Quantum algorithm for approximating partition functions

Pawel Wocjan,[1] Chen-Fu Chiang,[1] Daniel Nagaj,[2,3] and Anura Abeyesinghe[1]
[1]*School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, Florida 32816, USA*
[2]*Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Dúbravská Cesta 9, 84215 Bratislava, Slovakia*
[3]*Quniverse, Líščie Údolie 116, 84104 Bratislava, Slovakia*
(Received 2 June 2009; published 31 August 2009)

We present a quantum algorithm based on classical fully polynomial randomized approximation schemes (FPRASs) for estimating partition functions that combine simulated annealing with the Monte Carlo Markov chain method and use nonadaptive cooling schedules. We achieve a twofold polynomial improvement in time complexity: a quadratic reduction with respect to the spectral gap of the underlying Markov chains and a quadratic reduction with respect to the parameter characterizing the desired accuracy of the estimate output by the FPRAS. Both reductions are intimately related and cannot be achieved separately. First, we use Grover's fixed-point search, quantum walks, and phase estimation to efficiently prepare approximate coherent encodings of stationary distributions of the Markov chains. The speed up we obtain in this way is due to the quadratic relation between the spectral and phase gaps of classical and quantum walks. The second speed up with respect to accuracy comes from generalized quantum counting used instead of classical sampling to estimate expected values of quantum observables.

DOI: 10.1103/PhysRevA.80.022340 PACS number(s): 03.67.Ac

I. INTRODUCTION

Quantization of classical Markov chains has been crucial in the design of efficient quantum algorithms for a wide range of search problems that outperform their classical counterparts. We refer the reader to the survey article [1] for a detailed account of the rapidly growing collection of quantum-walk-based search algorithms. In this context, we also point to the work [2], where the authors apply quantized Markov chains to speed up search algorithms based on simulated annealing for finding low-energy states of (classical) Hamiltonians.

In this paper, we extend the scope of the use of quantized Markov chains beyond search problems. We show how to employ them to speed up fully polynomial-time randomized approximation schemes for partition functions based on simulated annealing and the Monte Carlo Markov chain method. To achieve this improvement, we rely on Szegedy's general method to quantize classical Markov chains [3,4], which we review in the Appendix. This method gives us a unitary *quantum-walk* operator $W(P)$ corresponding to one update step of the classical Markov chain P. The complexity of the classical algorithms we are speeding up is measured in the number of Markov chain invocations. Similarly, we express the complexity of our quantum algorithm as the number of times we have to apply a quantum-walk operator. As shown in [5], in the circuit model of quantum computation, this operator can be implemented precisely and efficiently.

Sampling from stationary distributions of Markov chains combined with simulated annealing is at the heart of many clever classical approximation algorithms. Notable examples include the algorithm for approximating the volume of convex bodies [6], the permanent of a non-negative matrix [7], and the partition function of statistical physics models such as the Ising model [8] and the Potts model [9]. Each of these algorithms is a *fully polynomial randomized approximation scheme* (FPRAS), outputting a random number $\hat{Z}$ within a factor of $(1 \pm \epsilon)$ of the real value Z, with probability greater than $\frac{3}{4}$, i.e.,

$$\mathcal{P}[(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z] \geq \frac{3}{4}, \tag{1}$$

in a number of steps polynomial in $1/\epsilon$ and the problem size.

We show how to use a quantum computer to speed up a class of FPRAS for estimating partition functions that rely on simulated annealing and the Monte Carlo Markov chain method (e.g., [8,9]). Let us start with an outline of these classical algorithms. Consider a physical system with state space Ω and an energy function $E:\Omega \to \mathbb{R}$, assigning each state $\sigma \in \Omega$ an energy $E(\sigma)$. The task is to estimate the Gibbs partition function

$$Z(T) = \sum_{\sigma \in \Omega} e^{-E(\sigma)/kT} \tag{2}$$

at a desired (usually very low) temperature T_F. We would like to know the value of Z at zero temperature because it is equal to the number of the system configurations with zero energy [17], and this could be a hard counting problem.

The partition function $Z(T)$ encodes the thermodynamical properties of the system in equilibrium at temperature T, where the probability of finding the system in state σ is given by the Boltzmann distribution,

$$\pi_i(\sigma) = \frac{1}{Z(T)} e^{-E(\sigma)/kT}. \tag{3}$$

It is hard to estimate $Z(T)$ directly. The schemes we want to speed up use the following trick. Consider a sequence of decreasing temperatures $T_0 \geq T_1 \geq \cdots \geq T_\ell$, where T_0 is a very high starting temperature and $T_\ell = T_F$ is the desired final temperature. Then, $Z(T_F)$ can be expressed as a telescoping product

$$Z(T_F) = Z_0 \frac{Z_1}{Z_0} \cdots \frac{Z_{\ell-1}}{Z_{\ell-2}} \frac{Z_\ell}{Z_{\ell-1}} = Z_0 \underbrace{(\alpha_0 \alpha_1 \cdots \alpha_{\ell-2} \alpha_{\ell-1})}_{\alpha}, \tag{4}$$

where $Z_i = Z(T_i)$ stands for the Gibbs partition function at temperature T_i and $\alpha_i = Z_{i+1}/Z_i$. It is easy to calculate the partition function $Z_0 = Z(T_0)$ at high temperature. Next, for each i, we can estimate the ratio α_i by sampling from a distribution that is sufficiently close to the Boltzmann distribution π_i (3) at temperature T_i (see Sec. II for more detail). This is possible by using a rapidly mixing Markov chain P_i whose stationary distribution is equal to the Boltzmann distribution π_i.

To be efficient, these classical schemes require that (1) we use a cooling schedule such that the resulting ratios $\alpha_i = Z(T_{i+1})/Z(T_i)$ are lower bounded by a constant c^{-1} (to simplify the presentation, we use $c=2$ from now on), and (2) the spectral gaps of the Markov chains P_i are bounded from below by δ.

The time complexity of such FPRAS, i.e., the number of times we have to invoke an update step for a Markov chain from $\{P_1, \ldots, P_{\ell-1}\}$, is

$$\tilde{O}\left(\frac{\ell^2}{\epsilon^2 \delta}\right), \tag{5}$$

where $\tilde{O}$ means up to logarithmic factors.

Our main result is a general method for "quantizing" such algorithms. Note that the method we present in this paper does not yet allow us to speed up the more complicated classical algorithm for the permanent (which requires to sample from the stationary distributions of the previously used Markov chains to decide which Markov chain to use next). Together with the algorithms using adaptive cooling schedules, it is a direction for further research.

Theorem 1. Consider a classical FPRAS for approximating the Gibbs partition function of a physical system at temperature T_F, satisfying the above conditions. Then, there exists a fully polynomial quantum approximation scheme that uses

$$\tilde{O}\left(\frac{\ell^2}{\epsilon \sqrt{\delta}}\right) \tag{6}$$

applications of a controlled version of a quantum-walk operator from $\{W(P_1), \ldots, W(P_{\ell-1})\}$.

The reduction in complexity for our quantum algorithm (in comparison to the classical FPRAS) is twofold. First, we reduce the factor $1/\delta$ to $1/\sqrt{\delta}$ by using quantum walks instead of classical Markov chains and utilizing the quadratic relation between spectral and phase gaps. As observed in [4], this relation is at the heart of many quantum search algorithms based on quantum walks (see, e.g., [1] for an overview of such quantum algorithms). Second, we speed up the way to determine the ratios α_i by using the quantum phase-estimation algorithm. This results in the reduction of the factor $1/\epsilon^2$ to $1/\epsilon$.

The quantum algorithm we present builds upon our previous work [10], where two of us have shown how to use quantum walks to approximately prepare coherent encodings

$$|\pi_i\rangle = \sum_{\sigma \in \Omega} \sqrt{\pi_i(\sigma)} |\sigma\rangle \tag{7}$$

of stationary distributions π_i of Markov chains P_i, provided that the Markov chains are slowly varying. Recall that a sequence of Markov chains is called slowly varying if the stationary distributions of two adjacent chains are sufficiently close to each other. As we will see later, this condition is automatically satisfied for Markov chains that are used in FPRAS for approximating partition functions.

Note that our objective of approximately preparing coherent encodings of stationary distributions is different from the objective in [14], where the author seeks to speed up the process of approximately preparing density operators encoding stationary distributions. For our purposes, we have to work with coherent encodings because otherwise we could not achieve the second reduction from $1/\epsilon^2$ to $1/\epsilon$.

The paper is organized as follows. In Sec. II we review the classical FPRAS in more detail. We present our quantum algorithm in two steps. First, in Sec. III B we explain how our quantum algorithm works, assuming that we can perfectly and efficiently prepare coherent encodings of the distributions [Eq. (3)]. Then, in Sec. III C we describe the full quantum algorithm, dropping the assumption of Sec. III B and using approximate procedures for quantum sample preparation and readout, which are based on the quantum walks. We perform a detailed analysis of accumulation of error due to the approximation procedures and show that the success probability remains high, establishing theorem 1. Finally, in Sec. IV we conclude with a discussion of open questions, the connection of our algorithm to simulated annealing, and the directions for future research.

II. STRUCTURE OF THE CLASSICAL ALGORITHM

Here we describe the classical approximation schemes in more detail, following closely the presentation in [9] (Sec. 2.1). Choosing a sequence of temperatures $T_0 \geq T_1 \geq \cdots \geq T_\ell$ starting with $T_0 = \infty$ and ending with the desired final (low) temperature $T_\ell = T_F$, we can express the Gibbs partition function (2) as a telescoping product (4). At $T_0 = \infty$, the partition function Z_0 is equal to

$$Z_0 = |\Omega|, \tag{8}$$

the size of the state space. On the other hand, for each $i = 0, \ldots, \ell - 1$, we can estimate the ratio

$$\alpha_i = \frac{Z_{i+1}}{Z_i} \tag{9}$$

in Eq. (4) as follows. Let $X_i \sim \pi_i$ denote a random state chosen according to the Boltzmann distribution π_i, i.e.,

$$\mathcal{P}[X_i = \sigma] = \pi_i(\sigma). \tag{10}$$

Define a new random variable Y_i by

$$Y_i = e^{-(\beta_{i+1} - \beta i) E(X_i)}, \tag{11}$$

where $\beta_i = (kT_i)^{-1}$ is the inverse temperature (k is the Boltzmann constant). This Y_i is an unbiased estimator for α_i since

$$\mathbf{E}(Y_i) = \sum_{\sigma\in\Omega} \pi_i(\sigma) e^{-(\beta_{i+1}-\beta i)E(\sigma)}, \tag{12}$$

$$= \sum_{\sigma\in\Omega} \frac{e^{-\beta_i E(\sigma)}}{Z_i} e^{-(\beta_{i+1}-\beta i)E(\sigma)}, \tag{13}$$

$$= \sum_{\sigma\in\Omega} \frac{e^{-\beta_{i+1}E(\sigma)}}{Z_i} = \frac{Z_{i+1}}{Z_i} = \alpha_i. \tag{14}$$

Assume now that we have an algorithm for generating states X_i according to π_i. We draw

$$m := 64\ell/\epsilon^2 \tag{15}$$

samples of X_i and take the mean $\bar{Y}_i$ of their corresponding estimators Y_i. Then, the mean $\bar{Y}_i$ satisfies

$$\frac{\mathbf{Var}(\bar{Y}_i)}{[\mathbf{E}(\bar{Y}_i)]^2} = \frac{\epsilon^2}{64\ell}\frac{\mathbf{Var}(Y_i)}{[\mathbf{E}(Y_i)]^2} \le \frac{\epsilon^2}{16\ell}. \tag{16}$$

(We have used the assumption $\frac{1}{2} \le \alpha_i \le 1$.) We can now compose such estimates of α_i. Define a new random variable $\bar{Y}$ by

$$\bar{Y} = \bar{Y}_{\ell-1}\bar{Y}_{\ell-2}\cdots\bar{Y}_0. \tag{17}$$

Since all $\bar{Y}_i$ are independent, we have

$$\mathbf{E}(\bar{Y}) = \mathbf{E}(Y_{\ell-1})\mathbf{E}(Y_{\ell-2})\cdots\mathbf{E}(Y_0) = \alpha_{\ell-1}\alpha_{\ell-2}\cdots\alpha_0 = \alpha. \tag{18}$$

Moreover, $\bar{Y}$ has the property

$$\begin{aligned}\frac{\mathbf{Var}(\bar{Y})}{[\mathbf{E}(\bar{Y})]^2} &= \frac{\mathbf{E}(\bar{Y}_{\ell-1}^2)\cdots\mathbf{E}(\bar{Y}_0^2) - \mathbf{E}(\bar{Y}_{\ell-1})^2\cdots\mathbf{E}(\bar{Y}_0)^2}{\mathbf{E}(\bar{Y}_{\ell-1})^2\cdots\mathbf{E}(\bar{Y}_0)^2} \\ &= \left(1+\frac{\mathbf{Var}(\bar{Y}_{\ell-1})}{[\mathbf{E}(\bar{Y}_{\ell-1})]^2}\right)\cdots\left(1+\frac{\mathbf{Var}(\bar{Y}_0)}{[\mathbf{E}(\bar{Y}_0)]^2}\right) - 1 \\ &\le (e^{\epsilon^2/16\ell})^\ell - 1 \le \epsilon^2/8,\end{aligned} \tag{19}$$

where we used $1+x \le e^x$ (true for all x) and $e^x - 1 \le 2x$ (true for all $x \in [0,1]$) in the last two steps, respectively. Chebyshev's inequality now implies that the value of $\bar{Y}$ is in the interval $[(1-\epsilon)\alpha, (1+\epsilon)\alpha]$ with probability at least $\frac{7}{8}$.

Of course, we are not able to obtain perfect samples X_i from π_i. Assume now that we have X_i' that are from a distribution with a variation distance from π_i smaller than

$$d := \epsilon^2/(512\ell^2). \tag{20}$$

Let $\bar{Y}'$ be defined as $\bar{Y}$ as above, but instead of X_i we use X_i'. Then, with probability at least $\frac{7}{8}$, we have $\bar{Y} = \bar{Y}'$. To derive this, observe that the algorithm can be thought to first take a sample from a product probability distribution π on the $(m\ell)$-fold direct product of Ω. We denote the probability distribution in the case of imperfect samples by π'. The total variation distance between π and π' is then bounded from above by

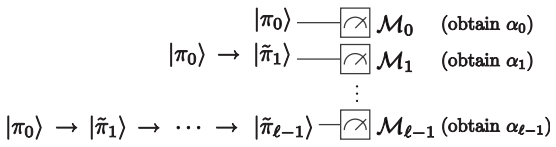

FIG. 1. Structure of the quantum algorithm.

$$dm\ell = \frac{\epsilon^2}{512\ell^2}\frac{64\ell}{\epsilon^2}\ell = \frac{1}{8}. \tag{21}$$

Therefore, $\bar{Y}'$ is in the interval $[(1-\epsilon)\mathbf{E}(Y), (1+\epsilon)\mathbf{E}(Y)]$ with probability at least $\frac{3}{4}$.

We obtain the samples X_i' by applying Markov chains P_i whose limiting distributions are equal to π_i. Constructing such rapidly mixing Markov chains is a hard task, but it has been done for the Ising model [8] and the Potts model [9].

III. QUANTUM ALGORITHM

A. Overview

The classical FPRAS we described in Sec. II consists of (1) preparing many samples from a distribution close to π_i by letting a suitable Markov chain mix, (2) using these samples to approximate the ratios α_i in Eq. (4), and (3) composing these estimates of α_i into an estimate of the partition function.

We build our quantum algorithm on this scheme, adding two quantum ingredients. First, instead of letting a Markov chain P_i mix toward its stationary distribution π_i, we choose to approximately prepare the state $|\pi_i\rangle = \Sigma_\sigma \sqrt{\pi_i(\sigma)}|\sigma\rangle$, a coherent encoding of the Boltzmann distribution. We use a preparation method [10] based on Grover's $\frac{\pi}{3}$-fixed-point search [13], efficiently driving the state $|\pi_0\rangle$ toward the desired state $|\pi_i\rangle$ through a sequence of intermediate states.

Second, instead of using classical samples from the distribution π_i, we approximate α_i by phase estimation of a certain unitary on the state $|\pi_i\rangle$. This is a new concept, going beyond our previous work [10]. This phase-estimation subroutine can be efficiently (albeit only approximately) applied by utilizing quantum walks.

The structure of our algorithm is depicted in Fig. 1. It consists of successive approximate preparations of $|\pi_i\rangle$ followed by a quantum circuit outputting a good approximation to α_i (with high probability). Our main result is the construction of a fast quantum version of a class of classical algorithms summed in theorem 1.

We arrive at our quantum algorithm in two steps. First, in Sec. III B, we explain how to quantize the classical algorithm in the perfect case, assuming that we can take perfect samples X_i from π_i. Then, in Sec. III C we release this assumption and describe the full quantum algorithm.

B. Perfect case

To estimate the ratios α_i in Eq. (4), the classical algorithm generates random states X_i from π_i and computes the mean $\bar{Y}_i$ of the random variables Y_i. The process of generating a

random state X_i from π_i is equivalent to preparing the mixed state

$$\rho_i = \sum_{\sigma\in\Omega} \pi_i(\sigma)|\sigma\rangle\langle\sigma|. \tag{22}$$

Instead of this, we choose to prepare the pure states

$$|\pi_i\rangle = \sum_{\sigma\in\Omega} \sqrt{\pi_i(\sigma)}|\sigma\rangle. \tag{23}$$

We call these states *quantum samples* since they coherently encode the probability distributions π_i. In this section, we assume that we can prepare these exactly and efficiently.

The random variable Y_i can be interpreted as the outcome of the measurement of the observable

$$A_i = \sum_{\sigma\in\Omega} y_i(\sigma)|\sigma\rangle\langle\sigma| \tag{24}$$

in the state ρ_i, where

$$y_i(\sigma) = e^{-(\beta_{i+1}-\beta_i)E(\sigma)}. \tag{25}$$

With this interpretation in mind, we see that to estimate α_i classically, we need to estimate the expected value $\mathrm{Tr}(A_i\rho_i)$ by repeating the above measurement several times and outputting the mean of the outcomes.

We now explain how to quantize this process. We add an ancilla qubit to our quantum system in which the quantum samples $|\pi_i\rangle$ live. For each $i=0,\ldots,\ell-1$, we define the unitary

$$V_i = \sum_{\sigma\in\Omega} |\sigma\rangle\langle\sigma| \otimes \begin{pmatrix} \sqrt{y_i(\sigma)} & \sqrt{1-y_i(\sigma)} \\ -\sqrt{1-y_i(\sigma)} & \sqrt{y_i(\sigma)} \end{pmatrix}. \tag{26}$$

This V_i can be efficiently implemented, it is a rotation on the extra qubit controlled by the state of the first tensor component. Let us label

$$|\psi_i\rangle = V_i(|\pi_i\rangle \otimes |0\rangle). \tag{27}$$

Consider now the expected value of the projector

$$T = \mathbb{I} \otimes |0\rangle\langle 0| \tag{28}$$

in the state $|\psi_i\rangle$. We find

$$\langle\psi_i|T|\psi_i\rangle = \langle\pi_i|A_i|\pi_i\rangle = \alpha_i. \tag{29}$$

We now show how to speed up the process of estimating α_i with a method that generalizes quantum counting [11]. As noted in the beginning of this section, we assume efficient preparation of $|\pi_i\rangle$, which in turn implies that we can efficiently implement the reflections

$$R_i = 2|\pi_i\rangle\langle\pi_i| - \mathbb{I}. \tag{30}$$

The result of this section, the existence of a quantum FPRAS for estimating the partition function assuming efficient and perfect preparation of $|\pi_i\rangle$, is summed in theorem 2.

Theorem 2. There is a fully polynomial quantum approximation scheme $\mathcal{A}$ for the partition function Z. Its output Q satisfies

$$\mathcal{P}[(1-\epsilon)Z \le Q \le (1+\epsilon)Z] \ge \frac{3}{4}. \tag{31}$$

For each $i=0,\ldots,\ell-1$, the scheme $\mathcal{A}$ uses

$$O(\log_2 \ell) \tag{32}$$

perfectly prepared quantum samples $|\pi_i\rangle$ and applies the controlled-R_i operator

$$O\left(\frac{\ell}{\epsilon}\log_2 \ell\right) \tag{33}$$

times, where R_i is as in Eq. (30).

To prove theorem 2, we need the following three technical results.

Lemma 1. (Quantum ratio estimation). Let $\epsilon_{pe}\in(0,1)$. For each $i=0,\ldots,\ell-1$, there exists a quantum approximation scheme $\mathcal{A}_i'$ for α_i. Its output Q_i' satisfies

$$\mathcal{P}[(1-\epsilon_{pe})\alpha_i \le Q_i' \le (1+\epsilon_{pe})\alpha_i] \ge \frac{7}{8}. \tag{34}$$

The scheme $\mathcal{A}_i'$ requires one copy of the quantum sample $|\pi_i\rangle$ and invokes the controlled-R_i operator $O(\epsilon_{pe}^{-1})$ times, where R_i is as in Eq. (30).

Proof. Let

$$G = (2|\psi_i\rangle\langle\psi_i| - \mathbb{I})(2T - \mathbb{I}). \tag{35}$$

Define the basis states

$$|\gamma_1\rangle = \frac{(\mathbb{I}-T)|\psi_i\rangle}{\sqrt{1-\alpha_i}}, \quad |\gamma_2\rangle = \frac{T|\psi_i\rangle}{\sqrt{\alpha_i}}. \tag{36}$$

Restricted to the plane spanned by $|\gamma_1\rangle$ and $|\gamma_2\rangle$, G acts as a rotation

$$G|_{\{|\gamma_1\rangle,|\gamma_2\rangle\}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{37}$$

where $\theta\in[0,\frac{\pi}{2}]$ satisfies

$$\cos\theta = 2\alpha_i - 1. \tag{38}$$

The eigenvectors and eigenvalues of G are

$$|G_\pm\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \pm i \end{bmatrix}, \quad \lambda_\pm = e^{\pm i\theta}. \tag{39}$$

We do not have direct access to one of these eigenvectors, as the state $|\psi_i\rangle$ is in a superposition of $|G_+\rangle$ and $|G_-\rangle$. Thus, when we apply the phase-estimation circuit for the unitary G to the state $|\psi_i\rangle$, we will sometimes obtain an estimate of θ, and sometimes an estimate of $2\pi-\theta$. However, this is not a problem since both θ and $2\pi-\theta$ plugged into Eq. (38) yield the same result for α_i.

We require that the estimate θ' satisfies

$$|\theta' - \theta| \le 2\epsilon_{pe}\alpha_i \le \epsilon_{pe}, \tag{40}$$

with probability at least $\frac{7}{8}$. Using the phase-estimation circuit in [12], this means that $\frac{\theta'}{2\pi}$ has to be an $n_a=\log_2\frac{2\pi}{\epsilon_{pe}}$ bit approximation of the phase and the failure probability p_f has to

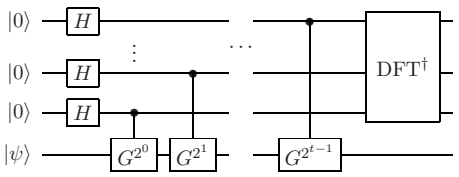

FIG. 2. A basic phase-estimation circuit with t ancilla qubits.

be less than $\frac{1}{8}$. To achieve this, it suffices to use a phase-estimation circuit (see Fig. 2) with

$$t = \log_2 \frac{2\pi}{\epsilon_{pe}} + \log_2\left(2 + \frac{1}{2p_f}\right) = O(\log_2 \epsilon_{pe}^{-1})$$

ancilla qubits. This circuit invokes the controlled-G operation $O(2^t)=O(\epsilon_{pe}^{-1})$ times.

Let α_i' denote the value we compute from the estimate θ'. We have

$$|\alpha_i - \alpha_i'| = \frac{1}{2}|\cos\theta - \cos\theta'| \le \frac{1}{2}|\theta - \theta'| \le \epsilon_{pe}\alpha_i, \quad (41)$$

showing that the estimate α_i' is within $\pm\epsilon_{pe}\alpha_i$ of the exact value α_i with probability at least $\frac{7}{8}$. This completes the proof that the random variable Q_i' corresponding to the output satisfies the desired properties on estimation accuracy and success probability. ■

We can boost the success probability of the above quantum approximation scheme for the ratio α_i by applying the *powering lemma* from [15], which we state here for completeness.

Lemma 2. (Powering lemma for approximation schemes). Let $\mathcal{B}'$ be a (classical or quantum) approximation scheme whose estimate W' is within $\pm\epsilon_{pe}q$ to some value q with probability $\frac{1}{2}+\Omega(1)$. Then, there is an approximation scheme $\mathcal{B}$ whose estimate W satisfies

$$\mathcal{P}[(1-\epsilon_{pe})q \le W \le (1+\epsilon_{pe})q] \ge 1 - \delta_{boost}. \quad (42)$$

It invokes the scheme $\mathcal{B}'$ as a subroutine $O(\log_2 \delta_{boost}^{-1})$ times.

With the help of lemma 2, we now have the constituents required to compose the individual estimates of α_i into an approximation for the partition function (4).

Lemma 3. Let $\epsilon>0$. Assume we have approximation schemes $\mathcal{A}_0,\mathcal{A}_1,\ldots,\mathcal{A}_{\ell-1}$ such that their estimates $Q_0,Q_1,\ldots,Q_{\ell-1}$ satisfy

$$\mathcal{P}\left[\left(1-\frac{\epsilon}{2\ell}\right)\alpha_i \le Q_i \le \left(1+\frac{\epsilon}{2\ell}\right)\alpha_i\right] \ge 1 - \frac{1}{4\ell}. \quad (43)$$

Then, there is a simple approximation scheme $\mathcal{A}$ for the product $\alpha=\alpha_0\alpha_1\cdots\alpha_{\ell-1}$. The result $Q=Q_0Q_1\cdots Q_{\ell-1}$ satisfies

$$\mathcal{P}[(1-\epsilon)\alpha \le Q \le (1+\epsilon)\alpha] \ge \frac{3}{4}. \quad (44)$$

Proof. For each $i=0,\ldots,\ell-1$, the failure probability for estimating α_i is smaller than $1/(4\ell)$. The union bound implies that the overall failure probability is smaller than 1/4, proving the lower bound $\frac{3}{4}$ on the success probability in Eq. (44).

To obtain the upper bound on the deviation, we now assume that each Q_i takes the upper bound value. We have

$$\frac{Q-\alpha}{\alpha} \le \prod_{i=0}^{\ell-1}\left(1+\frac{\epsilon}{2\ell}\right) - 1 = \left(1+\frac{\epsilon}{2\ell}\right)^{\ell} - 1 \le e^{\epsilon/2} - 1 \le \epsilon,$$

where we have used $1+x\le e^x\le 1+2x$, which is true for all $x\in[0,1]$. Thus, in the case of success, we have $Q\le(1+\epsilon)\alpha$.

To obtain the lower bound on the deviation, we assume that each Q_i takes its lower bound value. We have

$$\frac{\alpha-Q}{\alpha} \le 1 - \prod_{i=0}^{\ell-1}\left(1-\frac{\epsilon}{2\ell}\right) \le \sum_{i=0}^{\ell-1}\frac{\epsilon}{2\ell} \le \epsilon, \quad (45)$$

where we have used $|\Pi_i x_i - \Pi_i y_i| \le \Sigma_i |x_i - y_i|$, true for arbitrary $x_i,y_i\in[0,1]$. Thus, in the case of success, we have $(1-\epsilon)\alpha\le Q$. ■

We are now ready to prove theorem 2.

Proof of Theorem 2. For each $i=0,\ldots,\ell-1$, we can apply lemma 1 with the state $|\psi_i\rangle$ (27) and the projector T (28). This gives us a quantum approximation scheme for α_i. Note that to prepare $|\psi_i\rangle$, it suffices to prepare $|\pi_i\rangle$ once. Also, to realize a controlled reflection around $|\psi_i\rangle$, it suffices to invoke the controlled reflection around $|\pi_i\rangle$ once.

We now use the reflection $2|\psi_i\rangle\langle\psi_i|-\mathbb{I}$ and set $\epsilon_{pe}=\epsilon/(2\ell)$ in lemma 1. With these settings, we can apply lemma 2 to the resulting approximation scheme for α_i with $\delta_{boost}=1/(4\ell)$. This gives us approximation schemes $\mathcal{A}_i$ outputting Q_i with high precision and probability of success that can be used in lemma 3. The composite result $Q=Q_0\cdots Q_{\ell-1}$ is thus an approximation for $\alpha=\alpha_0\cdots\alpha_{\ell-1}$ with the property

$$\mathcal{P}[(1-\epsilon)\alpha \le Q \le (1+\epsilon)\alpha] \ge \frac{3}{4}. \quad (46)$$

Finally, we obtain the estimate for Z by multiplying Q with Z_0. Let us summarize the costs from lemmas 1–3. For each $i=0,\ldots,\ell-1$, this scheme uses $\log_2\delta_{boost}^{-1}=O(\log_2\ell)$ copies of the state $|\pi_i\rangle$, and invokes $(\log_2\delta_{boost}^{-1})\epsilon_{pe}^{-1}=O(\frac{\ell}{\epsilon}\log_2\ell)$ reflections around $|\pi_i\rangle$. ■

C. Quantum FPRAS

In the previous section, we have assumed that we can prepare the quantum samples $|\pi_i\rangle$ and implement the controlled reflections $R_i=2|\pi_i\rangle\langle\pi_i|-\mathbb{I}$ about these states perfectly and efficiently. We now release these assumptions and show how to approximately accomplish both tasks with the help of quantum-walk operators. We then show that the errors arising from these approximate procedures do not significantly decrease the success probability of the algorithm. This will wrap up the proof of our main result, theorem 1.

In [10], two of us showed how to approximately prepare quantum samples $|\pi_i\rangle$ of stationary distributions of slowly varying Markov chains. Using the fact that the consecutive states $|\pi_i\rangle$ and $|\pi_{i+1}\rangle$ are close, we utilize Grover's $\frac{\pi}{3}$ fixed-

point search [13] to drive the starting state $|\pi_0\rangle$ toward the desired state $|\pi_i\rangle$ through multiple intermediate steps. Moreover, to be able to perform this kind of Grover search, we have to be able to apply selective phase shifts of the form $S_i=\omega|\pi_i\rangle\langle\pi_i|+(\mathbb{I}-|\pi_i\rangle\langle\pi_i|)$ for $\omega=e^{i\pi/3}$ and $\omega=e^{-i\pi/3}$. This is another assumption of Sec. III B that we have to drop here. Nevertheless, an efficient way to apply these phase shifts approximately based on quantum walks and phase estimation exists [10].

Our task is to show that the approximation scheme from lemma 1 works even with approximate input states and using only approximate reflections about the states $|\pi_i\rangle$. Let us start with addressing the approximate state preparation. To be able to use the results of [10], we first have to establish an important condition. For their method to be efficient, the overlap of two consecutive quantum samples $|\pi_i\rangle$ and $|\pi_{i+1}\rangle$ has to be large. This is satisfied when $\alpha_i=Z_{i+1}/Z_i$ is bounded from below by $\frac{1}{2}$ since

$$|\langle\pi_i|\pi_{i+1}\rangle|^2 = \left|\sum_{\sigma\in\Omega}\frac{\sqrt{e^{-\beta_i E(\sigma)}e^{-\beta_{i+1}E(\sigma)}}}{\sqrt{Z_i Z_{i+1}}}\right|^2 \geq \left|\frac{\sum_{\sigma\in\Omega}e^{-\beta_{i+1}E(\sigma)}}{\sqrt{2Z_{i+1}}\sqrt{Z_{i+1}}}\right|^2 = \frac{1}{2}.$$

The following lemma then directly follows from the arguments used in [10] (theorem 2).

Lemma 4. For $\epsilon_S>0$ arbitrary and each $i=1,\ldots,\ell-1$, there is a quantum method preparing a state $|\tilde{\pi}_i\rangle$ with

$$\|\,|\tilde{\pi}_i\rangle-|\pi_i\rangle|0\rangle^{\otimes a}\|\leq\epsilon_S, \tag{47}$$

where $a=O(\frac{\ell}{\epsilon_S\sqrt{\delta}})$ is the number of ancilla qubits. The method invokes a controlled version of a walk operator from the set $\{W(P_1),\ldots,W(P_{\ell-1})\}$

$$O\left(\frac{\ell}{\sqrt{\delta}}\log_2^2\frac{\ell}{\epsilon_S}\right) \tag{48}$$

times.

We choose the preparation method from lemma 4 with $\epsilon_S=\frac{1}{32}$. The cost for this precision ϵ_S is

$$O\left(\frac{\ell}{\sqrt{\delta}}\log_2^2\,\ell\right) \tag{49}$$

applications of the quantum-walk operator. Recall that when we used lemma 1 in Sec. III B with the state $|\psi_i\rangle$ (coming from the perfect quantum sample $|\pi_i\rangle$) as input, the success probability of the resulting scheme was greater than $\frac{7}{8}$. We now use the method given in lemma 1 on the approximate input $|\tilde{\psi}_i\rangle=V_i(|\tilde{\pi}_i\rangle\otimes|0\rangle)$. With our chosen precision for preparing $|\tilde{\pi}_i\rangle$, the success probability of the approximation scheme of lemma 1 cannot decrease by more than $2\cdot\frac{1}{32}$.

The second assumption of lemma 1 we need to drop is the ability to perfectly implement the reflections $R_i=2|\pi_i\rangle\langle\pi_i|-\mathbb{I}$. We now show how to approximately implement these reflections. The following lemma follows directly from the arguments in [10].

Lemma 5. For $\epsilon_R>0$ arbitrary and each $i=1,\ldots,\ell-1$, there is an approximate reflection $\tilde{R}_i$ such that

$$\tilde{R}_i(|\varphi\rangle\otimes|0\rangle^{\otimes b})=(R_i|\varphi\rangle)\otimes|0\rangle^{\otimes b}+|\xi\rangle \tag{50}$$

where $|\varphi\rangle$ is an arbitrary state, $b=O(\log_2\epsilon_R^{-1}\log_2\frac{1}{\sqrt{\delta}})$ is the number of ancilla qubits, and $|\xi\rangle$ is some error vector with $\|\,|\xi\rangle\|\leq\epsilon_R$. It invokes the controlled version of a walk operator from $\{W(P_1),\ldots,W(P_{\ell-1})\}$

$$O\left(\frac{1}{\sqrt{\delta}}\log_2\frac{1}{\epsilon_R}\right) \tag{51}$$

times.

Recall that in lemma 1, the controlled reflection R_i is invoked $O(1/\epsilon_{pe})$ times. We now run this approximation scheme with $\tilde{R}_i$ instead of R_i. The norm of the accumulated error vector is

$$O\left(\frac{1}{\epsilon_{pe}}\right)\epsilon_R. \tag{52}$$

We choose

$$\epsilon_R=\Omega(\epsilon_{pe}) \tag{53}$$

to bound the norm of the accumulated error from above by $\frac{1}{32}$. The success probability can then decrease by at most $2\cdot\frac{1}{32}$.

Combining these arguments establishes a variant of lemma 1 without the unnecessary assumptions of Sec. III B:

Lemma 6. Let $\epsilon_{pe}\in(0,1)$. For each $i=0,\ldots,\ell-1$, there exists a quantum approximation scheme $\mathcal{A}_i''$ for α_i. Its estimate Q_i'' satisfies

$$\mathcal{P}[(1-\epsilon_{pe})\alpha_i\leq Q_i''\leq(1+\epsilon_{pe})\alpha_i]\geq\frac{3}{4}. \tag{54}$$

This scheme invokes the controlled version of a walk operator from $\{W(P_1),\ldots,W(P_{\ell-1})\}$

$$O\left(\frac{\ell}{\sqrt{\delta}}\log_2^2\,\ell+\frac{1}{\epsilon_{pe}\sqrt{\delta}}\log_2\,\epsilon_{pe}^{-1}\right). \tag{55}$$

Proof. The success probability of the scheme in lemma 1 was greater than $\frac{7}{8}$. Both the approximate state preparation and using approximate reflections reduce the overall probability of success by at most $\frac{1}{16}$. Thus, the probability of success of the method given in lemma 1 after dropping the unnecessary assumptions is at least $\frac{3}{4}$. ■

We can finally complete the proof of theorem 1 by following the procedure that led to the proof of Theorem 2 in Sec. III B.

Proof of Theorem 2. For each $i=0,\ldots,\ell-1$, we proceed as follows. We use the approximation scheme $\mathcal{A}_i''$ from lemma 6 with precision $\epsilon_{pe}=\epsilon/(2\ell)$. We then boost the success probability of each $\mathcal{A}_i''$ to $1-\frac{1}{4\ell}$ by applying the powering lemma (lemma 2) with $\delta_{boost}=1/(4\ell)$. This step increases the cost in Eq. (55) by the factor $O(\log_2\ell)$. This resulting scheme $\mathcal{A}_i$ now satisfies the properties required for lemma 3.

We can thus use it to obtain a composite approximation scheme whose output satisfies

$$\mathcal{P}[(1-\epsilon)Z \leq Q \leq (1+\epsilon)Z] \geq \frac{3}{4}. \tag{56}$$

The resulting cost of this scheme (the number of times we have to invoke the controlled quantum-walk operators) is

$$O\left[\frac{\ell^2}{\sqrt{\delta}}\log_2^3 \ell + \frac{\ell^2}{\epsilon\sqrt{\delta}}(\log_2 \ell)(\log_2 \ell + \log_2 \epsilon^{-1})\right] = \tilde{O}\left(\frac{\ell^2}{\epsilon\sqrt{\delta}}\right). \tag{57}$$

■

IV. CONCLUSIONS

We have shown that in the quantum circuit model, we can speed up a class of classical FPRAS for approximating partition functions, as measured in the number of times we have to invoke [18] a step of a quantum walk (instead of classical Markov chains). We obtained two reductions in complexity: $1/\delta \to 1/\sqrt{\delta}$ and $1/\epsilon^2 \to 1/\epsilon$. These two reductions are intimately related; they cannot occur separately. If we used quantum samples merely to obtain classical samples (i.e., if we tried to estimate the ratios without phase estimation) then this would lead to $O(\ell^3)$ dependence (for $\epsilon \propto \ell^{-1}$). This is because we would have to take $O(\frac{\ell}{\epsilon^2})$ classical samples for each i and producing a quantum sample costs at least $O(\ell)$. The advantage of our approximation procedure based on quantum phase estimation is that it requires only one quantum sample (or more precisely, $\log_2 \ell$, after using the powering lemma to boost the success probability). We cannot obtain the second speed up without using quantum samples (as mentioned in the introduction, this prevents us from using a procedure such as [14] that prepares density operators encoding stationary distributions). Also, the arguments employed in the error analysis in the quantum case are quite different from those in the classical error analysis.

Each classical FPRAS, we speed up uses the telescoping trick (4), a particular cooling schedule (decreasing sequence of temperatures), and slowly varying Markov chains which mix rapidly, with stationary distributions equal to the Boltzmann distributions at the intermediate temperatures. The classical FPRAS is useful only when we have the Markov chains with the required properties. Moreover, the cooling schedules need to be such that the ratios α_i [Eq. (9)] are lower bounded by some c^{-1}. In [16], the authors show that it is possible to use a cooling schedule $T_0=\infty > T_1' > \ldots > T_{\ell'-1}' = T_F$ for estimating the partition function $Z(T_F)$ as long as for each i,

$$\frac{\mathbf{E}(Y_i^2)}{[\mathbf{E}(Y_i)]^2} \leq b, \tag{58}$$

where b is some constant. Such a cooling schedule is called a Chebyshev cooling schedule. Note that the above condition is automatically satisfied in the situation we consider in this paper, but not vice versa [recall that we assume that we have a cooling schedule such that $\mathbf{E}(Y_i)$ is bounded from below by a constant for each i; we set it to $\frac{1}{2}$ for simplicity of presentation]. The advantage of Chebyshev cooling schedules is that they are provably shorter. The authors present an adaptive algorithm for constructing Chebyshev cooling schedule. We plan to explore if it is possible to speed up this process. But even if this is possible, a potential obstacle remains. It is not clear whether we can still obtain the reduction from $\frac{1}{\epsilon^2}$ to $\frac{1}{\epsilon}$ when we only know that the condition (58) is satisfied. It seems that the condition $\mathbf{E}(Y_i) > c^{-1}$ with c some constant is absolutely necessary for phase estimation to yield the quadratic speed up with respect to the accuracy parameter ϵ.

The combination of simulated annealing and the Monte Carlo Markov chain method used in approximating partition functions is the central piece of the best currently known algorithm for estimating permanents with non-negative entries [9]. We therefore plan to explore where our techniques can be used to speed up this breakthrough classical algorithm.

ACKNOWLEDGMENTS

A.A., C.C., and P.W. gratefully acknowledge the support by NSF Grants No. CCF-0726771 and No. CCF-0746600. D.N. gratefully acknowledges support by European Project QAP No. 2004-IST-FETPI-15848 and by the Slovak Research and Development Agency under the Contract No. APVV-0673-07.

APPENDIX: QUANTUM WALKS FROM CLASSICAL MARKOV CHAINS

The class of classical approximation schemes that we speed up uses reversible ergodic Markov chains P_i with stationary distributions π_i. Here, we briefly review the quantum analog of a Markov chain, describing the quantum-walk operator W corresponding to the classical Markov Chain P.

In each step of a Markov chain P with state space Ω, the probability of a state x to transition to another state y is given by the element p_{xy} of the $D \times D$ transition matrix, where $D=|\Omega|$. Following Szegedy [3], for each such Markov chain, we can define its quantum analog. The Hilbert space on which this quantum operation acts is $\mathbb{C}^D \otimes \mathbb{C}^D$, with two $\mathbb{C}^D$ registers. We start by defining the states

$$|p_x\rangle = \sum_{y \in \Omega} \sqrt{p_{xy}}|y\rangle. \tag{A1}$$

These states can be generated by a *quantum update*—any unitary U that satisfies

$$U|x\rangle|0\rangle = |x\rangle|p_x\rangle \tag{A2}$$

for some fixed state $0 \in \Omega$ and all $x \in \Omega$. The quantum analog of a Markov chain is then defined as follows.

Definition 1. (Quantum Walk). A quantum walk $W(P)$ based on a classical reversible Markov chain P is a unitary operation acting on the space $\mathbb{C}^D \otimes \mathbb{C}^D$ as

$$W(P) = R_{\mathcal{B}} R_{\mathcal{A}}, \tag{A3}$$

where $R_{\mathcal{B}}$ and $R_{\mathcal{A}}$ are reflections about the spaces

$$\mathcal{A} = \text{span}\{|x\rangle|0\rangle : x \in \Omega\}, \tag{A4}$$

$$\mathcal{B} = U^{\dagger} S U \mathcal{A}, \tag{A5}$$

and S is a swap of the two registers.

This particular definition of the quantum walk is suitable for making some of the proofs in [10] easier. It is equivalent to the standard definition of Szegedy [3] up to conjugation by U. Therefore, the spectral properties of our W and Szegedy's quantum walk are the same.

Let δ be the spectral gap of the classical Markov chain P. Let us write its eigenvalues as $\mu_0 = 1$ and $\mu_j = \cos(\theta_j)$, for $j = 1, \ldots, D-1$ and $\theta_j \in (0, \frac{\pi}{2})$. According to Szegedy [3], on the space $\mathcal{A}+\mathcal{B}$, the eigenvalues of the quantum walk $W(P)$ with nonzero imaginary part are $e^{\pm 2i\theta_j}$. The phase gap of the quantum walk $W(P)$ is then defined as $\Delta = 2\theta_1$ (with θ_1 the smallest of θ_j). When the Markov chain is ergodic and reversible, Szegedy proved that

$$\Delta \geq 2\sqrt{\delta}, \tag{A6}$$

a quadratic relation between the phase gap Δ of the quantum walk $W(P)$ and the spectral gap δ of the classical Markov chain P. This quadratic relation is behind the speed up of many of today's quantum-walk algorithms.

[1] M. Santha, Lect. Notes Comput. Sci. **4978**, 31 (2008).

[2] R. Somma, S. Boixo, H. Barnum, and E. Knill, Phys. Rev. Lett. **101**, 130504 (2008).

[3] M. Szegedy, *Quantum Speed-up of Markov Chain Based Algorithms*, Proceedings of 45th Annual IEEE Symposium on Foundations of Computer Science (IEEE Computer Society, Washington, DC, 2004), pp. 32–41.

[4] F. Magniez, A. Nayak, J. Roland, and M. Santha, *Search via Quantum Walk*, Proceedings of the 39th Annual ACM Symposium on Theory of Computing (ACM, New York, 2007), pp. 575–584.

[5] C. Chiang, D. Nagaj, and P. Wocjan, e-print arXiv:0903.3465.

[6] L. Lovász and S. Vempala, J. Comput. Syst. Sci. **72**, 392 (2006).

[7] M. Jerrum, A. Sinclair, and E. Vigoda, J. ACM **51**, 671 (2004).

[8] M. Jerrum and A. Sinclair, SIAM J. Comput. **22**, 1087 (1993).

[9] I. Bezáková, D. Štefankovič, V. Vazirani, and E. Vigoda, SIAM J. Comput. **37**, 1429 (2008).

[10] P. Wocjan and A. Abeyesinghe, Phys. Rev. A **78**, 042336 (2008).

[11] G. Brassard, P. Høyer, and A. Tapp, Lect. Notes Comput. Sci. **1443**, 820 (1998).

[12] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).

[13] L. Grover, e-print arXiv:quant-ph/0503205.

[14] P. C. Richter, Phys. Rev. A **76**, 042306 (2007).

[15] M. Jerrum, L. Valiant, and V. Vazirani, Theor. Comput. Sci. **43**, 169 (1986).

[16] D. Štefankovič, S. Vempala, and E. Vigoda, *Adaptive Simulated Annealing: A Near-Optimal Connection between Sampling and Counting*, Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (IEEE Computer Society, Washington, DC, 2007), pp. 183–193.

[17] This relationship is used, e.g., in the algorithm [7] for approximating the permanent of a non-negative matrix—one can find the value of the permanent by counting the number of perfect matchings of a particular bipartite graph, which in turn is equal to the zero-temperature partition function of a certain spin system.

[18] When the classical Markov chains can be implemented efficiently, each step of the corresponding quantum walks can also be applied efficiently and precisely, as shown, e.g., [5].