Review A* 65.3 (Feb. 2002), p. 032314. ISSN: 1094-1622. DOI: 10.1103/physreva.65.032314. arXiv: quant-ph/0102117 (page 4).

[Wei06] Dror Weitz. "Counting independent sets up to the tree threshold". In: *Proceedings of the thirty-eighth annual ACM symposium on Theory of Computing*. STOC06. ACM, May 2006. DOI: 10.1145/1132516.1132538 (page 4).

[WVHC08] Michael M. Wolf, Frank Verstraete, Matthew B. Hastings, and J. Ignacio Cirac. "Area laws in quantum systems: mutual information and correlations". In: *Physical Review Letters* 100.7 (Feb. 2008), p. 070502. ISSN: 1079-7114. DOI: 10.1103/physrevlett.100.070502. arXiv: 0704.3906 [quant-ph] (pages 1, 3).

[YE04] Ting Yu and J. H. Eberly. "Finite-time disentanglement via spontaneous emission". In: *Physical Review Letters* 93.14 (Sept. 2004), p. 140404. ISSN: 1079-7114. DOI: 10.1103/physrevlett.93.140404. arXiv: quant-ph/0404161 (page 4).

[YE09] Ting Yu and J. H. Eberly. "Sudden death of entanglement". In: *Science* 323.5914 (Jan. 2009), pp. 598–601. ISSN: 1095-9203. DOI: 10.1126/science.1167343. arXiv: 0910.1396 [quant-ph] (page 4).

[YL23] Chao Yin and Andrew Lucas. "Polynomial-time classical sampling of high-temperature quantum Gibbs states". May 29, 2023. DOI: 10.48550/ARXIV.2305.18514. arXiv: 2305.18514 [quant-ph] (page 5).