Quantum Walks on the Hypercube

CRISTOPHER MOORE
Computer Science Department
University of New Mexico, Albuquerque
and the Santa Fe Institute, Santa Fe, New Mexico
moore@cs.unm.edu

ALEXANDER RUSSELL
Department of Computer Science and Engineering
University of Connecticut
Storrs, Connecticut
acr@cse.uconn.edu

April 1, 2007

Abstract

Recently, it has been shown that one-dimensional quantum walks can mix more quickly than classical random walks, suggesting that quantum Monte Carlo algorithms can outperform their classical counterparts. We study two quantum walks on the n-dimensional hypercube, one in discrete time and one in continuous time. In both cases we show that the quantum walk mixes in $(\pi/4)n$ steps, faster than the $\Theta(n \log n)$ steps required by the classical walk. In the continuous-time case, the probability distribution is *exactly* uniform at this time. More importantly, these walks expose several subtleties in the definition of mixing time for quantum walks. Even though the continuous-time walk has an $O(n)$ instantaneous mixing time at which it is precisely uniform, it never approaches the uniform distribution when the stopping time is chosen randomly as in [AAKV01]. Our analysis treats interference between terms of different phase more carefully than is necessary for the walk on the cycle; previous general bounds predict an exponential, rather than linear, mixing time for the hypercube.

1 Introduction

Random walks form one of the cornerstones of theoretical computer science. As algorithmic tools, they have been applied to a variety of central problems, such as estimation of the volume of a convex body [DFK91, LK99], approximation of the permanent [JS89, JSV00], and discovery of satisfying assignments for Boolean formulae [Sch99]. Furthermore, the basic technical phenomena appearing in the study of random walks (e.g., spectral decomposition, couplings, and Fourier analysis) also support several other important areas such as pseudorandomness and derandomization (see, e.g., [AS92, (§9,§15)]).

The development of efficient *quantum* algorithms for problems believed to be intractable for (classical) randomized computation, like integer factoring and discrete logarithm [Sho97], has prompted the investigation of *quantum walks*. This is a natural generalization of the traditional notion discussed above where, roughly, the process evolves in a unitary rather than stochastic fashion.

The notion of "mixing time," the first time when the distribution induced by a random walk is sufficiently close to the stationary distribution, plays a central role in the theory of classical random walks. For a given graph, then, it is natural to ask if a quantum walk can mix more quickly than its classical counterpart. (Since a unitary process cannot be mixing, we define a stochastic process from a quantum one by performing a measurement at a given time or a distribution of times.) Several recent articles [AAKV01, ABN$^+$01, NV00] have answered this question in the affirmative, showing, for example, that a quantum walk on the n-cycle mixes in time $O(n \log n)$, a substantial improvement over the classical random walk which requires $\Theta(n^2)$ steps to mix. Quantum walks were also defined in [Wat01], and used to show that undirected graph connectivity is contained in a version of quantum LOGSPACE. These articles raise the exciting possibility

that quantum Monte Carlo algorithms could form a new family of quantum algorithms that work more quickly than their classical counterparts.

Two distinct notions of quantum walks exist in the literature. The first, introduced by [AAKV01, ABN$^+$01, NV00], studies the behavior of a "directed particle" on the graph; we refer to these as *discrete-time* quantum walks. The second, introduced by [CFG01], defines the dynamics by treating the adjacency matrix of the graph as a Hamiltonian; we refer to these as *continuous-time* quantum walks. The landscape is further complicated by the existence of two distinct notions of mixing time. The first "instantaneous" notion [ABN$^+$01, NV00] focuses on particular times at which measurement induces a desired distribution; the second "average" notion [AAKV01], another natural way to convert a quantum process into a stochastic one, focuses on measurement times selected at random.

In this article, we analyze both the continuous-time and a discrete-time quantum walk on the hypercube. In both cases, the walk is shown to have an instantaneous mixing time at $(\pi/4)n$. Recall that the classical walk on the hypercube mixes in time $\Theta(n \log n)$, so that the quantum walk is faster by a logarithmic factor. Moreover, in the discrete-time case the walk mixes in time less than the diameter of the graph, since $\pi/4 < 1$; and, astonishingly, in the continuous-time case the probability distribution at $t = (\pi/4)n$ is *exactly* uniform. Both of these things happen due to a marvelous conspiracy of destructive interference between terms of different phase.

These walks show *i.)* a similarity between the two notions of quantum walks, and *ii.)* a disparity between the two notions of quantum mixing times. As mentioned above, both walks have an instantaneous mixing time at time $(\pi/4)n$. On the other hand, we show that there is *no* time at which the continuous walk approaches the uniform distribution in the sense of [AAKV01]. Thus there are some real subtleties involved in defining mixing times for quantum walks.

The analysis of the hypercubic quantum walk exhibits a number of features markedly different from those appearing in previously studied walks. In particular, the dimension of the relevant Hilbert space is, for the hypercube, exponential in the length of the desired walk, while in the cycle these quantities are roughly equal. This requires that interference be handled in a more delicate way than is required for the walk on the cycle; in particular, the general bound of [AAKV01] predicts an exponentially large mixing time for the discrete-time walk.

We begin by defining quantum walks and discussing various notions of mixing time. We then analyze the two quantum walks on the hypercube in Sections 2 and 3. (Most of the technical details for the discrete-time walk are relegated to an appendix.) Finally, in Section 4, we discuss mixing times in the sense of [AAKV01].

1.1 Quantum walks and mixing times

Any graph $G = (V, E)$ gives rise to a familiar Markov chain by assigning probability $1/d$ to all edges leaving each vertex v of degree d. Let $P_u^t(v)$ be the probability of visiting a vertex v at step t of the random walk on G starting at u. If G is undirected, connected, and not bipartite, then $\lim_{t\to\infty} P_u^t$ exists[1] and is independent of u. A variety of well-developed techniques exist for establishing bounds on the rate at which P_u^t achieves this limit (e.g., [Vaz92]); if G happens to be the Cayley graph of a group (as are, for example, the cycle and the hypercube), then techniques from Fourier analysis can be applied (see [Dia88]). Below we will use some aspects of this approach, especially the Diaconis-Shahshahani bound on the total variation distance [DS81].

For simplicity, we restrict our discussion to quantum walks on Cayley graphs; more general treatments of quantum walks appear in [AAKV01, CFG01]. Before describing the quantum walk models we set down some notation.

[1] In fact, this limit exists under more general circumstances; see e.g. [MR95].

Notation. For a group G and a set of generators Γ such that $\Gamma = \Gamma^{-1}$, we let $X(G,\Gamma)$ denote the undirected Cayley graph of G with respect to Γ. For a finite set S, we let $L(S) = \{f : S \to \mathbb{C}\}$ denote the collection of $\mathbb{C}$-valued functions on S. This is a Hilbert space under the natural inner product $\langle f|g\rangle = \sum_{s\in S} f(s)\,g(s)^*$. For a Hilbert space V, an operator $U : V \to V$ is *unitary* if for all $\vec{v},\vec{w} \in V$, $\langle \vec{v}|\vec{w}\rangle = \langle U\vec{v}|U\vec{w}\rangle$; if U is represented as a matrix, this is equivalent to the condition that $U^\dagger = U^{-1}$ where $\dagger$ denotes the Hermitian conjugate.

There are two natural quantum walks that one can define for such graphs, which we now describe.

The discrete-time walk. This model, introduced by [AAKV01, ABN+01, NV00], augments the space $L(G)$ with a *direction space*, each basis vector of which corresponds one of the generators in Γ. A step of the walk then consists of the composition of two unitary transformations; a *shift* operator which leaves the direction unchanged while moving the particle in the appropriate direction, and a *local transformation* which operates on the direction while leaving the position unchanged. To be precise, the quantum walk on $X(G,\Gamma)$ is defined on the space $L(G\times\Gamma) \cong L(G)\otimes L(\Gamma)$. Let $\{\delta_\gamma \mid \gamma\in\Gamma\}$ be the natural basis for $L(\Gamma)$, and $\{\delta_g | g \in G\}$ the natural basis for $L(G)$. Then the shift operator is $S : (\delta_g \otimes \delta_\gamma) \mapsto (\delta_{g\gamma}\otimes\delta_\gamma)$, and the local transformation is $\check{D} = \mathbf{1}\otimes D$ where D is defined on $L(\Gamma)$ alone and $\mathbf{1}$ is the identity on $L(G)$. Then one "step" of the walk corresponds to the operator $U = \check{D}V$. If we measure the position of the particle, but not its direction, at time t, we observe a vertex v with probability $P_t(v) = \sum_{\gamma\in\Gamma} \left|\left\langle U^t\psi_0 \mid \delta_v\otimes\delta_\gamma\right\rangle\right|^2$ where $\psi_0 \in L(G\times\Gamma)$ is the initial state.

The continuous-time walk. This model, introduced by [CFG01], works directly with $L(G)$, the Hilbert space of $\mathbb{C}$-valued functions on G: $L(G) = \{f : G \to \mathbb{C}\}$. The walk evolves by treating the adjacency matrix of the graph as a Hamiltonian and using the Schrödinger equation. Specifically, if H is the adjacency matrix of $X(G,\Gamma)$, the evolution of the system at time t is given by U_t, where $U_t \stackrel{\text{eq}}{=} e^{iHt}$ (here we use the matrix exponential, and U_t is unitary since H is real and symmetric). Then if we measure the position of the particle at time t, we observe a vertex v with probability $P_t(v) = |\langle U_t\psi_0|e_v\rangle|^2$ where ψ_0 is the initial state.

In both cases we start with an initial wave function concentrated at a single vertex u. For the continuous-time walk, this corresponds to a wave function

$$\psi_u(v) = \langle\psi_u|\delta_v\rangle = \begin{cases} 1 & \text{if } u = v, \\ 0 & \text{otherwise.}\end{cases}$$

For the discrete-time walk, we start with a uniform superposition over all possible directions,

$$\psi_u(v,\gamma) = \left\langle\psi_u\right| e_v\otimes e_\gamma\rangle = \begin{cases} 1/\sqrt{|\Gamma|} & \text{if } u = v, \\ 0 & \text{otherwise.}\end{cases}$$

In order to define a discrete quantum walk, one must select a local operator D on the direction space. In principle, this introduces some arbitrariness into the definition. However, if we wish D to respect the permutation symmetry of the n-cube, and if we wish to maximize the operator distance between D and the identity, we show in Appendix A that we are forced to choose Grover's diffusion operator [Gro96], which we recall below. We call the resulting walk the "symmetric discrete-time quantum walk" on the n-cube. (Watrous [Wat01] also used Grover's operator to define quantum walks on undirected graphs.)

(Since for large n Grover's operator is close to the identity matrix, one might imagine that it would take $\Omega(n^{1/2})$ steps to even change direction, giving the quantum walk a mixing time of $\approx n^{3/2}$, slower than the classical random walk. However, like many intuitions about quantum mechanics, this is simply wrong.)

Since the evolution of the quantum walk is governed by a unitary operator rather than a stochastic one, unless P_t is constant for all t, there can be no "stationary distribution" $\lim_{t\to\infty} P_t$. In particular, for any $\varepsilon > 0$,

there are infinitely many (positive, integer) times t for which $\|U^t - \mathbf{1}\| \leq \varepsilon$ so that $\|U^t\psi_u - \psi_u\| \leq \varepsilon$ and P_t is close to the initial distribution. However, there may be particular stopping times t which induce distributions close to, say, the uniform distribution, and we call these *instantaneous mixing times*:

Definition 1 *We say that t is an* ε-instantaneous mixing time *for a quantum walk if $\|P_t - U\| \leq \varepsilon$, where*

$$\|A - B\| = \frac{1}{2}\sum_v |A(v) - B(v)|$$

denotes total variation distance and U denotes the uniform distribution.

For these walks we show:

Theorem 1 *For the symmetric discrete-time quantum walk on the n-cube, $t = \lceil k(\pi/4)n \rceil$ is an ε-instantaneous mixing time with $\varepsilon = O(n^{-7/6})$ for all odd k.*

and, even more surprisingly,

Theorem 2 *For the continuous-time quantum walk on the n-hypercube, $t = k(\pi/4)n$ is a 0-instantaneous mixing time for all odd k.*

Thus in both cases the mixing time is $\Theta(n)$, as opposed to $\Theta(n \log n)$ as it is in the classical case.

Aharonov et al. [AAKV01] define another natural notion of mixing time for quantum walks, in which the stopping time t is selected uniformly from the set $\{0, \ldots, T-1\}$. They show that the distributions $\bar{P}_T = \frac{1}{T}\sum_{t=0}^{T-1} P_t$ *do* converge as $T \to \infty$ and study the rate at which this occurs. For a continuous random walk, we analogously define the distribution $\bar{P}_T(v) = (1/T)\int_{0,T} P_t(v)\,\mathrm{d}t$. Then we call a time at which the resulting distribution $\bar{P}_T$ is close to uniform an *average mixing time*:

Definition 2 *We say that T is an* ε-average mixing time *for a quantum walk if $\|\bar{P}_T - U\| \leq \varepsilon$.*

The exact relationship between instantaneous and average mixing times is unclear. In fact, while the continuous walk on the hypercube possesses 0-instantaneous mixing times at all odd multiples of $(\pi/4)n$, the limiting distribution of $\bar{P}_T$ is *not* the uniform distribution, and we will show that an $\varepsilon > 0$ exists such that *no* time is an ε-average mixing time. For the discrete-time walk, the limiting distribution *is* uniform and we show that the general bound given in [AAKV01] predicts an exponential, rather than linear, average mixing time for the hypercube.

2 The symmetric discrete-time walk

In this section we prove Theorem 1. We treat the n-cube as the Cayley graph of $\mathbb{Z}_2^n$ with the regular basis vectors $\vec{e}_i = (0, \ldots, 1, \ldots, 0)$ with the 1 appearing in the ith place. Then the discrete-time walk takes place in the Hilbert space $L(\mathbb{Z}_2^n \times [n])$ where $[n] = \{1, \ldots, n\}$. Here the first component represents the position of the particle in the hypercube, and the second component represents the "direction" currently associated with the particle.

As in [AAKV01, NV00], we will not impose a group structure on the direction space, and will Fourier transform only over the position space. For this reason, we will express an element ψ in $L(\mathbb{Z}_2^n) \otimes L([n])$ as a function $\Psi : \mathbb{Z}_2^n \to \mathbb{C}^n$, where the ith coordinate of $\Psi(\vec{x})$ is the projection of ψ into $\delta_{\vec{x}} \otimes \delta_i$, i.e. the complex amplitude of the particle being at position $\vec{x}$ with direction i. The Fourier transform of such an element Ψ is $\tilde{\Psi} : \mathbb{Z}_2^n \to \mathbb{C}^n$, where

$$\tilde{\Psi}(\vec{k}) = \sum_{\vec{x}} (-1)^{\vec{k}\cdot\vec{x}} \Psi(\vec{x}).$$

Then the shift operator for the hypercube is

$$S:\Psi(x)\mapsto\sum_{i=1}^{n}\pi_i\Psi(\vec{x}\oplus\vec{e}_i)$$

where $\vec{e}_i$ is the ith basis vector in the n-cube, and π_i is the projection operator for the ith direction. The reason for considering the Fourier transform above is that the shift operator is locally diagonal in this basis: specifically it maps $\tilde{\Psi}(\vec{k})\mapsto S_{\vec{k}}\tilde{\Psi}(\vec{k})$ where

$$S_{\vec{k}}=\begin{pmatrix}(-1)^{k_1} & & & 0\\ & (-1)^{k_2} & & \\ & & \ddots & \\ 0 & & & (-1)^{k_n}\end{pmatrix}$$

For the local transformation, we use Grover's diffusion operator on n states, $D_{ij}=2/n-\delta_{ij}$.

The advantage of Grover's operator is that, like the n-cube itself, it is permutation symmetric. We use this symmetry to rearrange $U_{\vec{k}}=S_{\vec{k}}D$ to put the negated rows on the bottom,

$$U_{\vec{k}}=\left(\begin{array}{ccc|ccc}2/n-1 & 2/n & \cdots & & 2/n & \\ 2/n & 2/n-1 & & & & \\ \vdots & & \ddots & & & \\ \hline & & & -2/n+1 & -2/n & \cdots \\ & -2/n & & -2/n & -2/n+1 & \\ & & & \vdots & & \ddots\end{array}\right)$$

where the top and bottom blocks have $n-k$ and k rows respectively; here k is the Hamming weight of $\vec{k}$.

The eigenvalues of $U_{\vec{k}}$ then depend only on k. Specifically, $U_{\vec{k}}$ has the eigenvalues $+1$ and -1 with multiplicity $k-1$ and $n-k-1$ respectively, plus the eigenvalues λ,λ^* where

$$\lambda=1-\frac{2k}{n}+\frac{2i}{n}\sqrt{k(n-k)}=e^{i\omega_k}$$

and $\omega_k\in[0,\pi]$ is described by

$$\cos\omega_k=1-\frac{2k}{n},\quad \sin\omega_k=\frac{2}{n}\sqrt{k(n-k)}$$

Its eigenvectors with eigenvalue $+1$ span the $(k-1)$-dimensional subspace consisting of vectors with support on the k "flipped" directions that sum to zero, and similarly the eigenvectors with eigenvalue -1 span the $(n-k-1)$-dimensional subspace of vectors on the $n-k$ other directions that sum to zero. We call these the *trivial* eigenvectors. The eigenvectors of $\lambda,\lambda^*=e^{\pm i\omega_k}$ are

$$v_k,v_k^*=\frac{1}{\sqrt{2}}\Big(\underbrace{\frac{\mp i}{\sqrt{n-k}}}_{n-k},\underbrace{\frac{1}{\sqrt{k}}}_{k}\Big).$$

We call these the *non-trivial* eigenvectors for a given $\vec{k}$. Over the space of positions and directions these eigenvectors are multiplied by the Fourier coefficient $(-1)^{\vec{k}\cdot\vec{x}}$, so as a function of $\vec{x}$ and direction $1\le j\le n$ the two non-trivial eigenstates of the entire system, for a given $\vec{k}$, are

$$v_{\vec{k}}(\vec{x},j)=(-1)^{\vec{k}\cdot\vec{x}}\frac{2^{-n/2}}{\sqrt{2}}\times\begin{cases}1/\sqrt{k} & \text{if } \vec{k}_j=1\\ -i/\sqrt{n-k} & \text{if } \vec{k}_j=0\end{cases}$$

with eigenvalue $e^{i\omega_k}$, and its conjugate $v_{\vec{k}}^*$ with eigenvalue $e^{-i\omega_k}$.

We take for our initial wave function a particle at the origin $u = (0,\ldots,0)$ in an equal superposition of directions. Since its position is a δ-function in real space it is uniform in Fourier space as well as over the direction space, giving

$$\tilde{\Psi}_0(\vec{k}) = \frac{2^{-n/2}}{\sqrt{n}}(1,\ldots,1)$$

This is perpendicular to all the trivial eigenvectors, so their amplitudes are all zero. The amplitude of its component along the non-trivial eigenvector $v_{\vec{k}}$ is

$$a_{\vec{k}} = \langle \Psi_0 | v_{\vec{k}} \rangle = \frac{2^{-n/2}}{\sqrt{2}} \left(\sqrt{\frac{k}{n}} - i\sqrt{1-\frac{k}{n}} \right) \tag{1}$$

and the amplitude of $v_{\vec{k}}^*$ is $a_{\vec{k}}^*$. Note that $|a_{\vec{k}}|^2 = 2^{-n}/2$, so a particle is equally likely to appear in either non-trivial eigenstate with any given wave vector.

At this point, we note that there are an exponential number of eigenvectors in which the initial state has a non-zero amplitude. In Section 4, we show that the general bound of Aharonov et al. [AAKV01] predicts an exponential mixing time. In general, this bound performs poorly whenever the number of important eigenvalues is greater than the mixing time.

Instead, we will use the Diaconis-Shahshahani bound on the total variation distance in terms of the Fourier coefficients of the probability [Dia88]. If $P_t(\vec{x})$ is the probability of the particle being observed at position $\vec{x}$ at time t, and U is the uniform distribution, then the total variation distance is bounded by

$$\|P_t - U\|^2 \leq \frac{1}{4} \sum_{\substack{\vec{k} \neq (0,\ldots,0) \\ \vec{k} \neq (1,\ldots,1)}} \left| \tilde{P}_t(\vec{k}) \right|^2 = \frac{1}{4} \sum_{k=1}^{n-1} \binom{n}{k} \left| \tilde{P}_t(k) \right|^2. \tag{2}$$

Here we exclude both the constant term and the parity term $\vec{k} = (1,\ldots,1)$; since our walk changes position at every step, we only visit vertices with odd or even parity at odd or even times respectively. Thus U here means the uniform distribution with probability 2^{n-1} on the vertices of appropriate parity.

To find $\tilde{P}_t(\vec{k})$, we first need $\tilde{\Psi}_t(\vec{k})$. As Nayak and Vishwanath [NV00] did for the walk on the line, we start by calculating the tth matrix power of $U_{\vec{k}}$. This is

$$U_{\vec{k}}^t = \left(\begin{array}{ccc|ccc} a+(-1)^t & a & \cdots & & & \\ a & a+(-1)^t & & & c & \\ \vdots & & \ddots & & & \\ \hline & & & b-(-1)^t & b & \cdots \\ & -c & & b & b-(-1)^t & \\ & & & \vdots & & \ddots \end{array} \right)$$

where

$$a = \frac{\cos\omega_k t - (-1)^t}{n-k}, \quad b = \frac{\cos\omega_k t + (-1)^t}{k}, \quad \text{and} \quad c = \frac{\sin\omega_k t}{\sqrt{k(n-k)}}$$

Starting with the uniform initial state, the wave function after t steps is

$$\tilde{\Psi}_t(\vec{k}) = \frac{1}{\sqrt{n}} \Big(\underbrace{\cos\omega_k t + \sqrt{\frac{k}{n-k}}\sin\omega_k t}_{n-k}, \underbrace{\cos\omega_k t - \sqrt{\frac{n-k}{k}}\sin\omega_k t}_{k} \Big) \tag{3}$$

We could, at this point, calculate $\Psi_t(\vec{x})$ by Fourier transforming this back to real space. However, this calculation turns out to be significantly more awkward than calculating the Fourier transform of the probability distribution, $\tilde{P}_t(\vec{k})$, which we need to apply the Diaconis-Shahshahani bound. Since $P_t(\vec{x}) = \Psi_t(\vec{x})\Psi_t(\vec{x})^*$, and since multiplications in real space are convolutions in Fourier space, we perform a convolution over $\mathbb{Z}_2^n$:

$$\tilde{P}_t(\vec{k}) = \sum_{\vec{k}'} \tilde{\Psi}_t(\vec{k}') \cdot \tilde{\Psi}_t(\vec{k} \oplus \vec{k}')$$

where the inner product is defined on the direction space, $u \cdot v = \sum_{i=1}^n u_i v_i^*$. We write this as a sum over j, the number of bits of overlap between $\vec{k}'$ and $\vec{k}$, and l, the number of bits of $\vec{k}'$ outside the bits of $\vec{k}$ (and so overlapping with $\vec{k} \oplus \vec{k}'$). Thus $\vec{k}'$ has weight $j+l$, and $\vec{k} \oplus \vec{k}'$ has weight $k-j+l$.

Calculating the dot product $\tilde{\Psi}_t(\vec{k}') \cdot \tilde{\Psi}_t(\vec{k} \oplus \vec{k}')$ explicitly from Equation 3 as a function of these weights and overlaps, we have

$$\tilde{P}_t(k) = \frac{1}{2^n} \sum_{j=0}^{k} \sum_{l=0}^{n-k} \binom{k}{j} \binom{n-k}{l} \left[\cos\omega_{j+l} t \cos\omega_{k-j+l} t + A \sin\omega_{j+l} t \sin\omega_{k-j+l} t \right] \tag{4}$$

where

$$A = \frac{\cos\omega_k - \cos\omega_{j+l}\cos\omega_{k-j+l}}{\sin\omega_{j+l}\sin\omega_{k-j+l}}$$

The reader can check that this gives $\tilde{P}_t(0) = 1$ for the trivial Fourier component where $k = 0$, and $\tilde{P}_t(n) = (-1)^t$ for the parity term where $k = n$.

Using the identities $\cos a \cos b = (1/2)(\cos(a-b) + \cos(a+b))$ and $\sin a \sin b = (1/2)(\cos(a-b) - \cos(a+b))$ we can re-write Equation 4 as

$$\tilde{P}_t(k) = \frac{1}{2^n} \sum_{j=0}^{k} \sum_{l=0}^{n-k} \binom{k}{j} \binom{n-k}{l} \left[\left(\frac{1-A}{2}\right) \cos\omega_+ t + \left(\frac{1+A}{2}\right) \cos\omega_- t \right] = \frac{1}{2^n} \sum_{j=0}^{k} \sum_{l=0}^{n-k} \binom{k}{j} \binom{n-k}{l} Y \tag{5}$$

where $\omega_\pm = \omega_{j+l} \pm \omega_{k-j+l}$.

The terms $\cos\omega_\pm t$ in Y are rapidly oscillating with a frequency that increases with t. Thus, unlike the walk on the cycle, the phase is rapidly oscillating everywhere, as a function of either l or j. This will make the dominant contribution to $\tilde{P}_t(k)$ exponentially small when $t/n = \pi/4$, giving us a small variation distance when we sum over all $\vec{k}$.

To give some intuition for the remainder of the proof, we pause here to note that if Equation 5 were an integral rather than a sum, we could immediately approximate the rate of oscillation of Y to first order at the peaks of the binomials, where $j = k/2$ and $l = (n-k)/2$. One can check that $\mathrm{d}\omega_k/\mathrm{d}k \geq 2/n$ and hence $\mathrm{d}\omega_+/\mathrm{d}l = \mathrm{d}\omega_-/\mathrm{d}j \geq 4/n$. Since $|A| \leq 1$, we would then write

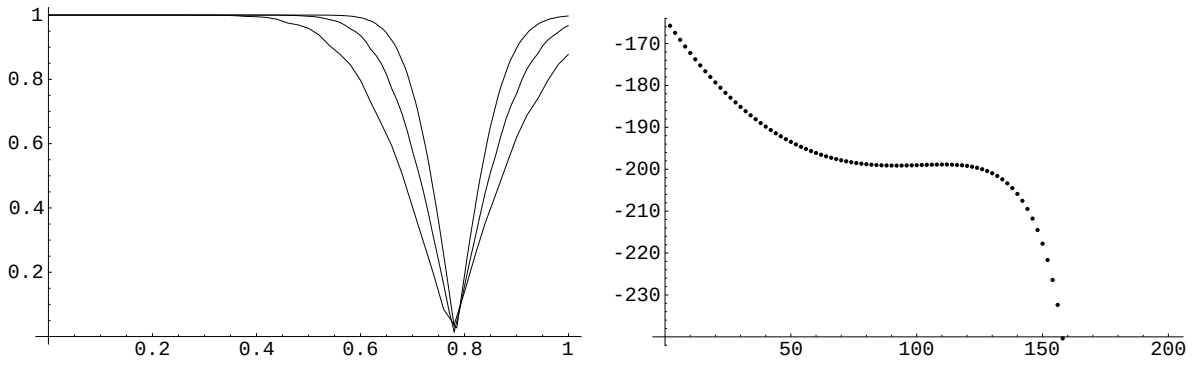
$$\tilde{P}_t(k) =_O \frac{1}{2^n} \sum_{j=0}^{k} \sum_{l=0}^{n-k} \binom{k}{j} \binom{n-k}{l} \left(\mathrm{e}^{4ijt/n} + \mathrm{e}^{4ilt/n} \right)$$

which, using the binomial theorem, would give

$$\left|\tilde{P}_t(k)\right| =_O \left| \frac{1+\mathrm{e}^{4it/n}}{2} \right|^k + \left| \frac{1+\mathrm{e}^{4it/n}}{2} \right|^{n-k} = \cos^k \frac{2t}{n} + \cos^{n-k} \frac{2t}{n} \tag{6}$$

In this case the Diaconis-Shahshahani bound and the binomial theorem give

$$\|P_t - U\|^2 \leq \frac{1}{4} \sum_{0<k<n} \binom{n}{k} \left(\cos^k \frac{2t}{n} + \cos^{n-k} \frac{2t}{n} \right)^2 \leq \frac{1}{2} \left[\left(2\cos^2 \frac{2t}{n} \right)^n + \left(1 + \cos^2 \frac{2t}{n} \right)^n - 1 \right]$$



(a) Variation distance at time t as a function of t/n.

(b) Probability as a function of Hamming weight.

Figure 1: Graph (a) plots an exact calculation of the total variation distance after t steps of the quantum walk for hypercubes of dimension 50, 100, and 200, as a function of t/n. At $t/n = \pi/4$ the variation distance is small even though the walk has not had time to cross the entire graph. This happens because the distribution is roughly uniform across the equator of the n-cube where the vast majority of the points are located. Graph (b) shows the probability distribution on the 200-dimensional hypercube after $157 \approx (\pi/4)n$ steps. The probability distribution has a plateau of 2^{-199} at the equator, matching the uniform distribution up to parity. Shown is the log of the probability as a function of Hamming distance from the starting point.

If we could take t to be the non-integer value $(\pi/4)n$, these cosines would be zero.

This will, in fact, turn out to be the right answer. But since Equation 5 is a sum, not an integral, we have to be wary of *resonances* where the oscillations are such that the phase changes by a multiple of 2π between adjacent terms, in which case these terms will interfere constructively rather than destructively. Thus to show that the first-order oscillation indeed dominates, we have a significant amount of work left to do. The details of managing these resonances can be found in Appendix B. The process can be summarized as follows: *i.)* we compute the Fourier transform of the quantity Y in Equation 5, since the sum of Equation 5 can be calculated for a single Fourier basis function using the binomial theorem; *ii.)* the Fourier transform of Y can be asymptotically bounded by the method of stationary phase. The dominant stationary point corresponds to the first-order oscillation, but there are an infinite number of other stationary points as well; so *iii.)* we use an entropy bound to show that the contribution of the other stationary points is exponentially small.

To illustrate our result, we have calculated the probability distribution, and the total variation distance from the uniform distribution (up to parity), as a function of time for hypercubes of dimension 50, 100, and 200. In order to do this exactly, we use the walk's permutation symmetry to collapse its dynamics to a function only of Hamming distance. In Figure 1(a) we see that the total variation distance becomes small when $t/n = \pi/4$, and in Figure 1(b) we see how the probability distribution is close to uniform on a "plateau" across the hypercube's equator. Since this is where the vast majority of the points are located, the total variation distance is small even though the walk has not yet had time to cross the entire graph.

3 The continuous-time walk

In this section we prove Theorem 2. Childs, Farhi and Gutmann [CFG01] define quantum walks in a different way, in which the unitary operator is generated from a Hamiltonian H using Schrödinger's equation. If H is simply the adjacency matrix of the graph, then $U_t = e^{iHt} = 1 + iHt + (iHt)^2/2 + \cdots$ giving a walk in

continuous time. The amplitude of making s steps is the coefficient $(it)^s/s!$ of H^s, which up to normalization is Poisson-distributed with mean t. They point out that this avoids the need to extend the Hilbert space of the particle with a direction space, and to define some local operation on it such as Grover's operator, in order to make the walk unitary. While this approach is less familiar in computer science, a quantum computer which is allowed to evolve in continuous time according to a certain Hamiltonian seems just as physically reasonable as one which uses a clock to evolve in discrete time as traditional computers do.

In the case of the hypercube, this walk turns out to be particularly easy to analyze. The adjacency matrix, normalized by the degree, is

$$H(\vec{x},\vec{y}) = \begin{cases} 1/n & d(\vec{x},\vec{y}) = 1 \\ 0 & d(\vec{x},\vec{y}) \neq 1 \end{cases} \tag{7}$$

where d is the Hamming distance. The eigenvectors of H and U_t are simply the Fourier basis functions: if $v_{\vec{k}}(\vec{x}) = (-1)^{\vec{k}\cdot\vec{x}}$ then $Hv_{\vec{k}} = (1-2k/n)\,v_{\vec{k}}$ and $U_t v_{\vec{k}} = \mathrm{e}^{it(1-2k/n)}\,v_{\vec{k}}$ where we again use k to denote the Hamming weight of $\vec{k}$. If our initial wave vector has a particle at $\vec{x} = (0,\ldots,0)$, then its initial Fourier spectrum is uniform, and at time t we have

$$\tilde{\Psi}_t(\vec{k}) = 2^{-n/2}\,\mathrm{e}^{it\left(1-\frac{2k}{n}\right)}.$$

Again writing the probability P as the convolution of Ψ with Ψ^* in Fourier space, we have

$$\tilde{P}_t(\vec{k}) = \sum_{\vec{k}'} \tilde{\Psi}_t(\vec{k}')\,\tilde{\Psi}_t^*(\vec{k}\oplus\vec{k}') = \frac{1}{2^n}\sum_{\vec{k}'} \mathrm{e}^{2it\left(|\vec{k}\oplus\vec{k}'|-k'\right)/n}$$

We write this as a sum over all possible overlaps j between $\vec{k}'$ and $\vec{k}$, and overlaps l between $\vec{k}'$ and $\vec{k}\oplus\vec{k}'$. Noting that $k' = j+l$ and $|\vec{k}\oplus\vec{k}'| = k-j+l$, this gives

$$\tilde{P}_t(k) = \frac{1}{2^n}\sum_{j=0}^{k}\sum_{l=0}^{n-k} \mathrm{e}^{2it(k-2j)/n} = \cos^k\frac{2t}{n} \tag{8}$$

Finally, the Diaconis-Shahshahani bound on the total variation distance between P_t and the uniform distribution is

$$\|P_t - U\|^2 \leq \frac{1}{4}\sum_{k=1}^{n}\binom{n}{k}\left|\tilde{P}_t(k)\right|^2 = \left(1+\cos^2\frac{2t}{n}\right)^n - 1$$

Astonishingly, at $t = (\pi/4)n$ and its odd multiples, this gives a total variation distance which is exactly zero, showing that if we sample at these times the probability distribution is *exactly* uniform. Note that this is possible even when $t < n$ since the continuous-time walk has some probability for taking more than t steps (and, in fact, paths with different numbers of steps interfere with each other). Thus the continuous-time walk has the same mixing time as the discrete-time one, but with such a beautiful conspiracy of interference that every position has an identical probability. This concludes the proof of Theorem 2. For an alternative derivation based on hypercube's structure as a product graph, see Appendix C.

4 Average mixing times

In this section we discuss the mixing time as defined in [AAKV01], where we choose to stop the quantum walk at a time t uniformly distributed in the interval $[0,T]$. As mentioned in the Introduction, this gives a probability distribution $\bar{P}_T = (1/T)\sum_{t=0}^{T-1} P_t$. Since the Fourier transform is a linear operation, we can look

at the Fourier transform of $\bar{P}_T$ instead. In the case of the symmetric discrete-time walk, Equation 5 shows that for $k > 0$, the Fourier coefficient of $\bar{P}_T$ consists of a sum of oscillating terms proportional to $\cos \omega_\pm t$. As $T \to \infty$, these oscillations cancel, so we are left with just the constant term $k = 0$ and $\bar{P}_T$ indeed approaches the uniform distribution.

One could calculate an average mixing time for the symmetric discrete-time walk using the methods of Appendix B. We do not do that here. However, we will now show that the general bound of [AAKV01] predicts an average mixing time for the n-cube which is exponential in n. The authors of that paper showed that the variation distance between $\bar{P}_T$ and the uniform distribution (or more generally, the limiting distribution $\lim_{T\to\infty} \bar{P}_T$) is bounded by a sum over distinct pairs of eigenvalues,

$$\|\bar{P}_T - U\| \leq \frac{2}{T} \sum_{i,j \text{ s.t. } \lambda_i \neq \lambda_j} \frac{|a_i|^2}{|\lambda_i - \lambda_j|} \tag{9}$$

where $a_i = \langle \psi_0 | v_i \rangle$ is the component of the initial state along the eigenvector v_i. Since this bound includes eigenvalues λ_j for which $a_j = 0$, we note that it also holds when we replace $|a_i|^2$ with $|a_i a_j^*|$, using the same reasoning as in [AAKV01].

For the quantum walk on the cycle of length n, this bound gives an average mixing time of $O(n \log n)$. For the n-cube, however, there are exponentially many pairs of eigenvectors with distinct eigenvalues, all of which have a non-zero component in the initial state. Specifically, for each Hamming weight k there are $\binom{n}{k}$ non-trivial eigenvectors each with eigenvalue $e^{i\omega_k}$ and $e^{-i\omega_k}$. These complex conjugates are distinct from each other for $0 < k < n$, and eigenvalues with distinct k are also distinct. The number of distinct pairs is then

$$\sum_{k=1}^{n-1} \binom{n}{k}^2 + 4 \sum_{k,k'=0}^{n} \binom{n}{k} \binom{n}{k'} = \Omega(4^n)$$

Taking $|a_k| = 2^{-n/2}/\sqrt{2}$ from Equation 1 and the fact that $|\lambda_i - \lambda_j| \leq 2$ since the λ_i are on the unit circle, we see that Equation 9 gives an upper bound on the ε-average mixing time of size $\Omega(2^n/\varepsilon)$. In general, this bound will give a mixing time of $\Omega(M/\varepsilon)$ whenever the initial state is distributed roughly equally over M eigenvectors, and when these are roughly equally distributed over $\omega(1)$ distinct eigenvalues.

For the continuous-time walk, on the other hand, Equation 8 shows that $\bar{P}_T$ approaches the average of $\cos^k 2t/n$. In fact, it is equal to this average whenever T is a multiple of $(\pi/2)n$. For k odd this average is zero, but for k even it is

$$\frac{1}{\pi} \int_0^\pi dx \cos^k x = \frac{2^k \pi}{\Gamma\left(\frac{1}{2} - \frac{k}{2}\right)^2 k!}$$

Since these Fourier coefficients do not vanish, $\bar{P}_T$ does not approach the uniform distribution even in the limit $T \to \infty$. In particular, the Fourier coefficient of $\bar{P}_T$ for $k = 2$ is

$$\tilde{\bar{P}}_T(2) = \frac{1}{T} \int_0^T dt \cos^2 \frac{2t}{n} = \frac{1}{2} + \frac{\sin 4T/n}{8T/n} \tag{10}$$

This integral is minimized when $T = 1.12335n$, at which point $\tilde{\bar{P}}_T(2) = 0.39138+$. Since $\tilde{\bar{P}}_T(2)$ is bounded below by this, it is easy to show that the total variation distance $\|\bar{P}_T - U\|$ is bounded away from zero as a result. Thus there exists $\varepsilon > 0$ such that no ε-average mixing time exists.

Acknowledgments. We are grateful to Dorit Aharonov, Mark Newman, Tony O'Connor, Leonard Schulman, and Umesh Vazirani for helpful conversations, and to McGill University and the Bellairs Research Institute for hosting a conference at which a significant part of this work was done. This work is partially supported by NSF grant PHY-0071139.

References

[AAKV01] Dorit Aharonov, Andris Ambainis, Julia Kempe, and Umesh Vazirani. Quantum walks on graphs. In ACM [ACM01].

[ABN+01] Andris Ambainis, Eric Bach, Ashwin Nayak, Ashvin Vishwanath, and John Watrous. One-dimensional quantum walks. In ACM [ACM01].

[ACM01] *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, Crete, Greece, 6–8 July 2001.

[AS92] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., 1992.

[BH75] Norman Bleistein and Richard Handelsman. *Asymptotic expansions of integrals*. Holt, Rinehart and Winston, 1975.

[CFG01] Andrew Childs, Edward Farhi, and Sam Gutmann. An example of the difference between quantum and classical random walks. Los Alamos preprint archive, `quant-ph/0103020`, 2001.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38(1):1–17, January 1991.

[Dia88] Persi Diaconis. *Group Representations in Probability and Statistics*. Lecture notes–Monograph series. Institute of Mathematical Statistics, 1988.

[DS81] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 57:159–179, 1981.

[GG81] Ofer Gabber and Zvi Galil. Explicit constructions of linear-sized superconcentrators. *Journal of Computer and System Sciences*, 22(3):407–420, June 1981.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 212–219, Philadelphia, Pennsylvania, 22–24 May 1996.

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, December 1989.

[JSV00] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. Technical Report TR00-079, The Electronic Colloquium on Computational Complexity, 2000.

[LK99] László Lovász and Ravi Kannan. Faster mixing via average conductance. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, Atlanta, Georgia, 1–4 May 1999.

[Lub94] Alexander Lubotzky. *Discrete Groups, Expanding Graphs, and Invariant Measures*, volume 125 of *Progress in Mathematics*. Birkhäuser Verlag, 1994.

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[Nis90] Noam Nisan. Pseudorandom generators for space-bounded computation. In *Proceedings of the Twenty Second Annual ACM Symposium on Theory of Computing*, pages 204–212, Baltimore, Maryland, 14–16 May 1990.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, August 1993.

[NV00] Ashwin Nayak and Ashvin Vishawanath. Quantum walk on the line. Los Alamos preprint archive, `quant-ph/0010117`, 2000.

[Per92] Rene Peralta. On the distribution of quadratic residues and nonresidues modulo a prime number. *Mathematics of Computation*, 58(197):433–440, 1992.

[Sch99] Uwe Schöning. A probabilistic algorithm for k-SAT and constraint satisfaction problems. In *40th Annual Symposium on Foundations of Computer Science*, pages 17–19. IEEE, 1999.

[Sho97] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, October 1997.

[Vaz92] Umesh Vazirani. Rapidly mixing markov chains. In Béla Bollobás, editor, *Probabilistic Combinatorics and Its Applications*, volume 44 of *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, 1992.

[Wat01] John Watrous. Quantum simulations of classical random walks and undirected graph connectivity. *Journal of Computer and System Sciences*, 62(2):376–391, 2001.

A Grover's diffusion operator

In general, the selection of the local operator D on the direction space appears to introduce a certain amount of artificiality into the definition of a discrete-time quantum walk. If we ask, however, that the operator obey the permutation symmetry of the hypercube, then there is a one-parameter family of such unitary operators up to multiplication by an overall phase.

To see this, suppose D is unitary and permutation-symmetric. Then it can have only two distinct entries, namely those on the diagonal and off it. Let $D_{ij} = a$ if $i = j$ and b if $i \neq j$. Then unitarity requires that $|a|^2 + (n-1)|b|^2 = 1$ and $2\,\mathrm{Re}\,ab^*) + (n-2)|b|^2 = 0$. The first of these two equations describes a circle, and their difference gives another, $|a-b|^2 = 1$. The intersection of these circles gives at most two values for b which differ only by a phase (and by conjugation if a is real). Solutions exist when $1 - 2/n \leq |a| \leq 1$.

To show that Grover's operator is the member of this family farthest from the family of diagonal unitary matrices $\{c\mathbf{1} : |c| = 1\}$, recall that the *operator norm* of a matrix A is $\|A\| = \mathrm{Tr}A^\dagger A$. Then the distance from D to this family is

$$\|D - c\mathbf{1}\| = n|a-c|^2 + (n^2-n)|b|^2 = 2n(1 - \mathrm{Re}\,ac^*)$$

When c has the same phase as a this is minimized at $2n(1-a)$, and this minimum is maximized when $|a| = 1 - 2/n$. This corresponds to Grover's operator times an overall phase; in this paper we take a to be real and negative.

B Resonances in the discrete-time walk

In order to evaluate Equation 5, we use Fourier analysis again — this time on functions of j and l, or rather on the rescaled variables

$$x = \cos\omega_j = 1 - \frac{2j}{n}, \quad y = \cos\omega_l = 1 - \frac{2l}{n}$$

We Fourier transform the quantity Y in Equation 5. Since we are interested in oscillations of frequency $\Theta(t)$, we write

$$Y(x,y) = \sum_{p_x,\,p_y \in \mathbb{Z}} \tilde{Y}\left(\frac{\pi p_x}{t}, \frac{\pi p_y}{t}\right) \mathrm{e}^{-i\pi(p_x x + p_y y)} \tag{11}$$

so that as t goes to infinity, we may treat this as the integral

$$Y(x,y) = \iint \tilde{Y}(\beta_x,\beta_y)\,\mathrm{e}^{-it(\beta_x x+\beta_y y)}\,\mathrm{d}\beta_x\,\mathrm{d}\beta_y. \tag{12}$$

Then, using the binomial theorem, we have

$$\tilde{P}_t(k) = \iint \mathrm{d}\beta_x\,\mathrm{d}\beta_y\,\tilde{Y}(\beta_x,\beta_y)\,\mathrm{e}^{-it\left(\left(1-\frac{k}{n}\right)\beta_x+\frac{k}{n}\beta_y\right)}\,\cos^k\frac{\beta_x t}{n}\,\cos^{n-k}\frac{\beta_y t}{n} \tag{13}$$

We will show that $\tilde{Y}$ peaks at values of β_x and β_y corresponding to the first-order oscillation, namely $(\beta_x,\beta_y) = (2,0)$ and $(0,2)$. This gives a form similar to Equation 6, so that if $2t/n = \pi/2$ the total variation distance will be exponentially small.

We calculate $\tilde{Y}$ by inverting Equation 12,

$$\tilde{Y}(\beta_x,\beta_y) = \frac{1}{4}\int_{-1}^{+1}\int_{-1}^{+1} \mathrm{d}x\,\mathrm{d}y\,Y(x,y)\,\mathrm{e}^{it(\beta_x x+\beta_y y)}$$

where the normalization is due to the range of x and y. We divide this integral into two terms, both of which are of the form

$$\iint \mathrm{d}x\,\mathrm{d}y\left(\frac{1\mp A}{2}\right)\cos\omega_\pm t\;\mathrm{e}^{it(\beta_x x+\beta_y y)} = O\left(\iint \mathrm{d}x\,\mathrm{d}y\left(\frac{1\mp A}{2}\right)\mathrm{e}^{it(\omega_\pm+\beta_x x+\beta_y y)}\right) \tag{14}$$

We can evaluate the right-hand integral in Equation 14 using the method of stationary phase, also known as steepest descent, which Nayak and Vishwanath [NV00] use to find the asymptotic form of the wave function on the line. In general, if f is a slowly varying function then the asymptotic integral

$$\lim_{t\to\infty}\iint f(x,y)e^{it\phi(x,y)}\mathrm{d}x\mathrm{d}y$$

is dominated by contributions from the points (x,y) in the domain of integration where ϕ has zero gradient. (See, e.g., [BH75].) If r is the smallest integer such that the rth derivative of ϕ at (x,y) is nonzero, we say that (x,y) is *rth-order*. In general, such asymptotic integrals are dominated by contributions from the stationary points of highest order.

In Equation 14 the slowly varying function is $(1\mp A)/2$, and the phase function is

$$\phi_\pm(x,y) = \omega_\pm + \beta_x x + \beta_y y$$

Its derivatives are

$$\begin{aligned}\frac{\partial\phi_\pm}{\partial x} &= -\frac{1}{\sin\omega_{j+l}} \pm \frac{1}{\sin\omega_{k-j+l}} + \beta_x \\ \frac{\partial\phi_\pm}{\partial y} &= -\frac{1}{\sin\omega_{j+l}} \mp \frac{1}{\sin\omega_{k-j+l}} + \beta_y\end{aligned}$$

For both ϕ_+ and ϕ_-, setting these to zero gives four stationary points (x_0, y_0), where the angles ω_{j+l}, ω_{k-j+l} are described by

$$\sin\omega_{j+l} = \frac{2}{\beta_x+\beta_y} \qquad \sin\omega_{k-j+l} = \frac{2}{|\beta_x-\beta_y|} \tag{15}$$

$$\cos\omega_{j+l} = x_0+y_0-1 = \pm\sqrt{1-\left(\frac{2}{\beta_x+\beta_y}\right)^2} \qquad \cos\omega_{k-j+l} = 1-\frac{2k}{n}-x_0+y_0 = \pm\sqrt{1-\left(\frac{2}{\beta_x-\beta_y}\right)^2}$$

Note that the signs of the cosines can be chosen independently, and all four possibilities exist for both ϕ_+ and ϕ_-. Choosing both cosines to be positive gives

$$\begin{aligned}x_0 &= \frac{1}{2}\left(\sqrt{1-\left(\frac{2}{\beta_x+\beta_y}\right)^2} - \sqrt{1-\left(\frac{2}{\beta_x-\beta_y}\right)^2}\right) + 1 - \frac{k}{n} \\ y_0 &= \frac{1}{2}\left(\sqrt{1-\left(\frac{2}{\beta_x+\beta_y}\right)^2} + \sqrt{1-\left(\frac{2}{\beta_x-\beta_y}\right)^2}\right) + \frac{k}{n}\end{aligned} \tag{16}$$

The other three solutions are given by choosing one or both of the cosines in Equation 15 to be negative, which affects the signs of the square roots in Equation 16. For these solutions to be real, we require $\beta_y \geq |\beta_x|+2$ for the stationary points of ϕ_+, and $\beta_x \geq |\beta_y|+2$ for the stationary points of ϕ_-. Thus $\beta_y-\beta_x \geq 2$ for ϕ_+ and $\beta_x-\beta_y \geq 2$ for ϕ_-, and in both cases $\beta_x+\beta_y \geq 2$.

To find the order of these stationary points, we calculate ϕ's second derivatives at (x_0, y_0):

$$\begin{aligned}\frac{\partial^2\phi_\pm}{\partial x^2} &= \frac{\partial^2\phi_\pm}{\partial y^2} = -\frac{\cos\omega_{j+l}}{\sin^3\omega_{j+l}} \mp \frac{\cos\omega_{k-j+l}}{\sin^3\omega_{k-j+l}} \\ \frac{\partial^2\phi_\pm}{\partial x\partial y} &= \frac{\partial^2\phi_\pm}{\partial y\partial x} = -\frac{\cos\omega_{j+l}}{\sin^3\omega_{j+l}} \pm \frac{\cos\omega_{k-j+l}}{\sin^3\omega_{k-j+l}}\end{aligned} \tag{17}$$

Given the restrictions on β_x and β_y for the stationary points to be real, for each of ϕ_+ and ϕ_- the second derivatives are zero at exactly one pair of frequencies, namely $\beta_x = 0$ and $\beta_y = 2$ for ϕ_+, and $\beta_x = 2$ and $\beta_y = 0$ for ϕ_-. We will call these the *dominant stationary points*. Note that at these frequencies we have $\omega_{j+l} = \omega_{k-j+l} = \pi/2$ and the four stationary points coincide at the peak of the binomials in Equation 5 where $j = k/2$ and $l = (n-k)/2$. Moreover, these frequencies are exactly the first-order oscillations of Y appearing in Equation 6.

Computing the third order derivatives at $\omega_{j+l} = \omega_{k-j+l} = \pi/2$ gives

$$\begin{aligned}\frac{\partial^3\phi_\pm}{\partial x^3} &= \frac{\partial^3\phi_\pm}{\partial x\partial y^2} = -\left[\frac{1}{\sin^3\omega_{j+l}} + \frac{3\cos^2\omega_{j+l}}{\sin^5\omega_{j+l}}\right] \pm \left[\frac{1}{\sin^3\omega_{k-j+l}} + \frac{3\cos^2\omega_{k-j+l}}{\sin^5\omega_{k-j+l}}\right] = -1\pm 1 \\ \frac{\partial^3\phi_\pm}{\partial y^3} &= \frac{\partial^3\phi_\pm}{\partial x^2\partial y} = -\left[\frac{1}{\sin^3\omega_{j+l}} + \frac{3\cos^2\omega_{j+l}}{\sin^5\omega_{j+l}}\right] \mp \left[\frac{1}{\sin^3\omega_{k-j+l}} + \frac{3\cos^2\omega_{k-j+l}}{\sin^5\omega_{k-j+l}}\right] = -1\mp 1\end{aligned}$$

Thus the dominant stationary points are third order, and in their vicinity $\phi_\pm$ takes the form

$$\phi_\pm = \frac{1}{6}\left(-(x+y)^3 \pm (x-y)^3\right) + o(x^4, y^4)$$

Thus if we rotate $\pi/4$ to new variables $a = x+y$ and $b = x-y$, we transform ϕ into the sum of two decoupled functions in the vicinity of the dominant stationary point, and write the integral of Equation 14 as the product of two one-dimensional integrals. For one-dimensional integrals with a third-order stationary point x_0, this takes the form [BH75, §7]

$$\lim_{t\to\infty} \int \mathrm{d}x\, f(x)\, \mathrm{e}^{it\phi(x)} = \frac{\Gamma(1/3)}{t^{1/3}}\, f(x_0)\, \mathrm{e}^{it\phi(x_0)}\, \frac{\mathrm{e}^{i\pi\,\mathrm{sgn}(\sigma)/6}}{3\,|\sigma|^{1/3}} + o(t^{-1/3})$$

where $\sigma = \phi'''(x_0)$ is the third derivative at x_0. Since we have the product of two such integrals, and since $f(x_0) = (1 \mp A)/2 = o(1)$ and $|\sigma| = 2$, the contribution of the dominant stationary point to $\tilde{P}_t(k)$ is

$$\left[\tilde{P}_t(k)\right]_{\text{dominant}} = o\left(t^{-2/3}\left(\cos^k \frac{2t}{n} + \cos^{n-k} \frac{2t}{n}\right)\right) \tag{18}$$

We now need to calculate the contribution of the other stationary points. These are second-order, and their contribution takes the form

$$\lim_{t\to\infty} \iint \mathrm{d}x\,\mathrm{d}y\, f(x,y)\, \mathrm{e}^{it\phi(x,y)} = \frac{2\pi}{t} \sum_{(x,y)} f(x,y)\, \mathrm{e}^{it\phi(x,y)}\, \frac{\mathrm{e}^{i\pi\delta_{x,y}/2}}{\sqrt{|\det \partial^2 \phi_{x,y}|}} + o\left(\frac{1}{t^2}\right) \tag{19}$$

where $\partial^2\phi_{x,y}$ is the matrix of second derivatives of ϕ at (x,y), and $\delta_{x,y}$ is $+1$, 0, or -1 depending on whether zero, one, or both of its eigenvalues are negative. From Equation 17 we have

$$\det \partial^2 \phi_\pm = \pm 4\, \frac{\cos\omega_{j+l}\, \cos\omega_{k-j+l}}{\sin^3\omega_{j+l}\, \sin^3\omega_{k-j+l}}$$

Focusing on the oscillating part of Equation 13, we have

$$\iint \mathrm{d}\beta_x\, \mathrm{d}\beta_y\, \mathrm{e}^{it\psi_\pm(\beta_x,\beta_y)}\, \cos^k \frac{\beta_x t}{n}\, \cos^{n-k} \frac{\beta_y t}{n} \tag{20}$$

where

$$\psi_\pm(\beta_x, \beta_y) = \phi_\pm(x_0, y_0) - \left(1 - \frac{k}{n}\right)\beta_x - \frac{k}{n}\beta_y$$

Since this really is an integral in the limit $n \to \infty$, the $\cos^k$, $\cos^{n-k}$ terms create sharper and sharper peaks where β_x, β_y are multiples of 4. We can approximate ψ at each peak to first order as a function of β_x and β_y. For the stationary point of $\phi_\pm$ where the sign of both cosines is positive, $\psi_\pm$ is given by

$$\psi_\pm(\beta_x,\beta_y) = \sin^{-1}\frac{2}{\beta_x+\beta_y} - \sin^{-1}\frac{2}{\beta_x-\beta_y} + \sqrt{\left(\frac{\beta_x+\beta_y}{2}\right)^2 - 1} \pm \sqrt{\left(\frac{\beta_x-\beta_y}{2}\right)^2 - 1}$$

Its derivatives with respect to β_x and β_y are

$$\begin{aligned}
\frac{\partial\psi_\pm}{\partial\beta_x} &= \frac{1}{2}\left(\sqrt{1-\left(\frac{2}{\beta_x+\beta_y}\right)^2} - \sqrt{1-\left(\frac{2}{\beta_x-\beta_y}\right)^2}\right) = x_0 - \left(1-\frac{k}{n}\right) = \frac{k-2j_0}{n} \\
\frac{\partial\psi_\pm}{\partial\beta_y} &= \frac{1}{2}\left(\sqrt{1-\left(\frac{2}{\beta_x+\beta_y}\right)^2} + \sqrt{1-\left(\frac{2}{\beta_x-\beta_y}\right)^2}\right) = y_0 - \frac{k}{n} = \frac{n-k-2l_0}{n}
\end{aligned} \tag{21}$$

and similarly for the other stationary points (x_0, y_0); we can also derive this directly from the definition of $\psi_\pm$ and the fact that we are at a stationary point of $\phi_\pm$. In other words, the derivatives of ψ are proportional to the distance of the stationary points off the binomial peaks.

The entire (β_x, β_y)-plane can be tiled with 4×4 squares centered on these peaks. Integrating Equation 20 on one such tile, say around the peak $\beta_x = 4p$, $\beta_y = 4q$, gives

$$\begin{aligned} & \int_{4p-2}^{4p+2} \int_{4q-2}^{4q+2} d\beta_x \, d\beta_y \, e^{it\left(\frac{\partial\psi}{\partial\beta_x}\beta_x + \frac{\partial\psi}{\partial\beta_y}\beta_y\right)} \cos^k \frac{\beta_x t}{n} \cos^{n-k} \frac{\beta_y t}{n} \\ = \quad & \frac{\pi^2 n^2}{2^n t^2} \binom{k}{\frac{1}{2}(k - n\frac{\partial\psi}{\partial\beta_x})} \binom{n-k}{\frac{1}{2}(n-k-n\frac{\partial\psi}{\partial\beta_y})} = \frac{\pi^2 n^2}{2^n t^2} \binom{k}{j_0} \binom{n-k}{l_0} \\ =_O \quad & \exp\left[n \left(\frac{k}{n} h\left(\frac{j_0}{k}\right) + \left(1 - \frac{k}{n}\right) h\left(\frac{l_0}{n-k}\right) - \ln 2 \right) \right] = \exp(nZ) \end{aligned} \tag{22}$$

where $h(z) = -z \ln z - (1-z)\ln(1-z)$ is the entropy function. Note that if the quantity Z in Equation 22 is less than $-\ln\sqrt{2}$ for all stationary points other than the dominant ones, then their contribution to $|\tilde{P}(k)|^2$ will be $2^{-\gamma n}$ where $\gamma > 1$, in which case summing over all k will give an exponentially small contribution, $O(2^{(1-\gamma)n})$, to the total variation distance. To confirm this, note that Z is maximized by the other stationary points closest to the origin, such as the stationary point of ϕ_+, with both cosines positive, where $\beta_x = 0$ and $\beta_y = 4$. From Equation 21 this gives $\partial\psi/\partial\beta_x = 0$ and $\partial\psi/\partial\beta_y = \sqrt{3}/2$, and so $j_0 = k/2$ and $l_0 = ((1 - \frac{\sqrt{3}}{2})n - k)/2$. Both binomials are non-zero only in the interval $k \in \left(0, (1 - \frac{\sqrt{3}}{2})n\right)$ and Z is maximized at $k = 0$, where

$$Z = h\left(\frac{1}{2} - \frac{\sqrt{3}}{4}\right) - \ln 2 = -0.447 < \ln\frac{1}{\sqrt{2}} = -0.346$$

The other second-order stationary points are this far or farther from the origin, giving values of j_0 and l_0 farther off the binomial peaks, and therefore smaller entropies.

Recalling Equation 19 above, our final concern is the sum of the heights of these peaks,

$$\sum_{\beta_x, \beta_y} \frac{1}{\sqrt{\left|\det \partial^2 \phi_{\beta_x, \beta_y}\right|}}$$

taken over all second-order stationary points (β_x, β_y). Since these occur when β_x, β_y are multiples of 4, from Equation 15 we have $\left|\cos\omega_{j+l} \cos\omega_{k-j+l}\right| \geq 3/4$. Then

$$\left|\det \partial^2 \phi_\pm(\beta_x, \beta_y)\right| \geq \frac{3}{\left|\sin^3 \omega_{j+l} \sin^3 \omega_{k-j+l}\right|} = \frac{3}{64} \left|\beta_x + \beta_y\right|^3 \left|\beta_x - \beta_y\right|^3$$

and it is sufficient to show that the sum

$$\sum_{\beta_x \neq \beta_y} |\beta_x + \beta_y|^{-3/2} |\beta_x - \beta_y|^{-3/2}$$

converges. Again rotating by $\pi/4$ to variables $a = \beta_x + \beta_y$ and $b = \beta_x - \beta_y$, we get the sum

$$\sum_{a,b} |a|^{-3/2} |b|^{-3/2} \leq \left(\sum_a |a|^{-3/2} \right)^2$$

Observing that $\sum_{a>0} a^{-3/2}$ converges shows that the contribution of the second-order stationary points is exponentially small.

Now we return to the dominant contribution to $\tilde{P}_t(k)$, Equation 18. If we could have $t = (\pi/4)n$ exactly, this dominant term would be zero, leaving us with the second-order stationary points and an exponentially small total variation distance. However, in the discrete-time walk t must be an integer. Setting $t = \lceil (\pi/4)n \rceil$, we have $\cos 2t/n = o(1/n)$. Using the binomial theorem and Equation 18, the Diaconis-Shahshahani bound gives

$$\begin{aligned} \|P_t - U\|^2 &=_o n^{-4/3} \sum_{0<k<n} \binom{n}{k} \left(2\cos^{2k}\frac{2t}{n} + 2\cos^n \frac{2t}{n} \right) \\ &\leq 2n^{-4/3} \left[\left(2\cos\frac{2t}{n} \right)^n + \left(1 + \cos^2 \frac{2t}{n} \right)^n - 1 \right] = o(n^{-7/3}) \end{aligned}$$

and so the total variation distance is $\|P_t - U\| = o(n^{-7/6})$, completing the proof of Theorem 1.

C A graph product derivation of the continuous-time walk

As an alternate derivation for the continuous-time walk, we can calculate the wave function ψ_t directly by exploiting the hypercube's simple structure as a product graph. Let σ_x be the Pauli matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right)$. Then we can rewrite Equation 7 as

$$H = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1} \otimes \cdots \otimes \sigma_x \otimes \cdots \otimes \mathbf{1}$$

where the jth term in the sum has σ_x appearing in the jth place in the tensor product. Then using the identity $(A \otimes B)(C \otimes D) = AB \otimes CD$, and the fact that $e^{A+B} = e^A e^B$ when A and B commute, we have

$$U = e^{iHt} = \prod_{j=1}^{n} \mathbf{1} \otimes \cdots \otimes e^{it\sigma_x/n} \otimes \cdots \otimes \mathbf{1} = \left[e^{it\sigma_x/n} \right]^{\otimes n} = \begin{pmatrix} \cos t/n & i \sin t/n \\ i \sin t/n & \cos t/n \end{pmatrix}^{\otimes n}$$

where $A^{\otimes n}$ is the tensor product of n copies of A. If $\psi_0 = |0 \cdots 0\rangle = |0\rangle^{\otimes n}$, then

$$\psi_t = U_t \psi_0 = \left[\left(\cos\frac{t}{n} \right) |0\rangle + \left(i \sin\frac{t}{n} \right) |1\rangle \right]^{\otimes n}$$

and we see that the continuous-time walk is equivalent to n non-interacting one-qubit systems. Then the amplitude for observing the particle at a position $\vec{x}$ with Hamming weight x is

$$\psi_t(\vec{x}) = \left(\cos\frac{t}{n} \right)^{n-x} \left(i \sin\frac{t}{n} \right)^x$$

which when $t = k(\pi/4)n$ for k odd gives $|\psi_t(x)|^2 = 2^{-n}$, the uniform distribution.