

Rapid mixing of path integral Monte Carlo for 1D stoquastic Hamiltonians

Elizabeth Crosson¹ and Aram W. Harrow²

¹Center for Quantum Information and Control, University of New Mexico.

²Center for Theoretical Physics, Massachusetts Institute of Technology.

Path integral quantum Monte Carlo (PIMC) is a method for estimating thermal equilibrium properties of stoquastic quantum spin systems by sampling from a classical Gibbs distribution using Markov chain Monte Carlo. The PIMC method has been widely used to study the physics of materials and for simulated quantum annealing, but these successful applications are rarely accompanied by formal proofs that the Markov chains underlying PIMC rapidly converge to the desired equilibrium distribution.

In this work we analyze the mixing time of PIMC for 1D stoquastic Hamiltonians, including disordered transverse Ising models (TIM) with long-range algebraically decaying interactions as well as disordered XY spin chains with nearest-neighbor interactions. By bounding the convergence time to the equilibrium distribution we rigorously justify the use of PIMC to approximate partition functions and expectations of observables for these models at inverse temperatures that scale at most logarithmically with the number of qubits.

The mixing time analysis is based on the canonical paths method applied to the single-site Metropolis Markov chain for the Gibbs distribution of 2D classical spin models with couplings related to the interactions in the quantum Hamiltonian. Since the system has strongly nonisotropic couplings that grow with system size, it does not fall into the known cases where 2D classical spin models are known to mix rapidly.

1 Introduction

The application of Markov chain Monte Carlo methods to thermal states of a limited class of quantum systems was first proposed by Suzuki, Miyashita and Kuroda 40 years ago [43]. Their method applied to Hamiltonians with real and nonpositive off-diagonal matrix elements in a particular basis, a property that was called “being free of the sign problem” for many years until the term “stoquastic” was eventually adopted [12] to emphasize a connection with nonnegative matrices. Since then there have been a growing number of works on rigorous classical simulations for various stoquastic systems [8, 9, 11, 15]. Many quantum systems of physical and computational interest are stoquastic, including spinless particles moving on arbitrary graphs with position dependent interactions, as well as generalized transverse Ising models, which are particularly notable for their use in quantum annealing [1, 19, 31].

Elizabeth Crosson: crosson@unm.edu

Aram W. Harrow : aram@mit.edu

Statement of results. A general n -qubit 1D stoquastic Hamiltonian with nearest-neighbor interactions has the form,

$$\mathcal{H} = \sum_{j=1}^n H_{j,j+1}, \quad (1)$$

where each $H_{j,j+1}$ acts non-trivially on (at most) the sites $j, j+1$ (with $n+1$ identified with 1) and each $\|H_{j,j+1}\| \leq 1$. Define the “computational basis” $\{|1\rangle, |-1\rangle\}$ for each qubit to be the eigenstates of the σ^z operator. We assume $\mathcal{H}$ is stoquastic in the computational basis [11, 12], which means that in this basis each $H_{j,j+1}$ is a real symmetric matrix with nonpositive off-diagonal entries,

$$\langle z | H_{j,j+1} | z' \rangle \leq 0 \text{ for all } z, z' \in \{-1, 1\}^n \text{ with } z \neq z'. \quad (2)$$

For a given inverse temperature β we consider the quantum partition function $\mathcal{Z}_\beta$ and the quantum Gibbs state ρ_β ,

$$\rho_\beta := \frac{e^{-\beta\mathcal{H}}}{\mathcal{Z}_\beta}, \quad \mathcal{Z}_\beta := \text{tr}(e^{-\beta\mathcal{H}}), \quad (3)$$

and also the expectation value of observables $\mathcal{Q}$ in the thermal state, $\langle \mathcal{Q} \rangle_\beta := \text{tr}[\mathcal{Q}\rho_\beta]$. For the special class of systems called 1D generalized transverse Ising models (TIM) we go beyond (1) and allow systems with long-range interactions that decay sufficiently quickly (faster than inverse square) with distance,

$$\mathcal{H} = - \sum_{j=1}^n \Gamma_j \sigma_j^x + \sum_{j=1}^n K_j^z \sigma_j^z + \sum_{j,k=1}^n K_{jk}^{zz} \sigma_j^z \sigma_k^z, \quad \Gamma := \min_{j=1,\dots,n} \Gamma_j > 0, \quad (4)$$

$$K_j^z \in [-1, 1], \quad \text{and } |K_{jk}^{zz}| \leq |i-j|^{-(2+\xi)} \text{ for } \xi > 0. \quad (5)$$

The form (5) simplifies both the presentation of the PIMC method and our proof that the mixing time of PIMC for these models is at most $\text{poly}(n, e^\beta, \Gamma^{-1})$. We also show the mixing time of PIMC is at most $\text{poly}(n, e^\beta, \Gamma^{-1})$ for generalized TIM with $\sigma^x \sigma^x$ and $\sigma^y \sigma^y$ interactions,

$$H = - \sum_{j=1}^n \Gamma_j \sigma_j^x - \sum_{j=1}^n K_j^{xx} \sigma_j^x \sigma_{j+1}^x - \sum_{j=1}^n K_j^{yy} \sigma_j^y \sigma_{j+1}^y + \sum_{j=1}^n \left(K_j^{zz} \sigma_j^z \sigma_{j+1}^z + K_j^z \sigma_j^z \right), \quad (6)$$

$$K_j^{zz}, K_j^z \in [-1, 1], \quad K_j^{xx} \geq 0, \quad K_j^{yy} \in [-K_j^{xx}, K_j^{xx}] \text{ and } \Gamma := \min_{j=1,\dots,n} \Gamma_j > 0.$$

The mixing time bounds can be used to construct an FPRAS (fully polynomial-time randomized approximation scheme) for the partition function of models of the form (5) and (6) and inverse temperatures $\beta = \mathcal{O}(\log n)$. In addition to the partition function there is a similar FPRAS for the expectation value of observables that are sparse and row-computable (in the computational basis) in the thermal state ρ_β for $\beta = \mathcal{O}(\log n)$. For general 1D stoquastic Hamiltonians (1) we prove that the PIMC mixing time is at most quasi-polynomial in the system size, which yields a quasi-polynomial time approximation algorithm for the partition function and for the approximation of observables.

Theorem 1. *There is an algorithm which takes as input an inverse temperature β and an n -qubit 1D stoquastic Hamiltonian H and outputs an estimate $\tilde{\mathcal{Z}}_\beta$ of the quantum partition function $\mathcal{Z}_\beta$.*

If also given an observable $\mathcal{Q} \succeq 0$ that is sparse and row-computable in the computational basis, the algorithm outputs an estimate $\tilde{\mathcal{Q}}$ of its expectation value in the state ρ_β .

These estimates satisfy

$$|\tilde{\mathcal{Z}}_\beta - \mathcal{Z}_\beta| \leq \delta_{mult} \mathcal{Z}_\beta + \delta_{add} \quad (7)$$

$$|\tilde{\mathcal{Q}} - \text{tr}[\mathcal{Q}\rho_\beta]| \leq \delta_{mult} \text{tr}[\mathcal{Q}\rho_\beta] + \delta_{add}. \quad (8)$$

with probability $\geq 1 - \delta_{fail}$. The algorithm runs in time at most

$$\text{poly}(n, e^\beta, \Gamma^{-1}, \delta_{mult}^{-1}, \log(\delta_{add}^{-1}), \log(\delta_{fail}^{-1}))$$

for models of the form (5) and (6), and in time at most

$$\text{poly}(2^{\beta \left(\log \frac{n}{\delta_{mult}}\right)^2}, \log(1/\delta_{add}), \log(1/\delta_{fail}))$$

for models of the form (1).

Relation to previous work. A variety of quantum Monte Carlo methods have been devised for estimating equilibrium properties of stoquastic Hamiltonians, which can be broadly divided into diffusion methods [21, 27, 40] and path integral methods [2, 36, 39, 42, 43]. Diffusion Monte Carlo methods use Hamiltonian matrix elements to define transition rates for a substochastic random process, along with birth and death population dynamics or importance sampling to control statistical fluctuations. Examples of proven efficient convergence for diffusion Monte Carlo methods include simulating frustration-free stoquastic adiabatic computation [11] and simulating the “go with the winners” method [26]. Disadvantages of diffusion methods include quantum amplitude vs quantum probability obstructions for some variants [13, 27], as well as the need for importance sampling with trial wave functions that may in general not be available [8].

PIMC methods, in contrast, are based on Markov chains defined on an enlarged state space consisting of many coupled copies of a high-temperature classical system along a so-called “imaginary-time” direction [43]. PIMC has been used to give an FPRAS for the ferromagnetic case of TIM [8] on general interaction graphs by a reduction to a classic result on sampling the Gibbs distribution of classical ferromagnetic Ising models [28]. Also, the Glauber dynamics for an infinite dimensional variant of PIMC has been shown to have optimal temporal mixing for ferromagnetic TIM at high temperature [14] and for ferromagnetic TIM on regular trees [35]. Here “ferromagnetic” means of the form in (5) but with each $K_j^z = 0$ and each $K_j^{zz} \leq 0$, i.e. lower energy is achieved by aligning adjacent spins. Since such interactions cannot be frustrated, these models are considered easier to analyze. Disadvantages of PIMC methods include the significant overhead created by adding the imaginary-time direction, and also the possibility of slow mixing due to topological obstructions [3, 23]. The general presence of obstructions to rapid mixing for the standard versions of both diffusion and path integral methods also motivates the continuing development of new algorithms for simulating stoquastic Hamiltonians [2, 9].

To our knowledge, the simulations in this work are the first provably efficient Monte Carlo methods for a large class of generalized TIM systems, which differ from ferromagnetic TIM by allowing couplings of both signs between neighboring spins, as well as local fields of both signs. A generalized TIM in 2D or higher dimensions can encode NP-complete constraint satisfaction problems, which is the basis for the use of these models in quantum annealing [1, 19, 31]. Besides their relative simplicity from the standpoint of experimental implementation, generalized TIM on interaction graphs of degree 3 are also known to be universal for stoquastic quantum annealing [10,

17]. The application of Monte Carlo methods to stoquastic quantum annealing is called simulated quantum annealing, and a significant open question is whether some variant of simulated quantum annealing can efficiently simulate stoquastic quantum annealing. This open question has motivated several recent proposals for non-stoquastic quantum annealing [16, 25, 41, 46]. Some recent progress has been made by analyzing specific models for which both quantum annealing and simulated quantum annealing can be exponentially faster than classical simulated annealing [6, 7, 15, 20, 30, 32, 37, 38].

Another class of simulation methods for 1D quantum systems are those based on matrix product states [45]. These methods do not require the Hamiltonian to be stoquastic and have been applied in rigorous classical simulations of systems with limited entanglement [4, 22, 33]. Their runtime scales exponentially with spectral gap, while our runtime bound for PIMC scales exponentially with the inverse temperature, which is qualitatively similar. These kinds of simulations are not expected to obsolete PIMC in general, however, as the latter is used in practice to simulate highly entangled stoquastic systems. Furthermore, the version of PIMC that we consider is closely related to the practical implementations of PIMC, and so this work can be seen as a rigorous analysis of a heuristic approximation algorithm that was originally developed and applied with great success by the physics community.

Organization of the remaining sections. Section 2 reviews the PIMC method and sketches the proof of our mixing time analysis. Section 3.1 defines the Suzuki-Trotter approximation to the partition function which is used to map 1D stoquastic Hamiltonians onto 2D classical spin systems. Section 3.2 and Section 3.3 describe the use of samples from the Gibbs distribution of the 2D classical system to approximate quantum observables and the quantum partition function. Section 3.4 proves a concentration theorem for PIMC that justifies the analysis of a restricted state space in later sections. Section 4.1 reviews the canonical paths method that we use to analyze the PIMC mixing time. This method is then applied to generalized TIM in Section 4.3, generalized TIM with XX interactions in Section 4.4, and general 1D stoquastic Hamiltonians in Section 4.5.

2 Overview

In the PIMC method the Hamiltonian (1) is related by Suzuki’s quantum-to-classical mapping [42, 44] to a system of $\{\pm 1\}$ classical spins on a 2D lattice $\Lambda = \{1, \dots, L\} \times \{1, \dots, n\}$. The new site index $\{1, \dots, L\}$ is known in physics as the “imaginary-time” direction, and it is often useful to think of the set of spin configurations $\Omega := \{-1, 1\}^{n \times L}$ as consisting of “time slices” along the L direction, i.e. for $\mathbf{z} \in \Omega$ we write $\mathbf{z} := (z_1, \dots, z_L)$ with each $z_i \in \{-1, 1\}^n$. The mapping introduces a local energy function $E_\beta : \Omega \rightarrow \mathbb{R}$ on the classical spin configurations in such a way that an estimate of the classical partition function $Z := \sum_{\mathbf{z} \in \Omega} e^{-\beta E_\beta(\mathbf{z})}$ can be used to estimate the quantum partition function $\mathcal{Z}$, and so that samples from the Gibbs distribution $\pi(\mathbf{z}) := e^{-\beta E_\beta(\mathbf{z})}/Z$ can be used to estimate expectation values in the quantum Gibbs state. For the nearest-neighbor version of the transverse Ising models in (5) the energy function E_β is

$$E_\beta(\mathbf{z}) = \sum_{(i,j) \in \Lambda} \left(\frac{K_{j,j+1}^{zz}}{L} z_{i,j} z_{i,j+1} + \frac{K_j^z}{L} z_{i,j} - \beta^{-1} J_i z_{i,j} z_{i+1,j} \right), \quad z_{L+1,j} := z_{1,j}, \quad (9)$$

where $z_{i,j}$ denotes the spin in configuration $\mathbf{z}$ at the site $(i,j) \in \Lambda$, and $J_i := \frac{1}{2} \log \coth \left(\frac{\beta \Gamma_i}{L} \right)$ (see Section 3.1). The couplings along the imaginary-time direction are ferromagnetic, but the couplings along the spatial direction can have varying signs as well as local fields. Our choices of parameters are such that $L = \text{poly}(n)$, and so the general case considers J_i with magnitude up to $\mathcal{O}(\log n)$. Therefore the couplings of the 2D model are highly anisotropic, as in figure 1.

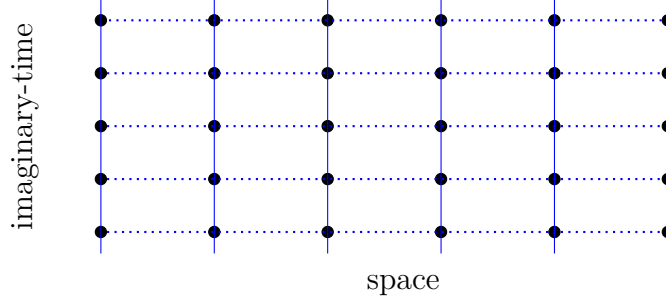


Figure 1: The local fields and the couplings along the spatial direction can be either positive or negative but have magnitudes reduced by $1/L$ (dotted lines), whereas the couplings along the imaginary-time direction can have magnitude growing logarithmically with system size but are always ferromagnetic (solid lines).

The PIMC method proceeds by attempting to sample from π and approximate Z using the Markov chain Monte Carlo method with single-site Metropolis transition probabilities. With this in mind, the primary question is to determine the *mixing time* τ for this Markov chain, which is the minimum number of steps of this Markov chain needed to sample from a distribution that is close to π . We will use the method of canonical paths that was originally introduced by Jerrum and Sinclair (reviewed in [29, 34]) in order to bound the mixing time. The idea of the method is that for every pair of states $\mathbf{x}, \mathbf{y} \in \Omega$ the Markov chain P needs to route an amount of “traffic” equal to $\pi(\mathbf{x})\pi(\mathbf{y})$ along some path $\gamma_{\mathbf{x},\mathbf{y}}$ which connects $\mathbf{x}$ to $\mathbf{y}$ using the transitions of the Markov chain. Define the *congestion* across a particular transition $e = (\mathbf{z}, \mathbf{z}')$ to be the ratio of the total traffic through it (i.e. $\sum_{\mathbf{x}, \mathbf{y}: e \in \gamma_{\mathbf{x},\mathbf{y}}} \pi(\mathbf{x})\pi(\mathbf{y})$) to the amount of probability flow across it at equilibrium, which equals $\pi(\mathbf{z})P(\mathbf{z}, \mathbf{z}')$. The mixing time can then be upper bounded in terms of the maximum congestion through any transition, using standard arguments which we review in Section 4.1.

Our problem now reduces to finding good sets of paths between every pair of configurations $\mathbf{x}, \mathbf{y} \in \Omega$. We use a set of paths inspired by those used to analyze the (unweighted) random walk on the boolean hypercube [29]. Specifically we will replace the entries of $\mathbf{x}$ with entries of $\mathbf{y}$ one at a time. A crucial requirement is that the energies of the intermediate states do not get too large, as this would create a large congestion. For the model (9) the paths first update the classical spins corresponding to the site of qubit 1 in consecutive imaginary-time order, and then proceed to do the same for the sites of qubits 2 through n . For each qubit $j \in \{1, \dots, n\}$ the line of sites $\{(1, j), (2, j), \dots, (L, j)\}$ is called the *worldline* of the j -th qubit, and so these canonical paths have the form of updating each qubit worldline consecutively, always finishing one worldline before starting the next.

An example of an intermediate step along this path is depicted in Fig. 2. Notice that intermediate steps have only introduced a constant number of broken bonds between spins that are neighbors

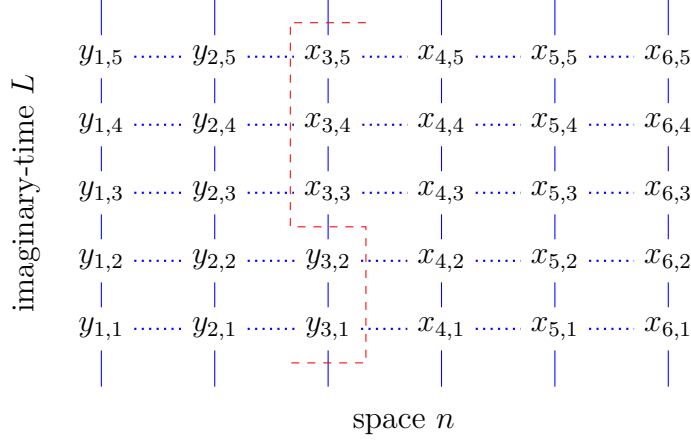


Figure 2: An intermediate point along a canonical path from x to y . We have finished updating qubits 1 and 2 and are in the middle of updating qubit 3. The red line indicates the domain wall formed between the x and y configurations. The change in the classical energy created by this domain wall determines the Markov chain congestion for this set of canonical paths.

in the imaginary time direction, each contributing an amount of energy that is $\mathcal{O}(\beta^{-1} \log n)$, and at most L broken bonds between spins that are neighbors in the space direction, each contributing energy $\mathcal{O}(1/L)$. The relatively low energy of the intermediate configurations along the path is the key to the formal proof of rapid mixing for PIMC applied to 1D generalized TIM in Section 4.3. For 1D generalized TIM with long-range interactions, the energy of the intermediate configuration $\mathbf{z}$ will include contributions from spins that are arbitrarily far away from the cut. Using the assumption on the rate of decay of the magnitude of these interactions, the total contribution to the energy is upper bounded by

$$\sum_{j=1}^{n/2} \sum_{k=\frac{n}{2}+1}^n |j-k|^{-(2+\xi)} = \mathcal{O}(\xi^{-2} + n^{-\xi}),$$

where the bound is proven by viewing the summation as a right Riemann sum of the corresponding integral, which yields an upper bound because the integrand is a monotonically decreasing function. Since we regard ξ as a constant this contribution to the energy is $\mathcal{O}(1)$.

For Hamiltonians of the form (6) the 2-local off-diagonal terms give rise to a classical energy function with interactions involving four spins at a time,

$$E_{\beta}(\mathbf{z}) := \sum_{(i,j) \in \Lambda} \left(\frac{K_j^{zz}}{L} z_{i,j} z_{i,j+1} + \frac{K_j^z}{L} z_{i,j} - \beta^{-1} h_{i,j} (z_{i,j}, z_{i,j+1}, z_{i+1,j}, z_{i+1,j+1}) \right), \quad (10)$$

The $h_{i,j}$ terms (which are implicitly defined by the Gibbs distribution (22)) correspond to strong ferromagnetic interactions between neighboring pairs of spins in neighboring time slices, and violating them increases the energy by an amount that is $\mathcal{O}(\log n)$. If $z_{i,j} \neq z_{i+1,j}$ and $z_{i,j+1} \neq z_{i+1,j+1}$ then we say there is a 2-local jump between time slice i and time slice $i+1$, because the resulting energy contribution from the 4-body interaction will depend on the magnitude of the 2-local off-diagonal quantum terms i.e. K_j^{xx} and K_j^{yy} .

The canonical paths we have described so far will lead to intermediate configurations that potentially create or destroy a non-constant number of 2-local jumps across the domain wall pictured in figure 2. If K_j^{xx} is larger than K_j^x , then configurations $\mathbf{x}$ and $\mathbf{y}$ which contain a non-constant number of 2-local jumps will result in the canonical paths described so far overloading some edges with large congestion. To overcome this we include additional path segments along the canonical paths that remove 2-local jumps from the worldlines near the domain wall in figure 2. This greatly increases the number of paths passing through each edge, but we are still able to obtain a $\text{poly}(n, e^\beta)$ upper bound on the congestion by noticing that most of these paths have endpoints with very small stationary probabilities.

For Hamiltonians of the form (1) it is possible that the off-diagonal terms in (1) do not allow for the PIMC Markov chain with single-site transitions to be ergodic. In this situation some configurations $\mathbf{z} \in \Omega$ will have $\pi(\mathbf{z}) = 0$, which creates a new obstacle to the canonical paths we have described so far. To restore ergodicity under single-site updates we add a fictitious transverse field to the Hamiltonian with a weak coupling that goes to zero with system size:

$$H \rightarrow H - \Gamma \sum_{i=1}^n \sigma_i^x \quad (11)$$

with $\Gamma \leq \delta_{\text{mult}}/n$ so that the subsequent estimate for the partition function will not be further than the allowed error tolerance away from its intended value.

Once the fictitious field is added it once again makes sense to attempt to use the same canonical paths that were applied to systems of the form (6). But now the fact that the off-diagonal matrix elements of the Hamiltonian are no longer symmetric under the global operation of flipping all spins means that removing a single (1-local or 2-local) jump from a given worldline can decrease the stationary probability of that configuration by a factor scaling exponentially in the number of jumps remaining in that worldline. To overcome this problem we show in Section 3.4 that typical configurations distributed according to π will have at most $\mathcal{O}(\beta \log n)$ jumps in any particular worldline. This means that the worst-case intermediate configurations just described may have a stationary probability reduced by $n^{\mathcal{O}(\beta \log n)}$, and this factor bottlenecks the mixing time estimate. Using these facts we describe canonical paths in Section 4.5 within a restricted state space Ω^* in which every configuration has $\mathcal{O}(\beta \log n)$ jumps per worldline, and by either restricting the PIMC Markov chain to Ω^* or applying the leaky random walk conductance bound [15] we obtain a bound on the mixing time within the subset Ω^* that establishes theorem 1.

3 Path integral Monte Carlo

Sections 3.1 through 3.3 rigorously present the underpinnings of the standard PIMC method: the quantum-to-classical mapping, and its application to approximating expectation values of observables as well as the partition function. Section 3.4 proves a general concentration of measure bound which shows that the number of jumps along any worldline in PIMC is $\mathcal{O}(\beta \log n)$ with high probability. Note that this result applies to general stoquastic Hamiltonians (not just the 1D case). This result is used to analyze a restricted state space in Section 4.5.

3.1 Quantum-to-classical mapping

The first step of the quantum-to-classical mapping is to split H as in (1) into a sum of commuting Hamiltonians. For each local term H_i we construct local operators A_i, B_i, C_i such that $-\beta H_i =$

$A_i + B_i + C_i$ and $[A_i, A_j] = [B_i, B_j] = [C_i, C_j] = 0$ for all i, j . Let $A = \sum_i A_i$, $B = \sum_i B_i$, $C = \sum_i C_i$, and hereafter let A_i match the diagonal part of H_i in the computational basis, while B_i and C_i contain the off-diagonal terms of H_i when i is odd or even, respectively. For positive integers L define the Suzuki-Trotter approximation of the partition function $Z = \text{tr} [e^{A+B+C}]$,

$$Z_{\beta,L} := \text{tr} \left[\left(e^{\frac{A}{2L}} e^{\frac{B}{L}} e^{\frac{A}{2L}} e^{\frac{C}{L}} \right)^L \right]. \quad (12)$$

The fact that $\lim_{L \rightarrow \infty} Z_{\beta,L} = \mathcal{Z}_\beta$ is the content of the 1875 Lie product formula, while for finite L the Suzuki-Trotter approximation is related to the quantum partition function (3) by

$$e^{-\delta} \mathcal{Z}_\beta \leq Z_{\beta,L} \leq e^\delta \mathcal{Z}_\beta, \quad (13)$$

for some δ that is $\mathcal{O}(\|H\|^3/L^3)$. This error bound follows from two facts about the matrix exponential: a variant of the usual Suzuki-Trotter expansion, and a continuity bound. Both appear in [8] for transverse Ising models, and the proof for three commuting layers A, B, C is similar and so we omit the details.

Before considering the general case of H as in (1) it is useful to consider the simplified case of (12) for generalized TIM of the form (5) by taking $A = -\beta \text{diag}(H)$, $B = -\beta(H - \text{diag}(H))$, and $C = 0$ so that the Suzuki-Trotter approximation reduces to

$$Z_{\beta,L} = \text{tr} \left[\left(e^{\frac{A}{2L}} e^{\frac{B}{L}} e^{\frac{A}{2L}} \right)^L \right] = \text{tr} \left[\left(e^{\frac{A}{L}} e^{\frac{B}{L}} \right)^L \right]. \quad (14)$$

For transverse Ising models, the next step is to expand the Suzuki-Trotter approximation (14) to the partition function by inserting several complete sets of states,

$$Z_{\beta,L} = \sum_{z_1 \in \{-1,1\}^n} \langle z_1 | \left(e^{\frac{A}{L}} e^{\frac{B}{L}} \right)^L | z_1 \rangle \quad (15)$$

$$= \sum_{z_1, \dots, z_L} \prod_{i=1}^L \langle z_i | e^{\frac{A}{L}} e^{\frac{B}{L}} | z_{i+1} \rangle, \quad z_{L+1} := z_1 \quad (16)$$

$$= \sum_{z_1, \dots, z_L} \prod_{i=1}^L e^{\frac{A(z_i)}{L}} \langle z_i | e^{\frac{B}{L}} | z_{i+1} \rangle. \quad (17)$$

where in the last step we have used the fact that $A = -\beta \text{diag}(H)$ is diagonal in the computational basis and defined $A(z_i) := \langle z_i | A | z_i \rangle$. The Gibbs distribution of the classical 2D spin system is defined by equating $e^{-\beta E_\beta(z_1, \dots, z_L)}$ with the summand of (17). The closed form (9) is obtained applying the identity $e^{\omega \sigma^x} = \cosh(\omega)I + \sinh(\omega)\sigma^x$ to

$$\langle z_i | e^{\frac{B}{L}} | z_{i+1} \rangle = \langle z_i | e^{-\frac{\beta \Gamma}{L} \sum_{j=1}^n \sigma_j^x} | z_{i+1} \rangle = \prod_{j=1}^n \langle z_{i,j} | e^{-\frac{\beta \Gamma}{L} \sigma_j^x} | z_{i+1,j} \rangle.$$

For a general 1D stoquastic Hamiltonian of the form in (1) we will apply (12) with

$$A = -\beta \text{diag}(H) \quad , \quad B = -\beta \sum_{i \in \text{Odd}} [H_i - \text{diag}(H_i)] \quad , \quad C = -\beta \sum_{i \in \text{Even}} [H_i - \text{diag}(H_i)], \quad (18)$$

Expanding (12) with $2L$ complete sets of states,

$$Z_{\beta,L} = \sum_{z_1 \in \{-1,1\}^n} \langle z_1 | \left(e^{\frac{A}{2L}} e^{\frac{B}{L}} e^{\frac{A}{2L}} e^{\frac{C}{L}} \right)^L | z_1 \rangle \quad (19)$$

$$= \sum_{z_1, \dots, z_{2L}} \prod_{i \in \text{Odd}}^{2L-1} \langle z_i | e^{\frac{A}{2L}} e^{\frac{B}{L}} | z_{i+1} \rangle \langle z_{i+1} | e^{\frac{A}{2L}} e^{\frac{C}{L}} | z_{i+2} \rangle \quad (20)$$

$$= \sum_{z_1, \dots, z_{2L}} e^{\sum_{i=1}^{2L} \frac{A(z_i)}{2L}} \prod_{i \in \text{Odd}}^{2L-1} \langle z_i | e^{\frac{B}{L}} | z_{i+1} \rangle \langle z_{i+1} | e^{\frac{C}{L}} | z_{i+2} \rangle, \quad (21)$$

where periodic boundary conditions $z_{2L+1} := z_1$ are taken in (20). At this point the essential difference from the transverse Ising case is clear: the off-diagonal part of the Hamiltonian is split into two commuting layers, these layers alternate along the imaginary-time direction, and this causes the coupling between classical spins at neighboring time slices to alternate as well. For ease of notation define B_i to be $2B$ if i is odd, and $2C$ if i is even, and rescale $L \rightarrow L/2$ to obtain a distribution π from (21),

$$\pi(\mathbf{z}) := \frac{1}{Z_{\beta,L}} e^{\sum_{i=1}^L \frac{A(z_i)}{L}} \prod_{i=1}^L \langle z_i | e^{\frac{B_i}{L}} | z_{i+1} \rangle. \quad (22)$$

3.2 Approximating observables

In this section we consider an observable $\mathcal{O}$ for which we would like to estimate $\langle \mathcal{O} \rangle_\beta := \text{tr}[\mathcal{O} \rho_\beta]$. We justify this by demonstrating closeness of $\langle \mathcal{O} \rangle_\beta$ to a value $\langle \mathcal{O} \rangle_{\beta,L}$ that is defined in terms of the Suzuki-Trotter approximation in equation (34). The expectation $\langle \mathcal{O} \rangle_\beta$ can be expressed as the derivative of a partition function by defining

$$\mathcal{Z}_\beta(\zeta) := \text{tr}[e^{-\beta \mathcal{H} + \zeta \mathcal{O}}], \quad (23)$$

and using the fact that

$$\langle \mathcal{O} \rangle_\beta = \frac{1}{\mathcal{Z}_\beta(0)} \frac{\partial \mathcal{Z}_\beta}{\partial \zeta}(0). \quad (24)$$

The next proposition shows that $\langle \mathcal{O} \rangle_\beta$ can be approximated by a finite difference.

Proposition 2. *Suppose that $\hat{\mathcal{Z}}_\beta(0)$ and $\hat{\mathcal{Z}}_\beta(\zeta)$ satisfy*

$$e^{-\delta} \leq \frac{\hat{\mathcal{Z}}_\beta(0)}{\mathcal{Z}_\beta(0)}, \frac{\hat{\mathcal{Z}}_\beta(\zeta)}{\mathcal{Z}_\beta(\zeta)} \leq e^\delta,$$

for $\zeta = \sqrt{3\delta}/\|\mathcal{O}\|$. Then the estimator

$$\widehat{\langle \mathcal{O} \rangle}_\beta := \frac{\hat{\mathcal{Z}}_\beta(\zeta) - \hat{\mathcal{Z}}_\beta(0)}{\zeta \hat{\mathcal{Z}}_\beta(0)}. \quad (25)$$

satisfies the bound

$$|\widehat{\langle \mathcal{O} \rangle}_\beta - \langle \mathcal{O} \rangle_\beta| \leq 2\sqrt{\delta} \|\mathcal{O}\| \quad (26)$$

Proof. We use first Taylor's theorem, then the bound $\text{tr } AB \leq \|A\| \|B\|_1$ to obtain:

$$\left| \mathcal{Z}_\beta(\zeta) - \mathcal{Z}_\beta(0) - \zeta \frac{\partial \mathcal{Z}_\beta}{\partial \zeta}(0) \right| \leq \frac{\zeta^2}{2} \max_{\zeta' \in [0, \zeta]} \frac{\partial^2 \mathcal{Z}_\beta}{\partial \zeta^2}(\zeta') \quad (27a)$$

$$\leq \zeta^2 \|\mathcal{O}\|^2 \max_{\zeta' \in [0, \zeta]} \mathcal{Z}_\beta(\zeta') \quad (27b)$$

$$\leq \zeta^2 \|\mathcal{O}\|^2 \mathcal{Z}_\beta(0) e^{\zeta \|\mathcal{O}\|}. \quad (27c)$$

Dividing both sides of (27) by $\zeta \mathcal{Z}_\beta(0)$ we have

$$\left| \langle \mathcal{O} \rangle_\beta - \frac{\mathcal{Z}_\beta(\zeta) - \mathcal{Z}_\beta(0)}{\zeta \mathcal{Z}_\beta(0)} \right| = \left| \langle \mathcal{O} \rangle_\beta - \left(\frac{\mathcal{Z}_\beta(\zeta)}{\zeta \mathcal{Z}_\beta(0)} - \frac{1}{\zeta} \right) \right| \leq \zeta \|\mathcal{O}\|^2 e^{\zeta \|\mathcal{O}\|}. \quad (28)$$

Then

$$\widehat{\langle \mathcal{O} \rangle}_\beta - \langle \mathcal{O} \rangle_\beta = \frac{\hat{\mathcal{Z}}_\beta(\zeta)}{\zeta \hat{\mathcal{Z}}_\beta(0)} - \frac{1}{\zeta} - \langle \mathcal{O} \rangle_\beta \quad (29)$$

$$\leq e^{2\delta} \frac{\mathcal{Z}_\beta(\zeta)}{\zeta \mathcal{Z}_\beta(0)} - \frac{1}{\zeta} - \langle \mathcal{O} \rangle_\beta \quad (30)$$

$$\leq (e^{2\delta} - 1) \frac{\mathcal{Z}_\beta(\zeta)}{\zeta \mathcal{Z}_\beta(0)} + \zeta \|\mathcal{O}\|^2 e^{\zeta \|\mathcal{O}\|} \quad (31)$$

$$\leq (3\delta/\zeta + \zeta \|\mathcal{O}\|^2) e^{\zeta \|\mathcal{O}\|} \quad (32)$$

$$\leq 2\sqrt{\delta} \|\mathcal{O}\| \quad (33)$$

In the last step we have used the assumption that $\delta \leq 1/21$ to simplify the numerical constants. A similar argument shows that $\widehat{\langle \mathcal{O} \rangle}_\beta - \langle \mathcal{O} \rangle_\beta \geq -2\sqrt{\delta} \|\mathcal{O}\|$, thus completing the proof. $\square$

By (25), (13), and the triangle inequality, the expectation $\langle \mathcal{O} \rangle_\beta$ can be approximated by

$$\langle \mathcal{O} \rangle_{\beta, L} = \frac{1}{Z_{\beta, L}(0)} \frac{\partial}{\partial \zeta} Z_{\beta, L}(0), \quad (34)$$

where $Z_{\beta, L}(\zeta)$ is a Suzuki-Trotter approximation to (23). If $\mathcal{O}$ is diagonal use the form in Section 3.1 with $A \rightarrow A + \zeta \mathcal{O}$,

$$Z_{\beta, L}(\zeta) = \sum_{\mathbf{z} \in \Omega} e^{\frac{1}{L} \sum_{i=1}^L (A(z_i) + \zeta \mathcal{O}(z_i))} \prod_{i=1}^L \langle z_i | e^{\frac{\mathcal{B}_i}{L}} | z_{i+1} \rangle. \quad (35)$$

Evaluating the derivative at $\zeta = 0$,

$$\frac{1}{Z_{\beta, L}(0)} \frac{\partial}{\partial \zeta} Z_{\beta, L}(0) = \frac{1}{Z_{\beta, L}} \sum_{\mathbf{z} \in \Omega} \left(\frac{1}{L} \sum_{i=1}^L \mathcal{O}(z_i) \right) e^{\frac{1}{L} \sum_{i=1}^L A(z_i)} \prod_{i=1}^L \langle z_i | e^{\frac{\mathcal{B}_i}{L}} | z_{i+1} \rangle \quad (36)$$

$$= \sum_{\mathbf{z} \in \Omega} \pi(\mathbf{z}) \left(\frac{1}{L} \sum_{i=1}^L \mathcal{O}(z_i) \right) = \frac{1}{L} \sum_{i=1}^L \langle \mathcal{O}(z_i) \rangle_\pi = \langle \mathcal{O}(z_1) \rangle_\pi \quad (37)$$

where the final expression follows by the symmetry of π under cyclic permutations of the time slices. For off-diagonal $\mathcal{O}$,

$$Z_{\beta,L}(\zeta) = \sum_{\mathbf{z} \in \Omega} e^{\frac{1}{L} \sum_{i=1}^L A(z_i)} \prod_{i=1}^L \langle z_i | e^{\frac{1}{L} (\mathcal{B}_i + \zeta \mathcal{O}/2)} | z_{i+1} \rangle. \quad (38)$$

Evaluating the derivative at $\zeta = 0$,

$$\langle \mathcal{O} \rangle_{\beta,L} = \frac{1}{Z_{\beta,L}(0)} \sum_{\mathbf{z} \in \Omega} e^{\frac{1}{L} \sum_{i=1}^L A(z_i)} \sum_{k=1}^L \langle z_k | \left(\frac{\mathcal{O}}{2L} \right) e^{\frac{\mathcal{B}_k}{L}} | z_{k+1} \rangle \prod_{\substack{i=1 \\ i \neq k}}^L \langle z_i | e^{\frac{\mathcal{B}_i}{L}} | z_{i+1} \rangle \quad (39)$$

$$= \sum_{\mathbf{z} \in \Omega} \left(\frac{1}{2L} \sum_{k=1}^L \frac{\langle z_k | \mathcal{O} e^{\frac{\mathcal{B}_k}{L}} | z_{k+1} \rangle}{\langle z_k | e^{\frac{\mathcal{B}_k}{L}} | z_{k+1} \rangle} \right) \pi(\mathbf{z}) \quad (40)$$

$$= \frac{1}{2L} \sum_{k=1}^L \left\langle \frac{\langle z_k | \mathcal{O} e^{\frac{\mathcal{B}_k}{L}} | z_{k+1} \rangle}{\langle z_k | e^{\frac{\mathcal{B}_k}{L}} | z_{k+1} \rangle} \right\rangle_{\pi} \quad (41)$$

$$= \frac{1}{2} \sum_{\mathbf{z} \in \Omega} \left[\frac{\langle z_1 | \mathcal{O} e^{\frac{\mathcal{B}_1}{L}} | z_2 \rangle}{\langle z_1 | e^{\frac{\mathcal{B}_1}{L}} | z_2 \rangle} + \frac{\langle z_2 | \mathcal{O} e^{\frac{\mathcal{B}_2}{L}} | z_3 \rangle}{\langle z_2 | e^{\frac{\mathcal{B}_2}{L}} | z_3 \rangle} \right] \pi(\mathbf{z}) \quad (42)$$

where the last step follows by symmetry of π under even cyclic permutations of the time slices. Since the summand of (42) is efficiently computable and has bounded magnitude we can estimate it using a standard Monte Carlo procedure (see lemma 3) once one has an efficient means of obtaining independent samples from π .

3.3 Approximating partition functions

In Section 4 we will prove that for any tolerance $\delta > 0$ the PIMC method can efficiently output an element of Ω sampled from a distribution $\tilde{\pi}$ which is within $|\tilde{\pi} - \pi| \leq \delta$. Here we explain how these samples can be used to approximate the partition function with an application of the standard telescoping product technique [5, 28]¹. We express the partition function $Z(\beta)$ as a product of terms involving the partition function at higher temperatures,

$$Z(\beta) = Z(0) \frac{Z(\beta_1)}{Z(\beta_0)} \cdots \frac{Z(\beta_k)}{Z(\beta_{k-1})} \quad (43)$$

where the temperature schedule $\beta_0 < \beta_1 < \dots < \beta_k$ with $\beta_0 = 0$ and $\beta_k = \beta$ is chosen so that each ratio in the telescoping product is bounded. Defining $w_{\beta_i}(\mathbf{z}) := Z(\beta_i) \pi_{\beta_i}(\mathbf{z})$, note that

$$\left\langle \frac{w_{\beta_i}}{w_{\beta_{i-1}}} \right\rangle_{\pi_{\beta_{i-1}}} = \sum_{\mathbf{z} \in \Omega} \frac{w_{\beta_i}(\mathbf{z})}{w_{\beta_{i-1}}(\mathbf{z})} \pi_{\beta_{i-1}}(\mathbf{z}) = \frac{Z(\beta_i)}{Z(\beta_{i-1})}. \quad (44)$$

¹The techniques used in this section to estimate the partition function (or equivalently the free energy) are necessary because the entropy term in the free energy cannot be directly estimated as an observable. If one is only interested in the estimation of specific observables then it suffices to use the end result(42).

At infinite temperature the quantum partition function is $\mathcal{Z}(0) = 2^n$. Further, a uniform schedule with $k = \mathcal{O}(\beta\|H\|\log(\beta\|H\|))$ ensures that

$$\frac{1}{e} \leq \frac{\mathcal{Z}(\beta_i)}{\mathcal{Z}(\beta_{i-1})} \leq 1. \quad (45)$$

Note that in infinite temperature limit $\beta = 0$ the classical effective systems (9) and (10) have infinite coupling strength between spins in the imaginary-time direction, and so the Markov chain we analyze in Section 4 will not be ergodic. This is unusual from the perspective of classical spin systems, but it happens in this case because of the temperature dependence of the couplings. To deal with this the first expectation $\langle w_{\beta_1}/w_{\beta_0} \rangle_{\pi_{\beta_0}}$ in the telescoping product can be computed using the fact that π_{β_0} is equivalent to the uniform distribution on the subset $\Omega_{\text{classical}} \subseteq \Omega$ consisting of configurations $(z_1, \dots, z_L)$ which have $z_i = z_j$ for all $i, j = 1, \dots, L$.

Finally, in order to bound the number of samples needed to compute the expectation values in the telescoping product we will make use of the Hoeffding bound.

Lemma 3 ([24]). *Let $X_1, \dots, X_t$ be independent random variables satisfying $|X_i| \leq 1$ and $\mathbb{E}[X_i] = \bar{X}$. Then*

$$\mathbb{P}\left[\left|\frac{1}{t} \sum_{i=1}^t X_i - \bar{X}\right| \geq \delta\right] \leq 2e^{-t\delta^2/2} \quad (46)$$

Define $M(z) = e^{-2\delta} w_{\beta_i}(z)/w_{\beta_{i+1}}(z)$, so that $|M(z)| \leq 1$. Observe that $|\langle M \rangle_{\pi} - \langle M \rangle_{\tilde{\pi}}| \leq \|\pi - \tilde{\pi}\|_1 \|M\| \leq \delta$. We will estimate $\langle M \rangle_{\tilde{\pi}}$ by drawing t samples from $\tilde{\pi}$, which we call $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(t)}$. Our estimator will simply be the sample mean, namely

$$\hat{M} := \frac{1}{t} \sum_{i=1}^t M(\mathbf{z}_i). \quad (47)$$

By Lemma 3 we have

$$\mathbb{P}[|\hat{M} - \langle M \rangle_{\tilde{\pi}}| \geq \delta] \leq 2e^{-t\delta^2/2}. \quad (48)$$

We conclude that taking $t = \mathcal{O}(1/\delta^2)$ samples yields an $\mathcal{O}(\delta)$ approximation to $\langle M \rangle_{\tilde{\pi}}$ with high probability. Therefore a polynomial number of samples will suffice to estimate each term of the telescoping product.

3.4 Restricting the PIMC state space

In this section we justify restricting PIMC to a subset of the state space that limits the number of jumps in any particular worldline to be $\mathcal{O}(\beta \log n)$ (this restricted state space will be used for the mixing time analysis in Section 4.5). For a configuration $\mathbf{z} = (z_1, \dots, z_L)$ the number of jumps in the worldline of the j -th qubit can be expressed in terms of the Pauli operator X_j ,

$$d_j(\mathbf{z}) := |\{i : z_{i,j} \neq z_{i+1,j}\}| = \sum_{i=1}^L \langle z_i | X_j | z_{i+1} \rangle.$$

The strategy is to bound the k -th moment of the random variable d_j with respect to the QMC stationary distribution π and use the fact that

$$\Pr[d_j \geq a]_{\pi} \leq \frac{\mathbb{E}_{\pi} d_j^k}{a^k}.$$

These moments satisfy

$$\begin{aligned}
\mathbb{E}_\pi d_j^k &:= \sum_{\mathbf{z} \in \Omega} d_j^k(\mathbf{z}) \pi(\mathbf{z}) \\
&= \sum_{\mathbf{z} \in \Omega} \left(\sum_{i=1}^L \langle z_i | X_j | z_{i+1} \rangle \right)^k \pi(\mathbf{z}) \\
&= \sum_{\mathbf{z} \in \Omega} \sum_{i_1, \dots, i_k=1}^L \prod_{r=1}^k \langle z_{i_r} | X_j | z_{i_r+1} \rangle \pi(\mathbf{z}) \\
&= \frac{1}{Z_{\beta, L}} \sum_{\mathbf{z} \in \Omega} \sum_{i_1, \dots, i_k=1}^L \prod_{r=1}^k \langle z_{i_r} | X_j | z_{i_r+1} \rangle \left(e^{\sum_{i=1}^L \frac{A(z_i)}{L}} \prod_{i=1}^L \langle z_i | e^{\frac{B_i}{L}} | z_{i+1} \rangle \right). \tag{49}
\end{aligned}$$

where A, B_i are defined in Section 3.1. To merge the two products in the expression above into a simpler trace expression, we want to show the following inequality for any neighboring time slices z_i and z_{i+1} ,

$$\langle z_i | X_j | z_{i+1} \rangle \leq 4 \frac{\beta}{L} \frac{\langle z_i | X_j e^{B_i/L} | z_{i+1} \rangle}{\langle z_i | e^{B_i/L} | z_{i+1} \rangle}, \tag{50}$$

where we assume the local terms are normalized so that for some $\Gamma > 0$

$$\Gamma \leq |\langle z_i | H | z_{i+1} \rangle| \leq 1 \quad \forall z_i, z_{i+1} : |z_i - z_{i+1}| = 1.$$

If $\langle z_i | X_j | z_{i+1} \rangle = 0$ then (50) is satisfied. Otherwise if $\langle z_i | X_j | z_{i+1} \rangle = 1$ then the numerator on the RHS of (50) satisfies

$$\langle z_i | X_j e^{B_i/L} | z_{i+1} \rangle = \langle z_{i+1} | e^{B_i/L} | z_{i+1} \rangle \geq e^{-\beta \|H\|/L} \geq \frac{1}{2}. \tag{51}$$

To upper bound the denominator of the RHS of (50), the fact that B_i is a commuting Hamiltonian implies

$$\begin{aligned}
\langle z_i | e^{B_i/L} | z_{i+1} \rangle &= \langle z_{i,j} z_{i,j+1} | e^{B_{i,j}/L} | z_{i+1,j}, z_{i+1,j+1} \rangle \\
&= \sum_{q=1}^{\infty} \frac{1}{q!} \langle z_{i,j} z_{i,j+1} | (B_{i,j}/L)^q | z_{i+1,j}, z_{i+1,j+1} \rangle \\
&\leq e^{\frac{\beta \|H_j\|}{L}} - 1 \\
&\leq \frac{1}{\ln(2)} \frac{\beta}{L},
\end{aligned}$$

where in the last step we have used $\|H_j\| \leq 1$, $\frac{\beta}{L} \leq \ln(2)$ and the convexity of e^x . Using $\ln(2)/2 \geq 1/4$ we have established (50).

To apply this result to (49) we will use the permutation symmetry of the $i_1, \dots, i_k$ variables to restrict to a sum over ordered k -tuples $(i_1, \dots, i_k)$ which satisfy $i_1 \leq i_2 \leq \dots \leq i_k$.

$$\mathbb{E}_\pi d_j^k \leq \frac{k!}{Z} \sum_{\mathbf{z} \in \Omega} \sum_{1 \leq i_1 \leq \dots \leq i_k \leq L} \prod_{r=1}^k \langle z_{i_r} | X_j | z_{i_r+1} \rangle \left(e^{\sum_{i=1}^L \frac{A(z_i)}{L}} \prod_{i=1}^L \langle z_i | e^{\frac{B_i}{L}} | z_{i+1} \rangle \right) \tag{52}$$

$$\leq \frac{4^k k! \beta^k}{Z L^k} \sum_{\mathbf{z} \in \Omega} \sum_{1 \leq i_1 \leq \dots \leq i_k \leq L} \prod_{r=1}^k \frac{\langle z_{i_r} | X_j e^{\frac{B_{i_r}}{L}} | z_{i_r+1} \rangle}{\langle z_{i_r} | e^{\frac{B_{i_r}}{L}} | z_{i_r+1} \rangle} \left(e^{\sum_{i=1}^L \frac{A(z_i)}{L}} \prod_{i=1}^L \langle z_i | e^{\frac{B_i}{L}} | z_{i+1} \rangle \right) \tag{53}$$

Defining $t_r = (i_r - i_{r-1})\beta/L$ (as well as $t_1 = r_1\beta/L, t_k = (L - i_k)\beta/L$) this expression can be seen as the trace of a product of operators,

$$\mathbb{E}_\pi d_j^k \leq \frac{k!(4\beta)^k}{ZL^k} \sum_{\substack{t_1, \dots, t_k \\ t_1 + \dots + t_k = \beta}}^L \text{tr} [M_{t_1} R_{t_1} X_j L_{t_2} M_{t_2} R_{t_2} X_j \dots X_j L_{t_k} M_{t_k}] \quad (54)$$

where

$$M_{t_i} := \left(e^{B_{\text{even}}/L} e^{A/L} e^{B_{\text{odd}}/L} e^{A/L} e^{B_{\text{even}}/L} \right)^{\lfloor t_i \beta / L \rfloor - 2}$$

and L_i, R_{t_i} can each denote one of several products of operators depending on t_i , which is determined by where the X_j operators disrupt the sequence,

$$L_{t_i} = \begin{cases} e^{A/L} \\ e^{A/L} e^{B_{\text{odd}}/L} \\ e^{A/L} e^{B_{\text{odd}}/L} e^{A/L} \\ e^{A/L} e^{B_{\text{odd}}/L} e^{A/L} e^{B_{\text{even}}/L} \end{cases}, \quad R_{t_i} = \begin{cases} e^{B_{\text{even}}/L} \\ e^{A/L} e^{B_{\text{even}}/L} \\ e^{B_{\text{odd}}/L} e^{A/L} e^{B_{\text{even}}/L} \\ e^{A/L} e^{B_{\text{odd}}/L} e^{A/L} e^{B_{\text{even}}/L} \end{cases}.$$

The purpose of this decomposition is that each M_{t_i} is a PSD operator raised to an positive integer power, and the rest of the Matrices L_{t_i}, R_{t_i}, X_j are all either PSD operators or are explicit products of 2,3, or 4 PSD operators. Now the traces in (54) can be bounded using classic inequality due to Ky Fan [18, Theorem 1]. Let $P_1 \dots P_N$ satisfy $P_i \succeq 0$ for each i , and let $\sigma(P_i)$ be a diagonal matrix with the eigenvalues of P_i along the diagonal in descending order. Then

$$\text{tr} [P_1 \dots P_N] \leq \text{tr} [\sigma(P_1) \dots \sigma(P_N)]$$

For our purposes define the notation $\sigma(R_{t_i})$ to be the product of diagonal matrices of eigenvalues of the operators that make up R_{t_i} e.g. if $R_{t_i} = e^{A/L} e^{B_{\text{even}}/L}$ then $\sigma(R_{t_i}) = \sigma(e^{A/L}) \sigma(e^{B_{\text{even}}/L})$. Furthermore we may take $B_{\text{even}}, B_{\text{odd}} \succeq 0$ without loss of generality (by shifting the overall Hamiltonian by a multiple of the identity), so that $\mathbb{1} \succeq \sigma(R_{t_i}), \sigma(L_{t_i})$ for each t_i . Therefore each trace in (54) satisfies

$$\begin{aligned} & \text{tr} [M_{t_1} R_{t_1} X_j L_{t_2} M_{t_2} R_{t_2} X_j \dots X_j L_{t_k} M_{t_k}] \\ & \leq \text{tr} [\sigma(M_{t_1}) \sigma(R_{t_1}) \sigma(X_j) \sigma(L_{t_2}) \sigma(M_{t_2}) \sigma(R_{t_2}) \sigma(X_j) \dots \sigma(X_j) \sigma(L_{t_k}) \sigma(M_{t_k})] \end{aligned} \quad (55)$$

$$\leq \text{tr} [\sigma(M_{t_1}) \sigma(M_{t_2}) \dots \sigma(M_{t_k})] \quad (56)$$

$$\begin{aligned} & = \text{tr} \left[\left(e^{B_{\text{even}}/L} e^{A/L} e^{B_{\text{odd}}/L} e^{A/L} e^{B_{\text{even}}/L} \right)^{L-2k} \right] \\ & = Z_{\beta, L-2k} \end{aligned}$$

where the fact that $\mathbb{1} \succeq \sigma(R_{t_i}), \sigma(L_{t_i})$ is used to go from (55) to (56). Now for $k \ll L$ we have $Z_{\beta, L} \approx Z_{\beta, L-2k}$, so this implies that $\mathbb{E}_\pi d_j^k \leq k! \beta^k$. Therefore taking $k = c \log n$ for some $c > 0$ and using $k! \leq (k/e)^k$ yields

$$\Pr_\pi [d_j \geq c\beta \log n] \leq \frac{k! \beta^k}{c^k \beta^k \log^k b} = \frac{k!}{k^k} \leq e^{-k} = n^{-c}.$$

4 Proofs of rapid mixing

Having established the correctness of the PIMC method in section 3 we now proceed to analyze the mixing time of the associated Markov chains. The mixing time analysis is divided into cases. In section 4.1 we review the main techniques that are common to each case in our analysis. Section 4.3 contains the mixing analysis for transverse Ising models, while sections 4.4 and 4.5 analyze Hamiltonians of the form (6) and (1), respectively.

4.1 Markov chains, mixing times, and canonical paths

Let P be the transition matrix of a reversible Markov chain defined on a state space Ω with stationary distribution π . Let $\mathbf{z}$ be the state of the chain at time $t = 0$, and define $P^t(\mathbf{z}, \cdot)$ to be the distribution of the state of the chain at time t . The variation distance of the chain at time t starting from the initial state $\mathbf{z}$ at time $t = 0$ is

$$d_{\mathbf{z}}(t) := \max_{A \subseteq \Omega} |P^t(\mathbf{z}, A) - \pi(A)| = \frac{1}{2} \sum_{\mathbf{z}' \in \Omega} |P^t(\mathbf{z}, \mathbf{z}') - \pi(\mathbf{z}')| \quad (57)$$

Define $\tau(\epsilon)$ to be the worst-case time needed to be within variation distance ϵ of the stationary distribution,

$$\tau(\epsilon) := \max_{\mathbf{z} \in \Omega} \min_t \{t : d_{\mathbf{z}}(t') \leq \epsilon \quad \forall t' \geq t\}. \quad (58)$$

Define the mixing time to be $\tau_{\text{mix}} := \tau(1/4)$, and note that $\tau(\epsilon) = \tau_{\text{mix}} \log \epsilon^{-1}$ [34].

To bound the mixing time of P we use the method of canonical paths [29]. A path from $x \in \Omega$ to $y \in \Omega$ is a sequence $\gamma_{x,y} = (v_1, \dots, v_k)$ of states with $v_1 = x, v_k = y$ and $P(v_i, v_{i+1}) > 0$ for $i = 1, \dots, k-1$. A set $\mathcal{P} = \{\gamma_{x,y}\}$ containing a path that connects every pair of states is called a set of canonical paths. If $\ell = \max_{\gamma \in \mathcal{P}} |\gamma|$ is the maximum length of any path in $\mathcal{P}$ and $\pi_{\min} := \min_{\mathbf{z} \in \Omega} \pi(\mathbf{z})$, then the mixing time τ_{mix} is $\mathcal{O}(R \ell \ln \pi_{\min}^{-1})$, where the congestion R is defined as

$$R := \max_{(v,v') \in E(\Omega)} \frac{1}{\pi(v)P(v,v')} \sum_{\substack{\gamma_{x,y} \in \mathcal{P} \\ (v,v') \in \gamma_{x,y}}} \pi(x)\pi(y) \quad (59)$$

In order to compute the congestion we will use a standard technique called encoding. For an edge $e = (v, v')$, let $\mathcal{P}(e)$ be the set of paths in $\mathcal{P}$ which pass through e . An encoding is an assignment of an injective function $\eta_e : \mathcal{P}(e) \rightarrow \Omega$ to every edge $e \in E(\Omega)$. More generally, it can be useful to consider injective encodings $\eta_e : \mathcal{P}(e) \rightarrow \Omega \times \Theta$ that map at most $G = |\Theta|$ paths through the e to each state in Ω [28]. Let $\eta_e(\gamma_{x,y}) = (\eta_e(x, y), \theta_e(x, y))$, and suppose there is some M such that

$$\pi(x)\pi(y) \leq M\pi(v)\pi(\eta_e(x, y)) \quad \forall x, y \in \Omega, \quad (60)$$

then the congestion satisfies

$$R = \max_{(v,v') \in E(\Omega)} \frac{1}{\pi(v)P(v,v')} \sum_{\substack{\gamma_{x,y} \in \mathcal{P} \\ (v,v') \in \gamma_{x,y}}} \pi(x)\pi(y) \quad (61)$$

$$\leq \max_{(v,v') \in E(\Omega)} \frac{M \cdot G}{\pi(v)P(v,v')} \sum_{\substack{\gamma_{x,y} \in \mathcal{P} \\ (v,v') \in \gamma_{x,y}}} \pi(v)\pi(\eta_e(x,y)) \quad (62)$$

$$\leq M \cdot G \cdot P_{\min}^{-1} \quad (63)$$

where $P_{\min} := \min_{(\mathbf{z}, \mathbf{z}') \in E(\Omega)} P(\mathbf{z}, \mathbf{z}')$, (62) follows from property (60), and (63) follows from the injectivity property of the encoding.

4.2 PIMC transition probabilities

To sample from the classical Gibbs distribution (22) the PIMC method uses a Markov chain which chooses a site in the classical lattice Λ uniformly at random and proposes to flip the bit at that site to make a transition $\mathbf{z} \rightarrow \mathbf{z}'$ with an acceptance probability $P(\mathbf{z}, \mathbf{z}')$ given by the Metropolis rule [34],

$$P(\mathbf{z}, \mathbf{z}') := \begin{cases} \frac{1}{2nL} \min \left\{ 1, \frac{\pi(\mathbf{z}')}{\pi(\mathbf{z})} \right\}, & \mathbf{z} \neq \mathbf{z}' \\ 1 - \sum_{\mathbf{z}'' \neq \mathbf{z}} P(\mathbf{z}, \mathbf{z}''), & \mathbf{z} = \mathbf{z}' \end{cases}. \quad (64)$$

As long as the state space is a connected graph, the transition matrix P with transitions satisfying (64) will converge to the stationary distribution π .

For the sake of lower bounding the minimum transition probability $P_{\min}$ from the previous section it therefore suffices to lower bound $\pi(\mathbf{z}')/\pi(\mathbf{z})$ where $\mathbf{z}$ and $\mathbf{z}'$ differ only on a single site. Suppose $\mathbf{z}$ differs from $\mathbf{z}'$ at position (i, j) , then after canceling common factors in (22) we have

$$\frac{\pi(\mathbf{z}')}{\pi(\mathbf{z})} = e^{\frac{1}{L}[A(z'_i) - A(z_i)]} \frac{\langle z_{i-1} | e^{\frac{B_{i-1}}{L}} | z'_i \rangle \langle z'_i | e^{\frac{B_i}{L}} | z_{i+1} \rangle}{\langle z_{i-1} | e^{\frac{B_{i-1}}{L}} | z_i \rangle \langle z_i | e^{\frac{B_i}{L}} | z_{i+1} \rangle}. \quad (65)$$

Depending on the off-diagonal matrix elements in B_i , the numerator of (65) can potentially be zero. This means that the state space Ω may not be connected by single bit flips, because the stationary distribution does not have support on all of Ω . This can be avoided in general by adding a 1-local transverse field with sufficiently small magnitude Γ to avoid disturbing the resulting approximation. With this we can lower bound the ratio (65) using the fact that B_i is a nonnegative matrix, which implies

$$\langle z | \left(I + \frac{B_i}{L} \right) | z' \rangle \leq \langle z | e^{\frac{B_i}{L}} | z' \rangle \quad \text{for all } z, z' \in \{-1, 1\}^n. \quad (66)$$

The denominator of (65) is upper bounded by 1, and (66) implies that the numerator is at least $(\beta\Gamma/L)^2$, therefore

$$\frac{\pi(\mathbf{z}')}{\pi(\mathbf{z})} \geq e^{-\frac{4\beta}{L}} \left(\frac{\beta\Gamma}{L} \right)^2. \quad (67)$$

4.3 Transverse Ising models

In this section we will describe a set of canonical paths which will be used to show rapid mixing of PIMC applied to Hamiltonians of the form (5). Let $\mathbf{x}, \mathbf{y} \in \Omega$ be spin configurations with $\mathbf{x} = (x_{1,1}, \dots, x_{L,1}, \dots, x_{1,n}, \dots, x_{L,n})$ and $\mathbf{y} = (y_{1,1}, \dots, y_{L,1}, \dots, y_{1,n}, \dots, y_{L,n})$.

It will be convenient to choose an ordering on the space $[L] \times [n]$. We order bits first by the second coordinate and then the first, so that $(r, l) \preceq (s, m)$ if $l \leq m$ or if $r \leq s$ and $l = m$. For any $1 \leq r \leq L$ and $1 \leq m \leq n$ define

$$[\leq (r, m)] := (\{1, \dots, L\} \times \{1, \dots, m-1\}) \cup (\{1, \dots, r\} \times \{m\}) \quad (68)$$

$$[> (r, m)] := (\{r+1, \dots, L\} \times \{m\}) \cup (\{1, \dots, L\} \times \{m+1, \dots, n\}). \quad (69)$$

The canonical path $\gamma_{\mathbf{x}, \mathbf{y}}$ from $\mathbf{x}$ to $\mathbf{y}$ consists of simply replacing each bit of $\mathbf{x}$ with the corresponding bit of $\mathbf{y}$ following the above ordering. This can be thought of as n steps, each of which consists of L substeps that correspond to transitions of the Markov chain. The r -th substep of the m -th step consists of flipping the bit in the r -th Trotter slice of the m -th qubit if $x_{r,m} \neq y_{r,m}$ or leaving it alone if $x_{r,m} = y_{r,m}$. In the former case, this contributes an edge $(\mathbf{v}, \mathbf{v}') \in E(\Omega)$ with

$$\mathbf{v} = (y_{1,1}, \dots, y_{r-1,m}, x_{r,m}, x_{r+1,m}, \dots, x_{L,n}) \quad (70)$$

$$\mathbf{v}' = (y_{1,1}, \dots, y_{r-1,m}, y_{r,m}, x_{r+1,m}, \dots, x_{L,n}). \quad (71)$$

In the latter case (i.e. $x_{r,m} = y_{r,m}$), this step does not add an edge to the path. Thus the number of edges in $\gamma_{\mathbf{x}, \mathbf{y}}$ equals the number of positions on which $\mathbf{x}$ and $\mathbf{y}$ disagree, which is always $\leq nL$. In Section 4.1 we defined an *encoding* of a path to be an injective map η_e from the set of paths through e to the state space Ω . The encoding we'll use for this path is

$$\eta_{(\mathbf{v}, \mathbf{v}')}(\gamma_{\mathbf{x}, \mathbf{y}}) := \mathbf{x} \cdot \mathbf{y} \cdot \mathbf{v}, \quad (72)$$

where $\cdot$ to denote elementwise multiplication. To see that $\mathbf{x}, \mathbf{y}$ are uniquely determined by $\mathbf{v}$ together with $\boldsymbol{\eta} = \eta_{(\mathbf{v}, \mathbf{v}')}(\gamma_{\mathbf{x}, \mathbf{y}})$, observe that $\mathbf{v} \cdot \boldsymbol{\eta} = \mathbf{x} \cdot \mathbf{y}$, which specifies the bits that are flipped along the path $\gamma_{\mathbf{x}, \mathbf{y}}$. Moreover, the edge $(\mathbf{v}, \mathbf{v}')$ specifies the location (r, m) of the bit being flipped in the transition from $\mathbf{v}$ to $\mathbf{v}'$. Let $(\mathbf{v} \cdot \boldsymbol{\eta})_{\leq (r, m)}$ denote the vector which equals $\mathbf{v} \cdot \boldsymbol{\eta} = \mathbf{x} \cdot \mathbf{y}$ on the coordinates in $[\leq (r, m)]$ and equals 1 on the coordinates in $[> (r, m)]$; define $(\mathbf{v} \cdot \boldsymbol{\eta})_{> (r, m)}$ analogously. Then we can recover $\mathbf{x}$ by calculating $\mathbf{v}' \cdot (\mathbf{v} \cdot \boldsymbol{\eta})_{\leq (r, m)}$, and can recover $\mathbf{y}$ by calculating $\mathbf{v}' \cdot (\mathbf{v} \cdot \boldsymbol{\eta})_{> (r, m)}$.

For any subset of the lattice sites $A \subseteq \Lambda$, define the restriction E_β^A of the classical energy function E_β by restricting the sum in (9) to sites $(i, j) \in A$, so that for all $\mathbf{z} \in \Omega$ we have

$$E_\beta^A(\mathbf{z}) = \sum_{(i,j) \in A} \left(\frac{K_j^z}{L} z_{i,j} - \beta^{-1} J_i z_{i,j} z_{i+1,j} + \sum_{k \in A} \frac{K_{jk}^{zz}}{L} z_{i,j} z_{i,k} \right). \quad (73)$$

Now we compute

$$E_\beta(\mathbf{v}) = E_\beta^{\leq (r, m)}(\mathbf{y}) + E_\beta^{> (r, m)}(\mathbf{x}) + B \quad (74)$$

$$E_\beta(\boldsymbol{\eta}) = E_\beta^{\leq (r, m)}(\mathbf{x}) + E_\beta^{> (r, m)}(\mathbf{y}) + B' \quad (75)$$

where

$$\begin{aligned}
B &:= \beta^{-1} J_m (y_{1,m} x_{L,m} + y_{r,m} x_{r+1,m}) + \frac{1}{L} \sum_{i=1}^r K_m^z y_{i,m} + \frac{1}{L} \sum_{i=r}^L K_m^z x_{i,m} \\
&+ \frac{1}{L} \sum_{i=1}^r K_{m,m+1}^{zz} y_{i,m} x_{i,m+1} + \frac{1}{L} \sum_{i=r}^L K_{m-1,m}^{zz} y_{i,m-1} x_{i,m} \\
&+ \frac{1}{L} \sum_{i=1}^L \sum_{j=1}^{m-1} \sum_{k=m}^n K_{jk}^{zz} y_j x_k, \\
B' &:= \beta^{-1} J_m (x_{1,m} x_{L,m} + x_{r,m} x_{r+1,m}) + \frac{1}{L} \sum_{i=1}^r K_m^z x_{i,m} + \frac{1}{L} \sum_{i=r}^L K_m^z y_{i,m} \\
&+ \frac{1}{L} \sum_{i=1}^r K_{m,m+1}^{zz} x_{i,m} y_{i,m+1} + \frac{1}{L} \sum_{i=r}^L K_{m-1,m}^{zz} x_{i,m-1} y_{i,m} \\
&+ \frac{1}{L} \sum_{i=1}^L \sum_{j=1}^{m-1} \sum_{k=m}^n K_{jk}^{zz} x_j y_k.
\end{aligned}$$

We also need

$$E_\beta(\mathbf{x}) = E_\beta^{\leq(r,m)}(\mathbf{x}) + E_\beta^{>(r,m)}(\mathbf{x}) + B_X \quad (76)$$

$$E_\beta(\mathbf{y}) = E_\beta^{\leq(r,m)}(\mathbf{y}) + E_\beta^{>(r,m)}(\mathbf{y}) + B_Y \quad (77)$$

where

$$\begin{aligned}
B_X &:= \sum_{i=1}^L \left(\frac{K_m^z}{L} x_{i,m} + \beta^{-1} J_m x_{i,m} x_{i+1,m} \right) \\
&+ \frac{1}{L} \sum_{i=1}^r K_{m,m+1}^{zz} x_{i,m} x_{i,m+1} + \frac{1}{L} \sum_{i=r}^L K_{m-1,m}^{zz} x_{i,m-1} x_{i,m} \\
&+ \frac{1}{L} \sum_{i=1}^L \sum_{j=1}^{m-1} \sum_{k=m}^n K_{jk}^{zz} x_j x_k,
\end{aligned}$$

and B_Y is defined similarly. The value of M needed to satisfy (60) for this encoding can now be determined by

$$\frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{v})\pi(\boldsymbol{\eta})} = e^{-\beta(E_\beta(\mathbf{x})+E_\beta(\mathbf{y})-E_\beta(\mathbf{v})-E_\beta(\boldsymbol{\eta}))} = e^{\beta(B+B'-B_X-B_Y)} \quad (78)$$

To upper bound the exponent in the worst-case define $J := \max_{m=1,\dots,n} J_m$, then

$$\beta(B+B'-B_X-B_Y) \leq 8J + 4\beta + 4\beta \sum_{j=1}^{n/2} \sum_{k=\frac{n}{2}+1}^n K_{jk}^{zz}.$$

To upper bound the sum we use the assumption $K_{jk}^{zz} \leq |j-k|^{-(2+\xi)}$ and regard it as a lower Riemann sum for an integral,

$$\sum_{j=1}^{n/2} \sum_{k=\frac{n}{2}+1}^n |j-k|^{-(2+\xi)} \leq \int_{j=1}^{j=\frac{n}{2}} \int_{k=\frac{n}{2}+1}^{k=n} dj dk |j-k|^{-(2+\xi)} = \frac{1 - 2^{\xi+1} n^{-\xi} + (n-1)^{-\xi}}{\xi^2 + \xi},$$

and since $\xi > 0$ it suffices in (63) to take $M = \mathcal{O}(e^{8J+4\beta})$. If the 1D quantum system has periodic boundary conditions connecting qubit 1 to qubit n , then the same analysis would show that $M = \mathcal{O}(e^{8(J+\beta)})$ suffices for (60). Therefore by (61)-(63) the congestion R satisfies

$$R \leq M \cdot P_{\min}^{-1} = \mathcal{O}\left(e^{8(\beta+J)} \cdot 2nL \cdot e^{2(J+\frac{\beta}{L})} \cdot \max_{(\mathbf{z}, \mathbf{z}')} \left(\frac{\pi(\mathbf{z}')}{\pi(\mathbf{z})}\right)\right)$$

the factor of $2nL$ comes from the chain being lazy and the choice of random single-site updates. Note that $\max_{(\mathbf{z}, \mathbf{z}')} (\pi(\mathbf{z}')/\pi(\mathbf{z}))$ is $\mathcal{O}(L^2\Gamma^{-2})$ from (67) and also $\ell := \max_{\mathbf{x}, \mathbf{y}} |\gamma_{\mathbf{x}, \mathbf{y}}| = nL$ and $\ln \pi_{\min}^{-1} = \ln(Z \cdot e^{nLJ+n\beta}) \leq nL(J + \beta + \ln 2)$, using $Z \leq 2^{nL}$. In Section 3.1 we show that $L = \text{poly}(n, \delta_{\text{mult}}^{-1})$ suffices to approximate the partition function with the desired precision. Therefore the mixing time bounds in Section 4.1 imply that the PIMC method generates samples as was assumed in Section 3.3, which completes the proof of theorem 1 for models of the form (5).

4.4 Polynomial time mixing in theorem 1

In this section we will prove rapid mixing of PIMC for Hamiltonians of the form (6) by describing paths between arbitrary states $\mathbf{x}, \mathbf{y} \in \Omega$. It will be convenient to represent states in the form $\mathbf{z} = [\bar{z}_1, \dots, \bar{z}_n]$ with $\bar{z}_j := (z_{1,j}, \dots, z_{L,j})$ denoting the spin values along the j -th worldline in configuration $\mathbf{z} \in \Omega$. We say that a “double jump” occurs in configuration $\mathbf{z}$ at position (i, j) if $(1 + z_{i,j}z_{i+1,j})(1 + z_{i,j+1}z_{i+1,j+1}) \neq 0$. For any consecutive pair of worldlines $j, j+1$, define the number of double jumps $d(\bar{z}_j, \bar{z}_{j+1})$,

$$d(\bar{z}_j, \bar{z}_{j+1}) := \frac{1}{4} \sum_{i=1}^L (1 + z_{i,j}z_{i+1,j})(1 + z_{i,j+1}z_{i+1,j+1}). \quad (79)$$

For any configuration $\mathbf{z}$ and worldline j we will define a path segment to a new configuration $[\bar{z}_1, \dots, \bar{z}_{j-2}, \bar{z}'_{j-1}, \bar{z}''_j, \bar{z}'_{j+1}, \bar{z}_{j+2}, \dots, \bar{z}_n]$ with all of the double jumps removed from worldline j ,

$$d(\bar{z}'_{j-1}, \bar{z}''_j) = d(\bar{z}''_j, \bar{z}'_{j+1}) = 0 \quad (80)$$

The double prime notation indicates that the j -th worldline does not contain any double jumps, while the single prime indicates a worldline that shares no jumps with its double primed neighboring worldline.

There are many possible paths to choose from for removing double jumps in the i -th worldline. Due to the form of the Hamiltonian (6), changing the placement of a double jump along the imaginary-time direction has no effect on the energy of the effective classical system, except possibly for the part which arises from the diagonal part of the Hamiltonian. Similar reasoning holds for the ordering of the single and double jumps; changing the ordering can only change the effective classical energy through the diagonal part of the Hamiltonian. Similarly, changing the imaginary-time position of the double jumps in $\bar{x}_{i-1}, \bar{x}_i$ will have no effect on the contributions to the effective classical energy that comes from double jumps occurring in $\bar{x}_{i-2}, \bar{x}_{i-1}$. Finally, note that changing the position of a double jump along the imaginary-time direction can only affect the classical energy through the diagonal part of the Hamiltonian, reducing the probability of the configuration by at most $e^{-4\beta}$ (the factor of 4 is due to the potential change from the diagonal part of the Hamiltonian when 3 worldlines are changed arbitrarily).

From the above considerations it follows that the imaginary-time position of the double jumps in a particular worldline can be changed freely with the stationary weight of the resulting configuration decreased by at most a factor of $e^{-4\beta}$ (the factor of 4 bounds the potential change from the diagonal part of the Hamiltonian when 3 worldlines are changed arbitrarily). Furthermore, since double jumps arising from the term $\sigma_i^x \sigma_{i+1}^x$ are simply bit-flips, two such jumps will cancel out (“annihilate”) if they coincide. This allows for paths that remove these double jumps in pairs simply by moving them around, with the additional guarantee that this will not lower the stationary weight by more than $e^{-4\beta}$. Therefore the path from $[\bar{z}_1, \dots, \bar{z}_n]$ to $[\bar{z}_1, \dots, \bar{z}_{j-2}, \bar{z}'_{j-1}, \bar{z}''_j, \bar{z}'_{j+1}, \bar{z}_{j+2}, \dots, \bar{z}_n]$ is defined to remove double jumps in pairs by shifting the position of the double jump at the least imaginary-time position along the imaginary-time direction until it encounters the next double jump and the pair is annihilated.

The canonical path from $\mathbf{x}$ to $\mathbf{y}$ proceeds through the worldlines in order $1, \dots, n$, prepping each pair of consecutive worldlines to remove double jumps, then updating the j -th worldline from $\bar{x}''_j$ to $\bar{y}''_j$. The full update of the first worldline proceeds as,

$$\begin{aligned} [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n] &\rightarrow [\bar{x}_1'', \bar{x}_2', \bar{x}_3, \dots, \bar{x}_n] \rightarrow [\bar{y}_1'', \bar{x}_2', \bar{x}_3, \dots, \bar{x}_n] \rightarrow [\bar{y}_1'', \bar{x}_2'', \bar{x}_3', \bar{x}_4, \dots, \bar{x}_n] \\ &\rightarrow [\bar{y}_1'', \bar{y}_2'', \bar{x}_3', \bar{x}_4, \dots, \bar{x}_n] \rightarrow [\bar{y}_1, \bar{y}_2', \bar{x}_3', \bar{x}_4, \dots, \bar{x}_n]. \end{aligned}$$

The purpose of this series of steps (each of which consists of many single-site updates) is to avoid passing through intermediate configurations with too many additional double jumps occurring across the 1D domain wall formed near the worldlines which are being updated. The worst case intermediate configurations along the paths will have the form,

$$\mathbf{z} = [\bar{y}_1, \dots, \bar{y}_{j-2}, \bar{y}'_{j-1}, (y''_{1,j}, \dots, y''_{r,j}, x''_{r+1,j}, \dots, x''_{L,j}), \bar{x}'_{j+1}, \bar{x}_{j+2}, \dots, \bar{x}_n]. \quad (81)$$

This system of paths is highly degenerate (a particular configuration such as (81) will appear in many paths). To bound the congestion we will use an encoding function $\eta_{(\mathbf{z}, \mathbf{z}')} : \mathcal{P}(\mathbf{z}, \mathbf{z}') \rightarrow \Omega \times \Theta$. If $\eta_{(\mathbf{z}, \mathbf{z}')}(\gamma_{\mathbf{x}\mathbf{y}}) = (\eta, \theta)$, then

$$\eta = [\bar{x}_1, \dots, \bar{x}_{j-2}, \bar{x}'_{j-1}, (x''_{1,j}, \dots, x''_{r,j}, y''_{r+1,j}, \dots, y''_{L,j}), \bar{y}'_{j+1}, \bar{y}_{j+2}, \dots, \bar{y}_n]. \quad (82)$$

and θ contains all of the information needed to reconstruct $\bar{x}_{j-1}, \bar{x}_j, \bar{x}_{j+1}, \bar{y}_{j-1}, \bar{y}_j, \bar{y}_{j+1}$ from $\bar{x}'_{j-1}, \bar{x}''_j, \bar{x}'_{j+1}, \bar{y}'_{j-1}, \bar{y}''_j, \bar{y}'_{j+1}$. This lost information must specify the location of all of the jumps that have been removed from these 6 worldlines. This means that the encoding sends exponentially paths to each state, and so to obtain a polynomial time mixing bound we have to slightly generalize (63) to account for the fact that the stationary probability on the endpoints of most of these paths is sufficiently small.

For any $\mathbf{v} \in \Omega$ define $d_j(\mathbf{v}) := d(\bar{v}_{j-1}, \bar{v}_j) + d(\bar{v}_j, \bar{v}_{j+1})$ (which is just the total number of double jumps in the j -th worldline of $\mathbf{v}$). Let q be the number of additional double jumps in the j -th worldlines of $\mathbf{x}$ and $\mathbf{y}$ which are not present in the j -th worldlines of $\mathbf{z}$ and $\boldsymbol{\eta}$,

$$q = d_j(\mathbf{x}) + d_j(\mathbf{y}) - d_j(\mathbf{z}) - d_j(\boldsymbol{\eta}) \quad (83)$$

Define $\mathcal{P}_{\mathbf{z}, \mathbf{z}'}^q$ to be the subset of canonical paths passing through $(\mathbf{z}, \mathbf{z}')$ with endpoints $\mathbf{x}$ and $\mathbf{y}$ that satisfy (83). With these definitions the congestion (59) becomes

$$R \leq \max_{(\mathbf{z}, \mathbf{z}')} \frac{1}{\pi(\mathbf{z})P(\mathbf{z}, \mathbf{z}')} \sum_{q=0}^L \sum_{\gamma_{\mathbf{x}\mathbf{y}} \in \mathcal{P}_{\mathbf{z}, \mathbf{z}'}^q} \pi(\mathbf{x})\pi(\mathbf{y}), \quad (84)$$

Let $\Theta = \cup_{q=1}^{6L} \Theta_q$ where $G(q) = |\Theta_q| = \binom{6L}{q}$ is the number of ways to restore q jumps to the undetermined worldlines. Now suppose we find $M(q)$ such that

$$\pi(\mathbf{x})\pi(\mathbf{y}) \leq M(q) \cdot \pi(\mathbf{z})\pi(\boldsymbol{\eta}), \quad (85)$$

for all $\gamma_{\mathbf{xy}} \in \mathcal{P}_{\mathbf{z}, \mathbf{z}'}^q$, then this implies

$$R \leq P_{\min}^{-1} \sum_{q=0}^L M(q) \sum_{\gamma_{\mathbf{xy}} \in \mathcal{P}_{\mathbf{z}, \mathbf{z}'}^q} \pi(\boldsymbol{\eta}) \quad (86)$$

$$\leq P_{\min}^{-1} \sum_{q=0}^L M(q) G(q) \sum_{\boldsymbol{\eta} \in \Omega} \pi(\boldsymbol{\eta}) \quad (87)$$

$$= P_{\min}^{-1} \sum_{q=0}^L M(q) G(q) \quad (88)$$

To satisfy (85) we need $M(q)$ to satisfy

$$M(q) \geq \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})\pi(\boldsymbol{\eta})} \quad (89)$$

$$= \frac{\pi([\bar{x}_{j-2}, \bar{x}_{j-1}, \bar{x}_j, \bar{x}_{j+1}, \bar{x}_{j+2}])\pi([\bar{y}_{j-2}, \bar{y}_{j-1}, \bar{y}_j, \bar{y}_{j+1}, \bar{y}_{j+2}])}{\pi([\bar{y}_{j-2}, \bar{y}'_{j-1}, \bar{x}'_j + \bar{y}'_j, \bar{x}'_{j+1}, \bar{x}_{j+2}])\pi([\bar{x}_{j-2}, \bar{x}'_{j-1}, \bar{x}'_j + \bar{y}'_j, \bar{y}'_{j+1}, \bar{y}_{j+2}])} \quad (90)$$

$$\geq e^{4\beta} \frac{\pi([\bar{x}_{j-1}, \bar{x}_j, \bar{x}_{j+1}])\pi([\bar{y}_{j-1}, \bar{y}_j, \bar{y}_{j+1}])}{\pi([\bar{y}'_j, \bar{x}'_j + \bar{y}'_j, \bar{x}'_{j+1}])\pi([\bar{x}'_{j-1}, \bar{x}'_j + \bar{y}'_j, \bar{y}'_{j+1}])} \quad (91)$$

where (90) to (91) follows because the terms in the classical energy function which correspond to off-diagonal parts of the Hamiltonian involving qubits $j-2$ and $j+2$ are unaffected by the potentially changed spin values in the primed $j-1$ and $j+1$ worldlines. The factor of $e^{4\beta}$ comes from upper bounding the change in the classical energy function due to the diagonal terms of the Hamiltonian between qubits $j-2, j-1$ and $j+1, j+2$. The effect of the diagonal part of the Hamiltonian on the ratio expression remaining in (91) can be bounded by a factor of $e^{8\beta}$. The additional single jumps introduced near the domain wall can increase the ratio by at most L^{12} (using (67), and noting that the worst case occurs when 2 new single jumps are introduced in each of 3 worldlines, in each of $\mathbf{z}$ and $\boldsymbol{\eta}$). Each of the double jumps in the numerator which are not present in the denominator decreases the ratio by at least $(2\beta/L)$. Therefore we may take $M(q) = e^{12\beta}(2\beta)^q L^{-q}$ and (88) becomes

$$R \leq e^{12\beta} L^{12} \cdot P_{\min}^{-1} \sum_{q=0}^L \binom{6L}{q} \left(\frac{2\beta}{L}\right)^q \leq L^{12} \cdot e^{20\beta} \cdot P_{\min}^{-1}. \quad (92)$$

Using the fact that $P_{\min}^{-1}$ is $\text{poly}(n, \beta, \Gamma^{-1})$ from (67), and the fact that the maximum length of the canonical paths used in this section is $\mathcal{O}(n^2 L^2)$, this shows that τ_{mix} is $\text{poly}(n, e^\beta, \Gamma^{-1})$ as required to complete the proof of theorem 1.

4.5 Quasipolynomial time mixing in theorem 1

In this section we also work with a restricted state space Ω^* , which is defined as the set of configurations with fewer than $2B := 4\beta \log(n)$ jumps in each of the worldlines,

$$\Omega^* = \{\mathbf{z} = [\bar{z}_1, \dots, \bar{z}_n] \in \Omega : |\{i : z_{i,j} \neq z_{i+1,j}\}| \leq 2B \text{ for } j = 1, \dots, n\}. \quad (93)$$

In Section 3.4 we show that $Z^* = \sum_{\mathbf{z} \in \Omega^*} e^{-\beta E_\beta(\mathbf{z})}$ satisfies $|Z - Z^*| \leq \delta Z$. The PIMC dynamics can be restricted to Ω^* by rejecting moves which would leave Ω^* , or alternatively the canonical paths in this section bound the mixing time within Ω^* by the leaky random walk conductance bound in [15]. To show rapid mixing of PIMC within Ω^* we assign a canonical path $\gamma_{\mathbf{x}\mathbf{y}}$ to every $\mathbf{x}, \mathbf{y} \in \Omega^*$. The simple case for these paths occurs when $\mathbf{x}, \mathbf{y}$ are in a subset $\Omega_{\text{inner}}^* \subseteq \Omega^*$ which only allows half as many jumps per worldline as Ω^* does,

$$\Omega_{\text{inner}}^* = \{\mathbf{z} = [z_1, \dots, z_n] \in \Omega : |\{i : z_{i,j} \neq z_{i+1,j}\}| \leq B \text{ for } j = 1, \dots, n\}. \quad (94)$$

If $\mathbf{x}, \mathbf{y} \in \Omega_{\text{inner}}^*$ then we use the same paths which were used to show rapid mixing for TIM in Section 4.3. Various intermediate points $\mathbf{v}$ along this path $\gamma_{\mathbf{x},\mathbf{y}}$ may leave Ω_{inner}^* but will remain in Ω^* since

$$\mathbf{v} = (y_{1,1}, \dots, y_{L,1}, \dots, y_{L-r-1,m}, x_{L-r,m}, x_{L-r+1,m}, \dots, x_{1,n}, \dots, x_{L,n}), \quad (95)$$

may have as many as B jumps in the first r time slices of $\bar{y}_m$ and B jumps in the $L - r$ latter time slices of $\bar{x}_m$, so $\bar{v}_m$ will have at most $2B$ jumps in total and so $\mathbf{v} \in \Omega^*$.

For points $\mathbf{x}, \mathbf{y} \in \Omega^*$ which are not in Ω_{inner}^* , we must not produce intermediate points with too many jumps in any worldline. To accomplish this the paths will have a “clean up” step similar to the one which was used in the previous section, to reduce the number of 2-local jumps in a worldline. Consider an intermediate point of the path which has so far updated the worldlines of the first m qubits (which could be $m = 0$), $\mathbf{v} = [\bar{y}_1, \dots, \bar{y}_m, \bar{x}_{m+1}, \dots, \bar{x}_n]$. If $\bar{x}_{m+1}$ has more than B double jumps then we remove these in imaginary-time order. Since we can no longer “annihilate” these transitions in pairs as was done in the previous section, we instead dismantle them into 1-local jumps (relying on the fictitious transverse field) and then annihilate those against one another. The effect of this process on the stationary weight of the configuration will be analyzed after we have given the encoding function for these paths. The net result of this process results is a worldline $\bar{x}'_{m+1}$ with fewer than B jumps. Similarly let $\bar{y}'_{m+1}$ be defined from $\bar{y}_{m+1}$ with all but the first B jumps removed. Next we update $\bar{x}'_{m+1}$ to $\bar{y}'_{m+1}$ in the usual way, and since $\bar{x}'_{m+1}$ and $\bar{y}'_{m+1}$ each have fewer than B jumps, the intermediate configurations along this path will have less than $2B$ jumps in every worldline and therefore remain in Ω^* .

The encoding $\eta : \mathcal{P}_{(\mathbf{z}, \mathbf{z}')} \rightarrow \Omega^* \times \Theta$ for these paths is again defined by (82). This time the maximum degeneracy of the encoding is $\binom{4L}{B}$, since there are at most $4L$ positions to which at most B 2-local jumps can be restored in the undetermined worldlines. Unfortunately, this time we cannot have a inequality analogous to (85), because the additional double jumps present in $\mathbf{x}$ and $\mathbf{y}$ can completely change the bit-values in wordlines $j - 1$ and $j + 1$. Unlike the case in the previous section these altered spin values can reduce the classical energy drastically because of the h terms acting on worldlines $j - 2, j - 1$ and $j + 1, j + 2$. The stationary weight of $\mathbf{z}$ and $\boldsymbol{\eta}$ can be lowered from $\mathbf{x}$ and $\mathbf{y}$ by a factor which scales exponentially in the number of double jumps that $\mathbf{x}$ and $\mathbf{y}$

have in worldlines $j-2, j-1$ and $j+1, j+2$. Therefore,

$$M \geq \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})\pi(\boldsymbol{\eta})} \quad (96)$$

$$= \frac{\pi([\bar{x}_{i-2}, \bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1}, \bar{x}_{i+2}])\pi([\bar{y}_{i-2}, \bar{y}_{i-1}, \bar{y}_i, \bar{y}_{i+1}, \bar{y}_{i+2}])}{\pi([\bar{y}'_{i-2}, \bar{y}'_{i-1}, \bar{x}'_i + \bar{y}'_i, \bar{x}'_{i+1}, \bar{x}_{i+2}])\pi([\bar{x}_{i-2}, \bar{x}'_{i-1}, \bar{x}'_i + \bar{y}'_i, \bar{y}'_{i+1}, \bar{y}_{i+2}])} \quad (97)$$

$$\geq e^{4\beta} \cdot \left(\frac{\beta\Gamma}{L}\right)^{-4B} \frac{\pi([\bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1}])\pi([\bar{y}_{i-1}, \bar{y}_i, \bar{y}_{i+1}])}{\pi([\bar{y}'_{i-1}, \bar{x}'_i + \bar{y}'_i, \bar{x}'_{i+1}])\pi([\bar{x}'_{i-1}, \bar{x}'_i + \bar{y}'_i, \bar{y}'_{i+1}])} \quad (98)$$

$$\geq e^{12\beta} \cdot \left(\frac{\beta\Gamma}{L}\right)^{-8B}, \quad (99)$$

The fictitious transverse field has magnitude δ_{mult}/n , so the congestion satisfies

$$R \leq P_{\min}^{-1} M \cdot G \quad (100)$$

$$\leq P_{\min}^{-1} \cdot e^{12\beta} \cdot \left(\text{poly}(n) \cdot \delta_{\text{mult}}^{-1}\right)^{8\beta\Gamma \log n} \cdot \binom{4L}{4\beta\Gamma \log n}, \quad (101)$$

from which it follows that the mixing time τ_{mix} is $\mathcal{O}(2^{\beta(\log n \delta_{\text{mult}}^{-1})^2})$ and we obtain the overall PIMC runtime as claimed in theorem 1.

Acknowledgments

Part of this work was completed while EC was funded by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). AWH was funded by NSF grants CCF-1452616, CCF-1629809, CCF-1729369 and ARO contract W911NF-17-1-0433. The research is based upon work partially supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the U.S. Army Research Office contract W911NF-17-C-0050. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

- [1] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90: 015002, Jan 2018. DOI: [10.1103/RevModPhys.90.015002](https://doi.org/10.1103/RevModPhys.90.015002).
- [2] Tameem Albash, Gene Wagenbreth, and Itay Hen. Off-diagonal expansion quantum monte carlo. *Physical Review E*, 96(6):063309, 2017. DOI: [10.1103/PhysRevE.96.063309](https://doi.org/10.1103/PhysRevE.96.063309).
- [3] Evgeny Andriyash and Mohammad H Amin. Can quantum Monte Carlo simulate quantum annealing?, 2017, [arXiv:1703.09277](https://arxiv.org/abs/1703.09277).
- [4] Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick. Rigorous RG algorithms and area laws for low energy eigenstates in 1D. *Communications in Mathematical Physics*, 356(1): 65–105, 2017. DOI: [10.1007/s00220-017-2973-z](https://doi.org/10.1007/s00220-017-2973-z).

- [5] Ivona Bezáková, Daniel Štefankovič, Vijay V. Vazirani, and Eric Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. *SIAM Journal on Computing*, 37(5):1429–1454, 2008. DOI: [10.1137/050644033](https://doi.org/10.1137/050644033).
- [6] Lucas T Brady and Wim van Dam. Quantum Monte Carlo simulations of tunneling in quantum adiabatic optimization. *Physical Review A*, 93(3):032304, 2016, [arXiv:1509.02562](https://arxiv.org/abs/1509.02562). DOI: [10.1103/PhysRevA.93.032304](https://doi.org/10.1103/PhysRevA.93.032304).
- [7] Lucas T Brady and Wim van Dam. Spectral-gap analysis for efficient tunneling in quantum adiabatic optimization. *Physical Review A*, 94(3):032309, 2016, [arXiv:1601.01720](https://arxiv.org/abs/1601.01720). DOI: [10.1103/PhysRevA.94.032309](https://doi.org/10.1103/PhysRevA.94.032309).
- [8] Sergey Bravyi. Monte Carlo simulation of stoquastic Hamiltonians. *Quantum Information & Computation*, 15(13-14):1122–1140, 2015, [arXiv:1402.2295](https://arxiv.org/abs/1402.2295). DOI: [10.5555/2871363.2871366](https://doi.org/10.5555/2871363.2871366).
- [9] Sergey Bravyi and David Gosset. Polynomial-time classical simulation of quantum ferromagnets. *Physical review letters*, 119(10):100503, 2017, [arXiv:1612.05602](https://arxiv.org/abs/1612.05602). DOI: [10.1103/PhysRevLett.119.100503](https://doi.org/10.1103/PhysRevLett.119.100503).
- [10] Sergey Bravyi and Matthew Hastings. On complexity of the quantum Ising model. *Communications in Mathematical Physics*, 349(1):1–45, 2017, [arXiv:1410.0703](https://arxiv.org/abs/1410.0703). DOI: [10.1007/s00220-016-2787-4](https://doi.org/10.1007/s00220-016-2787-4).
- [11] Sergey Bravyi and Barbara M. Terhal. Complexity of stoquastic frustration-free Hamiltonians. *SIAM J. Comput.*, 39(4):1462–1485, 2009, [arXiv:0806.1746](https://arxiv.org/abs/0806.1746). DOI: [10.1137/08072689X](https://doi.org/10.1137/08072689X).
- [12] Sergey Bravyi, David P. DiVincenzo, Roberto I. Oliveira, and Barbara M. Terhal. The complexity of stoquastic local Hamiltonian problems. *Quant. Inf. Comp.*, 8(5):0361–0385, 2006, [arXiv:quant-ph/0606140](https://arxiv.org/abs/quant-ph/0606140). DOI: [10.5555/2011772.2011773](https://doi.org/10.5555/2011772.2011773).
- [13] Jacob Bringewatt, William Dorland, Stephen P. Jordan, and Alan Mink. Diffusion monte carlo approach versus adiabatic computation for local hamiltonians. *Phys. Rev. A*, 97:022323, Feb 2018. DOI: [10.1103/PhysRevA.97.022323](https://doi.org/10.1103/PhysRevA.97.022323).
- [14] Alessandra Cipriani and Paolo Dai Pra. Decay of correlations for quantum spin systems with a transverse field: A dynamic approach, 2010, [arXiv:1005.3547](https://arxiv.org/abs/1005.3547).
- [15] Elizabeth Crosson and Aram W Harrow. Simulated quantum annealing can be exponentially faster than classical simulated annealing. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 714–723. IEEE, 2016, [arXiv:1601.03030](https://arxiv.org/abs/1601.03030). DOI: [10.1109/FOCS.2016.81](https://doi.org/10.1109/FOCS.2016.81).
- [16] Elizabeth Crosson, Edward Farhi, Cedric Yen-Yu Lin, Han-Hsuan Lin, and Peter Shor. Different strategies for optimization using the quantum adiabatic algorithm, 2014, [arXiv:1401.7320](https://arxiv.org/abs/1401.7320).
- [17] Toby S. Cubitt, Ashley Montanaro, and Stephen Piddock. Universal quantum hamiltonians. *Proceedings of the National Academy of Sciences*, 115(38):9497–9502, 2018, [arXiv:1701.05182](https://arxiv.org/abs/1701.05182). ISSN 0027-8424. DOI: [10.1073/pnas.1804949115](https://doi.org/10.1073/pnas.1804949115). URL <https://www.pnas.org/content/115/38/9497>.
- [18] Ky Fan. Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proceedings of the National Academy of Sciences*, 37(11):760–766, 1951. ISSN 0027-8424. DOI: [10.1073/pnas.37.11.760](https://doi.org/10.1073/pnas.37.11.760).
- [19] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001. DOI: [10.1126/science.1057726](https://doi.org/10.1126/science.1057726).

- [20] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. Quantum adiabatic evolution algorithms versus simulated annealing, 2002, [arXiv:quant-ph/0201031](#).
- [21] WMC Foulkes, L Mitas, RJ Needs, and G Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33, 2001. DOI: [10.1103/RevModPhys.73.33](#).
- [22] M. B. Hastings. Quantum adiabatic computation with a constant gap is not useful in one dimension. *Phys. Rev. Lett.*, 103:050502, 2009, [arXiv:0902.2960](#). DOI: [10.1103/PhysRevLett.103.050502](#).
- [23] Matthew B Hastings and MH Freedman. Obstructions to classically simulating the quantum adiabatic algorithm. *Quantum Information & Computation*, 13(11-12):1038–1076, 2013, [arXiv:1302.5733](#). DOI: [10.5555/2535639.2535647](#).
- [24] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. DOI: [10.1080/01621459.1963.10500830](#).
- [25] Layla Hormozi, Ethan W. Brown, Giuseppe Carleo, and Matthias Troyer. Nonstoquastic hamiltonians and quantum annealing of an ising spin glass. *Phys. Rev. B*, 95:184416, May 2017. DOI: [10.1103/PhysRevB.95.184416](#).
- [26] Michael Jarret and Brad Lackey. Substochastic Monte Carlo algorithms, 2017, [arXiv:1704.09014](#).
- [27] Michael Jarret, Stephen P Jordan, and Brad Lackey. Adiabatic optimization versus diffusion monte carlo methods. *Physical Review A*, 94(4), 2016, [arXiv:1607.03389](#). DOI: [10.1103/PhysRevA.94.042318](#).
- [28] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993. DOI: [10.1137/0222066](#).
- [29] Mark Jerrum and Alistair Sinclair. The Markov chain Monte Carlo method: An approach to approximate counting and integration. In Dorit S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 482–520. PWS Publishing Co., Boston, MA, USA, 1997. ISBN 0-534-94968-1. DOI: [10.5555/241938.241950](#).
- [30] Zhang Jiang, Vadim N Smelyanskiy, Sergei V Isakov, Sergio Boixo, Guglielmo Mazzola, Matthias Troyer, and Hartmut Neven. Scaling analysis and instantons for thermally assisted tunneling and quantum Monte Carlo simulations. *Physical Review A*, 95(1):012322, 2017, [arXiv:1603.01293](#). DOI: [10.1103/PhysRevA.95.012322](#).
- [31] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse Ising model. *Phys. Rev. E*, 58:5355–5363, Nov 1998, [arXiv:cond-mat/9804280](#). DOI: [10.1103/PhysRevE.58.5355](#).
- [32] Linghang Kong and Elizabeth Crosson. The performance of the quantum adiabatic algorithm on spike Hamiltonians. *Int. J. Quantum Inform.*, 15(1750011), 2015, [arXiv:1511.06991](#). DOI: [10.1142/S0219749917500113](#).
- [33] Zeph Landau, Umesh Vazirani, and Thomas Vidick. A polynomial-time algorithm for the ground state of 1D gapped local Hamiltonians. *Nature Physics*, 2013, [arXiv:1307.5143](#). DOI: [10.1038/nphys3345](#).
- [34] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Soc., 2008. ISBN 9780821886274. DOI: [10.1090/mbk/107](#).
- [35] Fabio Martinelli and Marc Wouts. Glauber dynamics for the quantum Ising model in a transverse field on a regular tree. *Journal of Statistical Physics*, 146(5):1059–1088, 2012, [arXiv:1105.5970](#). DOI: [10.1007/s10955-012-0436-7](#).

- [36] Roman Martonák, Giuseppe E. Santoro, and Erio Tosatti. Quantum annealing by the path-integral monte carlo method: The two-dimensional random ising model. *Phys. Rev. B*, 66: 094203, Sep 2002. DOI: [10.1103/PhysRevB.66.094203](https://doi.org/10.1103/PhysRevB.66.094203).
- [37] Siddharth Muthukrishnan, Tameem Albash, and Daniel A Lidar. Tunneling and speedup in quantum optimization for permutation-symmetric problems. *Physical Review X*, 6(3):031010, 2016. DOI: [10.1103/PhysRevX.6.031010](https://doi.org/10.1103/PhysRevX.6.031010).
- [38] Ben W Reichardt. The quantum adiabatic optimization algorithm and local minima. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 502–510. ACM, 2004. DOI: [10.1145/1007352.1007428](https://doi.org/10.1145/1007352.1007428).
- [39] Anders W Sandvik. Stochastic series expansion method with operator-loop update. *Physical Review B*, 59(22):R14157, 1999. DOI: [10.1103/PhysRevB.59.R14157](https://doi.org/10.1103/PhysRevB.59.R14157).
- [40] Lorenzo Stella and Giuseppe E Santoro. Quantum annealing of an Ising spin-glass by Green’s function Monte Carlo. *Physical Review E*, 75(3):036703, 2007, [arXiv:cond-mat/0608420](https://arxiv.org/abs/cond-mat/0608420). DOI: [10.1103/PhysRevE.75.036703](https://doi.org/10.1103/PhysRevE.75.036703).
- [41] Yuki Susa, Yu Yamashiro, Masayuki Yamamoto, and Hidetoshi Nishimori. Exponential speedup of quantum annealing by inhomogeneous driving of the transverse field. *Journal of the Physical Society of Japan*, 87(2):023002, 2018. DOI: [10.7566/JPSJ.87.023002](https://doi.org/10.7566/JPSJ.87.023002).
- [42] Masuo Suzuki. Quantum statistical Monte Carlo methods and applications to spin systems. *Journal of Statistical Physics*, 43(5-6):883–909, 1986. ISSN 0022-4715. DOI: [10.1007/BF02628318](https://doi.org/10.1007/BF02628318).
- [43] Masuo Suzuki, Seiji Miyashita, and Akira Kuroda. Monte Carlo simulation of quantum spin systems. *Prog. Theor. Phys.*, 58(5):1377–1387, 1977. DOI: [10.1143/PTP.58.1377](https://doi.org/10.1143/PTP.58.1377).
- [44] S. Suzuki, J. Inoue, and B.K. Chakrabarti. *Quantum Ising Phases and Transitions in Transverse Ising Models*. Lecture notes in physics. Springer, 2013. ISBN 9783642330391. DOI: [10.1007/978-3-642-33039-1](https://doi.org/10.1007/978-3-642-33039-1).
- [45] Frank Verstraete, Valentin Murg, and J Ignacio Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008, [arXiv:0907.2796](https://arxiv.org/abs/0907.2796). DOI: [10.1080/14789940801912366](https://doi.org/10.1080/14789940801912366).
- [46] Walter Vinci and Daniel A Lidar. Non-stoquastic hamiltonians in quantum annealing via geometric phases. *npj Quantum Information*, 3(1):1–6, 2017. DOI: [10.1038/s41534-017-0037-z](https://doi.org/10.1038/s41534-017-0037-z).