# Quantum Speed-up for Approximating Partition Functions

Pawel Wocjan[*]   Chen-Fu Chiang[†]   Anura Abeyesinghe[‡]   Daniel Nagaj [§]

November 4, 2008

## Abstract

We achieve a quantum speed-up of fully polynomial randomized approximation schemes (FPRAS) for estimating partition functions that combine simulated annealing with the Monte-Carlo Markov Chain method and use non-adaptive cooling schedules. The improvement in time complexity is twofold: 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. Second, we generalize the method of quantum counting, showing how to estimate expected values of quantum observables. Using this method instead of classical sampling, we obtain the speed-up with respect to accuracy.

# 1   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

---

[*]School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816, USA. Email: `wocjan@eecs.ucf.edu`

[†]School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816, USA. Email: `cchiang@eecs.ucf.edu`

[‡]School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816, USA.

[§]Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 84215 Bratislava, Slovakia. Email: `daniel.nagaj@savba.sk`

work [14], 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 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 (MCMC) method. To achieve this improvement, we rely on Szegedy's general method to quantize classical Markov chains [2, 3], which we review in Appendix A. 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.

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 [4], the permanent of a non-negative matrix [5], and the partition function of statistical physics models such as the Ising model [6] and the Potts model [7]. 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.

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

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. [6, 7]). Let us start with an outline of these classical algorithms. Consider a physical system with state space $\Omega$ 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

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

at a desired (usually low) temperature $T_F$. 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 $\sigma$ is given by the Boltzmann distribution

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

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 $\alpha_i$ by sampling from a distribution that is sufficiently close to the Boltzmann distribution $\pi_i$ (3) at temperature $T_i$ (see Section 2 for more detail). This is possible by using a rapidly-mixing Markov chain $P_i$ whose stationary distribution is equal to the Boltzman distribution $\pi_i$.

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),

2. the spectral gaps of the Markov chains $P_i$ are bounded from below by $\delta$.

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

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

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

Our main result is a general method for 'quantizing' such algorithms.

**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*

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

*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 [3], 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 $\alpha_i$ by using the quantum phase estimation algorithm in a novel way. This results in the reduction of the factor $1/\epsilon^2$ to $1/\epsilon$.

The quantum algorithm we present builds upon our previous work [8], 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 $\pi_i$ 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 [12], 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 Section 2 we review the classical FPRAS in more detail. We present our quantum algorithm in two steps. First, in Section 3.2 we explain how our quantum algorithm works, assuming that we can perfectly and efficiently prepare coherent encodings of the distributions (3). Then, in Section 3.3 we describe the full quantum algorithm, dropping the assumption of Section 3.2 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 Section 4 we conclude with a discussion of open questions, the connection of our algorithm to simulated annealing, and the directions for future research.

# 2 Structure of the Classical Algorithm

Here we describe the classical approximation schemes in more detail, following closely the presentation in [7, Section 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 (4) as follows. Let $X_i \sim \pi_i$ denote a random state chosen according to the Boltzmann distribution $\pi_i$, i.e.,

$$\Pr(X_i = \sigma) = \pi_i(\sigma). \tag{10}$$

Define a new random variable $Y_i$ by

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

4

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

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

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

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

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

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

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

$$\frac{\mathbf{Var}\left(\overline{Y}_i\right)}{\left(\mathbf{E}\left(\overline{Y}_i\right)\right)^2} = \frac{\epsilon^2}{64\ell}\frac{\mathbf{Var}\left(Y_i\right)}{\left(\mathbf{E}\left(Y_i\right)\right)^2} \leq \frac{\epsilon^2}{16\ell}. \tag{16}$$

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

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

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

$$\mathbf{E}\left(\overline{Y}\right) = \mathbf{E}\left(Y_{\ell-1}\right)\mathbf{E}\left(Y_{\ell-2}\right)\cdots\mathbf{E}\left(Y_0\right) = \alpha_{\ell-1}\alpha_{\ell-2}\cdots\alpha_0 = \alpha,$$

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

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

where we used $1 + x \leq e^x$ (true for all $x$) and $e^x - 1 \leq 2x$ (true for all $x \in [0,1]$) in the last two steps, respectively. Chebyshev's inequality now implies that the value of $\overline{Y}$ 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 $\pi_i$. Assume now that we have $X_i'$ that are from a distribution with a variation distance from $\pi_i$ smaller than

$$d := \epsilon^2/(512\ell^2). \tag{18}$$

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

$$d \cdot m \cdot \ell = \frac{\epsilon^2}{512\ell^2} \cdot \frac{64\ell}{\epsilon^2} \cdot \ell = \frac{1}{8} \,. \tag{19}$$

Therefore, $\overline{Y}'$ 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 $\pi_i$. Constructing such rapidly-mixing Markov chains is a hard task, but it has been done for the Ising model [6] and the Potts model [7].

# 3 Quantum Algorithm

## 3.1 Overview

The classical FPRAS we described in Section 2 consists of

1. preparing many samples from a distribution close to $\pi_i$ by letting a suitable Markov chain mix,

2. using these samples to approximate the ratios $\alpha_i$ in (4), and

3. composing these estimates of $\alpha_i$ into an estimate of the partition function.

We build our quantum algorithm on this scheme, with two novel quantum ingredients. First, instead of letting a Markov chain $P_i$ mix towards its stationary distribution $\pi_i$, we choose to approximately prepare the state $|\pi_i\rangle = \sum_\sigma \sqrt{\pi_i(\sigma)}\,|\sigma\rangle$, a coherent encoding of the Boltzmann distribution. We use a preparation method [8] based on Grover's $\frac{\pi}{3}$-fixed-point search [11], efficiently driving the state $|\pi_0\rangle$ towards the desired state $|\pi_i\rangle$ through a sequence of intermediate states.

Second, instead of using classical samples from the distribution $\pi_i$, we approximate $\alpha_i$ by phase-estimation of a certain unitary on the state $|\pi_i\rangle$. This is a new concept, going beyond our previous work [8]. 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 $\alpha_i$
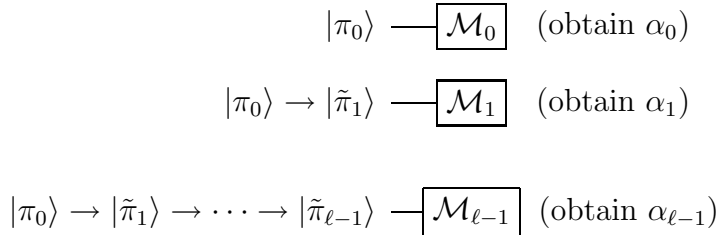
6

$$|\pi_0\rangle \longrightarrow \boxed{\mathcal{M}_0} \quad \text{(obtain } \alpha_0)$$

$$|\pi_0\rangle \rightarrow |\tilde{\pi}_1\rangle \longrightarrow \boxed{\mathcal{M}_1} \quad \text{(obtain } \alpha_1)$$

$$|\pi_0\rangle \rightarrow |\tilde{\pi}_1\rangle \rightarrow \cdots \rightarrow |\tilde{\pi}_{\ell-1}\rangle \longrightarrow \boxed{\mathcal{M}_{\ell-1}} \quad \text{(obtain } \alpha_{\ell-1})$$

Figure 1: Structure of the quantum algorithm.

(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 Section 3.2, we explain how to quantize the the classical algorithm in the perfect case, assuming that we can take perfect samples $X_i$ from $\pi_i$. Then, in Section 3.3 we release this assumption and describe the full quantum algorithm.

## 3.2   Perfect Case

To estimate the ratios $\alpha_i$ in (4), the classical algorithm generates random states $X_i$ from $\pi_i$ and computes the mean $\overline{Y}_i$ of the random variables $Y_i$. The process of generating a random state $X_i$ from $\pi_i$ is equivalent to preparing the mixed state

$$\rho_i = \sum_{\sigma \in \Omega} \pi_i(\sigma) |\sigma\rangle\langle\sigma| \, . \tag{20}$$

Instead of this, we choose to prepare the pure states

$$|\pi_i\rangle = \sum_{\sigma \in \Omega} \sqrt{\pi_i(\sigma)} |\sigma\rangle \, . \tag{21}$$

We call these states *quantum samples* since they coherently encode the probability distributions $\pi_i$. 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

$$A_i = \sum_{\sigma \in \Omega} y_i(\sigma) |\sigma\rangle\langle\sigma| \tag{22}$$

in the state $\rho_i$, where

$$y_i(\sigma) = e^{-(\beta_{i+1} - \beta_i)E(\sigma)} \, . \tag{23}$$

With this interpretation in mind, we see that to estimate $\alpha_i$ classically, we need to estimate the expected value $\text{Tr}(A_i\rho_i)$ 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

$$V_i = \sum_{\sigma \in \Omega} |\sigma\rangle \langle \sigma| \otimes \begin{pmatrix} \sqrt{y_i(\sigma)} & \sqrt{1 - y_i(\sigma)} \\ -\sqrt{1 - y_i(\sigma)} & \sqrt{y_i(\sigma)} \end{pmatrix} . \tag{24}$$

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

$$|\psi_i\rangle = V_i \big( |\pi_i\rangle \otimes |0\rangle \big). \tag{25}$$

Consider now the expected value of the projector

$$P = \mathbb{I} \otimes |0\rangle \langle 0| \tag{26}$$

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

$$\langle \psi_i | P | \psi_i \rangle = \langle \pi_i | A_i | \pi_i \rangle = \alpha_i . \tag{27}$$

We now show how to speed up the process of estimating $\alpha_i$ with a method that generalizes quantum counting [9]. 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

$$R_i = 2|\pi_i\rangle\langle\pi_i| - \mathbb{I} . \tag{28}$$

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*

$$\Pr\left[(1 - \epsilon)Z \leq Q \leq (1 + \epsilon)Z\right] \geq \frac{3}{4}. \tag{29}$$

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

$$O\left(\log \ell\right) \tag{30}$$

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

$$O\left(\frac{\ell}{\epsilon} \log \ell\right) \tag{31}$$

*times, where $R_i$ is as in (28).*

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 $\alpha_i$. Its output $Q'_i$ satisfies*

$$\Pr\left[(1 - \epsilon_{pe})\alpha_i \leq Q'_i \leq (1 + \epsilon_{pe})\alpha_i\right] \geq \frac{7}{8}. \tag{32}$$

*The scheme $\mathcal{A}'_i$ requires one copy of the quantum sample $|\pi_i\rangle$ and invokes the controlled-$R_i$ operator $O\left(\epsilon_{pe}^{-1}\right)$ times, where $R_i$ is as in (28).*

*Proof.* Let

$$G = (2|\psi_i\rangle\langle\psi_i| - \mathbb{I})(2P - \mathbb{I}). \tag{33}$$

Define the basis states

$$|\gamma_1\rangle = \frac{(\mathbb{I} - P)|\psi_i\rangle}{\sqrt{1 - \alpha_i}}, \quad \text{and} \quad |\gamma_2\rangle = \frac{P|\psi_i\rangle}{\sqrt{\alpha_i}}. \tag{34}$$

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

$$G\big|_{\{|\gamma_1\rangle, |\gamma_2\rangle\}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{35}$$

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

$$\cos\theta = 2\alpha_i - 1. \tag{36}$$

The eigenvectors and eigenvalues of $G$ are

$$|G_{\pm}\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \pm i \end{bmatrix}, \quad \lambda_{\pm} = e^{\pm i\theta}. \tag{37}$$

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 $\theta$, and sometimes an estimate of $2\pi - \theta$. However, this is not a problem since both $\theta$ and $2\pi - \theta$ plugged into (36) yield the same result for $\alpha_i$.

We require that the estimate $\theta'$ satisfies

$$|\theta' - \theta| \leq 2\epsilon_{pe}\,\alpha_i \leq \epsilon_{pe} \tag{38}$$

with probability at least $\frac{7}{8}$. Using the phase estimation circuit in [10], this means that $\frac{\theta'}{2\pi}$ has to be an $n_a = \log\frac{2\pi}{\epsilon_{pe}}$ bit approximation of the phase and the failure probability $p_f$ has to be less than $\frac{1}{8}$. To achieve this, it suffices to use a phase estimation circuit (see Fig. 2) with

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

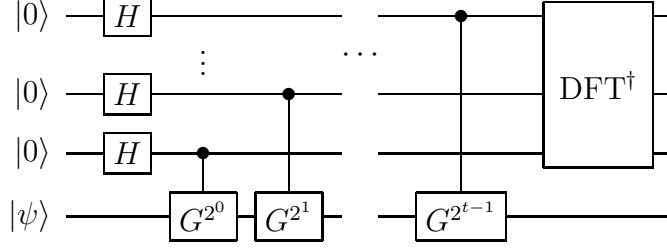ancilla qubits. This circuit invokes the controlled-$G$ operation $O(2^t) = O\left(\epsilon_{pe}^{-1}\right)$ times.

Figure 2: A basic phase estimation circuit with $t$ ancilla qubits

Let $\alpha_i'$ denote the value we compute from the estimate $\theta'$. We have

$$|\alpha_i - \alpha_i'| = \frac{1}{2}\left|\cos\theta - \cos\theta'\right| \leq \frac{1}{2}\left|\theta - \theta'\right| \leq \epsilon_{pe}\,\alpha_i\,, \tag{39}$$

showing that the estimate $\alpha_i'$ is within $\pm\epsilon_{pe}\alpha_i$ of the exact value $\alpha_i$ 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. $\square$

We can boost the success probability of the above quantum approximation scheme for the ratio $\alpha_i$ 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*

$$\Pr\left[(1 - \epsilon_{pe})q \leq W \leq (1 + \epsilon_{pe})q\right] \geq 1 - \delta_{boost}\,. \tag{40}$$

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

With the help of Lemma 2, we now have the constituents required to compose the individual estimates of $\alpha_i$ 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*

$$\Pr\left[\left(1 - \frac{\epsilon}{2\ell}\right)\alpha_i \leq Q_i \leq \left(1 + \frac{\epsilon}{2\ell}\right)\alpha_i\right] \geq 1 - \frac{1}{4\ell}\,. \tag{41}$$

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

$$\Pr\left[(1 - \epsilon)\alpha \leq Q \leq (1 + \epsilon)\alpha\right] \geq \frac{3}{4}\,. \tag{42}$$

10

*Proof.* For each $i = 0, \ldots, \ell - 1$, the failure probability for estimating $\alpha_i$ 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 succes probability in (42).

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

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

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

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

$$
\frac{\alpha - Q}{\alpha} \leq 1 - \prod_{i=0}^{\ell-1} \left(1 - \frac{\epsilon}{2\ell}\right) \leq \sum_{i=0}^{\ell-1} \frac{\epsilon}{2\ell} \leq \epsilon, \tag{43}
$$

where we have used $|\prod_i x_i - \prod_i y_i| \leq \sum_i |x_i - y_i|$, true for arbitrary $x_i, y_i \in [0, 1]$. Thus, in the case of success, we have $(1 - \epsilon)\alpha \leq Q$. $\square$

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 $|\psi_i\rangle$ (25) and the projector $P$ (26). This gives us a quantum approximation scheme for $\alpha_i$. 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 $\alpha_i$ 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

$$
\Pr\left[(1 - \epsilon)\alpha \leq Q \leq (1 + \epsilon)\alpha\right] \geq \frac{3}{4}. \tag{44}
$$

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 $\log \delta_{boost}^{-1} = O(\log \ell)$ copies of the state $|\pi_i\rangle$, and invokes $\left(\log \delta_{boost}^{-1}\right) \epsilon_{pe}^{-1} = O\left(\frac{\ell}{\epsilon} \log \ell\right)$ reflections around $|\pi_i\rangle$. $\square$

## 3.3   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 walks 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 [8], two of us 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 [11] to drive the starting state $|\pi_0\rangle$ towards 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 Section 3.2 that we have to drop here. Nevertheless, an efficient way to apply these phase shifts approximately, based on quantum walks and phase estimation, exists [8].

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 [8], 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

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

The following lemma then directly follows from the arguments used in [8, 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*

$$
\| |\tilde{\pi}_i\rangle - |\pi_i\rangle|0\rangle^{\otimes a} \| \leq \epsilon_S \, , \tag{45}
$$

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

$$
O\left( \frac{\ell}{\sqrt{\delta}} \log^2 \frac{\ell}{\epsilon_S} \right) \, . \tag{46}
$$

*times.*

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

$$
O\left( \frac{\ell}{\sqrt{\delta}} \log^2 \ell \right) \tag{47}
$$

12

applications of the quantum walk operator. Recall that when we used Lemma 1 in Section 3.2 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 [8, Lemma 2 and Corollary 2].

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

$$\tilde{R}_i\left(|\varphi\rangle \otimes |0\rangle^{\otimes b}\right) = \left(R_i|\varphi\rangle\right) \otimes |0\rangle^{\otimes b} + |\xi\rangle \tag{48}$$

*where $|\varphi\rangle$ is an arbitrary state, $b = O\left(\log \epsilon_R^{-1} \log \frac{1}{\sqrt{\delta}}\right)$ 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 $\{W(P_1), \ldots, W(P_{\ell-1})\}$*

$$O\left(\frac{1}{\sqrt{\delta}} \log \frac{1}{\epsilon_R}\right) \tag{49}$$

*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

$$O\left(\frac{1}{\epsilon_{pe}}\right) \cdot \epsilon_R. \tag{50}$$

We choose

$$\epsilon_R = \Omega(\epsilon_{pe}) \tag{51}$$

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 Section 3.2:

**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 $\alpha_i$. Its estimate $Q_i''$ satisfies*

$$\Pr\left[(1 - \epsilon_{pe})\alpha_i \leq Q_i'' \leq (1 + \epsilon_{pe})\alpha_i\right] \geq \frac{3}{4}. \tag{52}$$

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

$$O\left(\frac{\ell}{\sqrt{\delta}} \log^2 \ell + \frac{1}{\epsilon_{pe}\sqrt{\delta}} \log \epsilon_{pe}^{-1}\right). \tag{53}$$

*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 Section 3.2.

*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 (53) by the factor $O(\log \ell)$. 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

$$\Pr\left[(1 - \epsilon)Z \leq Q \leq (1 + \epsilon)Z\right] \geq \frac{3}{4}. \tag{54}$$

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

$$O\left(\frac{\ell^2}{\sqrt{\delta}} \log^3 \ell + \frac{\ell^2}{\epsilon\sqrt{\delta}}(\log \ell)\left(\log \ell + \log \epsilon^{-1}\right)\right) = \tilde{O}\left(\frac{\ell^2}{\epsilon\sqrt{\delta}}\right). \tag{55}$$

□

# 4  Conclusions

We have shown how to quantumly speed up a classical FPRAS for approximating partition functions, as measured in the number of times we have to invoke 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, $\log \ell$, 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 [12] 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 $\alpha_i$ (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 $i$,

$$\frac{\mathbf{E}\left(Y_i^2\right)}{\left(\mathbf{E}\left(Y_i\right)\right)^2} \leq b, \tag{56}$$

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 $\mathbf{E}\left(Y_i\right)$ 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 (56) is satisfied. It seems that the condition $\mathbf{E}\left(Y_i\right) > c^{-1}$ with $c$ some constant is absolutely necessary for phase estimation to yield the quadratic speed-up with respect to the accuracy parameter $\epsilon$.

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 [7]. We therefore plan to explore where our techniques can be used to speed up this breakthrough classical algorithm.

# 5   Acknowledgments

# A   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 $\pi_i$. Here we briefly review the quantum analogue 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 $\Omega$, 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 [2], for each such Markov Chain, we can define its quantum analogue. 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

$$|p_x\rangle = \sum_{y \in \Omega} \sqrt{p_{xy}} \, |y\rangle \, . \tag{57}$$

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

$$U \, |x\rangle \, |0\rangle = |x\rangle \, |p_x\rangle \tag{58}$$

for some fixed state $0 \in \Omega$ and all $x \in \Omega$. The quantum analogue 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*

$$W(P) = R_{\mathcal{B}} \cdot R_{\mathcal{A}}, \tag{59}$$

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

$$\begin{align}
\mathcal{A} &= \mathrm{span}\{|x\rangle \, |0\rangle : x \in \Omega\}, \tag{60}\\
\mathcal{B} &= U^{\dagger} S U \mathcal{A}, \tag{61}
\end{align}$$

*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 [8] easier. It is equivalent to the standard definition of Szegedy [2] up to conjugation by $U$. Therefore, the spectral properties of our $W$ and Szegedy's quantum walk are the same.

Let $\delta$ 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 $\theta_j \in \left(0, \frac{\pi}{2}\right)$. According to Szegedy [2], 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 $\theta_1$ the smallest of $\theta_j$). When the Markov chain is ergodic and reversible, Szegedy proved that

$$\Delta \geq 2\sqrt{\delta}, \tag{62}$$

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

# References

[1] M. Santha, *Quantum Walk Based Search Algorithms*, Proc. of 5th Theory and Applications of Models of Computation (TAMC08), Lectures Notes on Computer Science, vol. 4978, pp. 31–46, 2008.

[2] M. Szegedy, *Quantum Speed-up of Markov Chain Based Algorithms*, Proc. of 45th Annual IEEE Symposium on Foundations of Computer Science, pp. 32–41, 2004.

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

[4] L. Lovász and S. Vempala, *Simulated Annealing in Convex Bodies and an $O^*(n^4)$ Volume Algorithm*, Journal of Computer and System Sciences, vol. 72, issue 2, pp. 392–417, 2006.

[5] M. Jerrum, A. Sinclair, and E. Vigoda, *A Polynomial-Time Approximation Algorithm for the Permanent of a Matrix Non-Negative Entries*, Journal of the ACM, vol. 51, issue 4, pp. 671–697, 2004.

[6] M. Jerrum and A. Sinclair, *Polynomial-Time Approximation Algorithms for the Ising Model*, SIAM Journal on Computing, vol. 22, pp. 1087–1116, 1993.

[7] I. Bezáková, D. Štefankovič, V. Vazirani and E. Vigoda, *Accelerating Simulated Annealing for the Permanent and Combinatorial Counting Problems*, SIAM Journal on Computing, vol. 37, No. 5, pp. 1429–1454, 2008.

[8] P. Wocjan and A. Abeyesinghe, *Speed-up via Quantum Sampling*, Physical Review A, vol. 78, pp. 042336, 2008.

[9] G. Brassard, P. Høyer, and A. Tapp, *Quantum Counting*, Proc. of 25th International Colloquium on Automata, Languages and Programming, Lecture Notes in Computer Science, vol. 1443, pp. 820–831, 1998. quant-ph/9805082

[10] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.

[11] L. K. Grover *A Different Kind of Quantum Search*, arXiv:quant-ph/0503205

[12] P. Richter, *Quantum Speed-Up of Classical Mixing Processes*, Physical Review A, vol. 76, 042306, 2007.

[13] R. Somma, S. Boixo, and H. Barnum, *Quantum Simulated Annealing*, arXiv:0712.1008

[14] R. Somma, S. Boixo, H. Barnum, E. Knill, *Quantum Simulations of Classical Annealing Processes*, arXiv:0804.1571

[15] M. Jerrum, L. Valiant and V. Vazirani, *Random Generation of Combinatorial Structures from a Uniform Distribution*, Theoretical Computer Science, vol. 43, issue 2-3, pp. 169–188, 1986.

[16] D. Štefankovič, S. Vempala, and E. Vigoda, *Adaptive Simulated Annealing: A Near-Optimal Connection between Sampling and Counting*, Proc. of the 48th Annual IEEE Symbosium on Foundations of Computer Science, pp. 183-193, 2007.