J. Appl. Probab. **60,** 253–266 (2023)
*doi:*10.1017/jpr.2022.37

FAST MIXING OF A RANDOMIZED SHIFT-REGISTER MARKOV CHAIN

DAVID A. LEVIN,* ** AND
CHANDAN TANKALA,* *** *University of Oregon*

Abstract

We present a Markov chain on the n-dimensional hypercube $\{0, 1\}^n$ which satisfies $t_{\text{mix}}^{(n)}(\varepsilon) = n[1 + o(1)]$. This Markov chain alternates between random and deterministic moves, and we prove that the chain has a cutoff with a window of size at most $O(n^{0.5+\delta})$, where $\delta > 0$. The deterministic moves correspond to a linear shift register.

Keywords: Markov chains; fast mixing; cutoff; hypercube

2020 Mathematics Subject Classification: Primary 60J10
Secondary 94A55

1. Introduction

Developing Markov chain Monte Carlo (MCMC) algorithms with fast mixing times remains a problem of practical importance. We would like to make computationally tractable modifications to existing chains which decrease the time required to obtain near equilibrium samples.

The *mixing time* of an ergodic finite Markov chain (X_t) with stationary distribution π is defined as $t_{\text{mix}}(\varepsilon) = \min\{t \geq 0 : \max_x \|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\text{TV}} < \varepsilon\}$, and we write $t_{\text{mix}} = t_{\text{mix}}(1/4)$.

A theoretical algorithm for chains with uniform stationary distribution was analyzed by Chatterjee and Diaconis [7]. They proposed chains that alternate between random steps made according to a probability transition matrix and deterministic steps defined by a bijection f on the state space. Supposing the state space has size n, the transition matrix satisfies a one-step reversibility condition, and f obeys an *expansion condition*, they proved that $t_{\text{mix}} = O(\log n)$. However, they noted that finding an explicit bijection f satisfying the expansion condition can be difficult even for simple state spaces like $\mathbb{Z}_n$.

In this paper we analyze a Markov chain on the hypercube $\{0, 1\}^n$ of the form $P\Pi$ for an explicit Π, where P corresponds to the usual lazy random walk on $\{0, 1\}^n$. This chain may be of independent interest, as the deterministic transformation f on the state space is a ‘shift register’ operator. Such shift registers have many applications in cryptography, pseudo-random number generation, coding, and other fields. See, for example, [13] for background on shift registers.

The *lazy random walk* on $\{0, 1\}^n$ makes transitions as follows. When the current state is x, a coordinate from $i \in \{1, 2, \ldots, n\}$ is generated uniformly at random, and an independent random bit R is added (mod 2) to the bit x_i at coordinate i. The new state obtained is thus

$$x \mapsto x' = (x_1, \ldots, x_i \oplus R, \ldots, x_n). \tag{1}$$

Received 9 September 2021; revision received 3 March 2022.

* Postal address: Department of Mathematics, Fenton Hall, University of Oregon, Eugene, OR, 97403-1222 USA.
** Email address: dlevin@uoregon.edu
*** Email address: chandant@uoregon.edu

We denote the transition matrix of this chain by P. For a chain with transition probabilities Q on S and stationary distribution π, let $d(t) = d_n(t) = \max_{x\in S} \|Q^t(x, \cdot) - \pi\|_{\mathrm{TV}}$. A sequence of chains indexed by n has a *cutoff* if, for $t_n := t_{\mathrm{mix}}^{(n)}$, there exists a *window sequence* $\{w_n\}$ with $w_n = o(t_n)$ such that $\lim_{\alpha\to\infty} \limsup_{n\to\infty} d_n(t_n + \alpha w_n) = 0$, $\lim_{\alpha\to-\infty} \liminf_{n\to\infty} d_n(t_n + \alpha w_n) = 0$. For background on mixing times, cutoff, and related material, see, for example, [16].

It is well known that, for the lazy random walk on $\{0, 1\}^n$, $t_{\mathrm{mix}}(\varepsilon) = \frac{1}{2} n \log n[1 + o(1)]$ with a cutoff. (See [10], where precise information on the total variation distance is calculated. The difference of a factor of 2 above comes from the laziness in our version.)

A natural deterministic 'mixing' transformation on $\{0, 1\}^n$ is the 'linear shift register', which takes the xor sum of the bits in the current word $x = (x_1, \ldots, x_n)$ and appends it to the right-hand side, dropping the left-most bit:

$$x \mapsto f(x) = \big(x_2, \ldots, x_{n-1}, \oplus_{i=1}^n x_i\big). \tag{2}$$

Let Π denote the permutation matrix corresponding to this transformation, so that

$$\Pi_{i,j} = \begin{cases} 1 & \text{if } j = f(i), \\ 0 & \text{otherwise.} \end{cases}$$

The chain studied in the following has the transition matrix $Q_1 = P\Pi$, whose dynamics are simply described by combining the stochastic operation (1) with the deterministic f in (2): $x \mapsto x' \mapsto f(x')$.

Let log stand for the natural logarithm. The main result here is the following.

Theorem 1. *For the chain Q_1,*

(i) For $n \geq 5$, $d_n(n + 1) \leq 2/n$.

(ii) For any $\frac{1}{2} < \alpha < 1$, if $t_n = n - n^\alpha$, $d_n(t_n) \geq \|Q_1^{t_n}(0, \cdot) - \pi\|_{\mathrm{TV}} \geq 1 - o(1)$.

Thus, the sequence of chains has a cutoff at time n with a window of at most size $n^{1/2+\delta}$ for any $\delta > 0$.

Remark 1. If the transformation f obeys the expansion condition of [7], then the results therein yield a mixing time of order n. We were unable to directly verify that f does obey this condition. Moreover, the result in Theorem 1 establishes the stronger cutoff property.

Remark 2. Obviously a simple way to exactly randomize n bits in exactly n steps is to simply randomize in sequence, say from left to right, each bit. This is called a *systematic scan*, and avoids the extra factor of $\log n$ needed for *random updates* to touch a sufficient number of bits. (A 'coupon-collector' argument shows that to touch all but $O(\sqrt{n})$ bits using random updates, enough to achieve small total-variation distance from uniform, order $n \log n$ steps are required.) Thus, clearly our interest in analyzing this chain is not for direct simulation of n independent bits! Rather, we are motivated both by the potential for explicit deterministic moves to speed up Markov chains, and also by this particular chain which randomizes the well-known shift-register dynamical system.

This paper is organised as follows. In Section 2 we review some related results. The upper bound in Theorem 1 is proved in Section 3, and the lower bound is established in Section 4. In Section 5, a chain is analyzed that is similar to the chain of Theorem 1, but always updates the same location.

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

2. Related previous work

2.1. Markov chains on a hypercube

Previous work on combining deterministic transformation with random moves on a hypercube is described in [9], which studied the walk $\{X_t\}$ described by $X_{t+1} = AX_t + \epsilon_{t+1}$, where A is an $n \times n$ lower triangular matrix and ϵ_t are independent and identically distributed (i.i.d.) vectors having the following distribution: the variable $\epsilon_t = \mathbf{0}$ with probability $\theta \neq \frac{1}{2}$, while $\epsilon_t = e_1$ with probability $1 - \theta$. Here, $\mathbf{0}$ is a vector of zeros, and $\mathbf{e_1}$ is a vector with a one in the first coordinate and zeros elsewhere. Fourier analysis is used to show that $O(n \log n)$ steps are necessary and sufficient for mixing, and they prove a sharp result in both directions. This line of work is a specific case of a random walk on a finite group G described as $X_{t+1} = A(X_t)\epsilon_{t+1}$, where A is an automorphism of G and $\epsilon_1, \epsilon_2, \ldots$ are i.i.d. with some distribution μ on G. In the case of [9], $G = \mathbb{Z}_2^n$ and the automorphism A is a matrix. By comparison, the chain studied here mixes in only $n(1 + o(1))$ steps.

Another relevant (random) chain on $\{0, 1\}^n$ was analyzed in [17]. A subset of S size p from $\mathbb{Z}_2^n$ is chosen uniformly at random, and the graph G with vertex set $\mathbb{Z}_2^n$ is formed which contains an edge between vertices if and only if their difference is in S; [17] considered the random walk on the random graph G. It was shown that if $p = cn$, where $c > 1$ is a constant, then the mixing time is linear in n with high probability (over the choice of S) as $n \to \infty$. This Markov chain depends on the random environment to produce the speedup.

Finally, another example of cutoff for a Markov chain on a hypercube was [2]. This random walk moves by picking an ordered pair (i, j) of distinct coordinates uniformly at random and adding the bit at location i to the bit at location j, modulo 2; it was proved that this Markov chain has cutoff at time $\frac{3}{2}n \log n$ with a window of size n, so the mixing time is the same order as that of the ordinary random walk.

2.2. Related approaches to speeding up mixing

The results of [7] were refined in [3], proving further that, under mild assumptions on P, a 'typical' f yields a mixing time of order $\log n$ with a cutoff. In particular, if a permutation matrix Π is selected uniformly at random, then the (random) chain $Q = P\Pi$ has a cutoff at $(\log n)/\mathbf{h}$ with high probability (with respect to the selection of Π). Here, $\mathbf{h}$ is the entropy rate of P, and n is the size of the state space. Like the chain in [17], the random environment is critical to the analysis. However, in specific applications we would like to know an explicit deterministic permutation Π that mixes in $O(\log n)$ and does not require storage of the matrix Π, particularly when the state space increases exponentially with n.

A method for speeding up mixing called *lifting* was introduced in [11]. The idea behind this technique is to create 'long cycles' and introduce non-reversibility. For example, for a simple random walk on the n-path the mixing time of the lifting is $O(n)$, whereas the mixing time on the path is $\Theta(n^2)$. Thus, this method can provide a speedup of the square root of the mixing time of the original chain. An explicit lower bound on the mixing time of the lifted chain in terms of the original chain was given in [8]. The chain we study here has a similar flavor in that the transformation f creates non-reversibility and long cycles.

Another related speedup technique is *hit and run*, which introduces non-local moves in a chosen direction (see the survey in [1]). A recent application of a top-to-random shuffle is [5], where it is shown that a speedup in mixing by a constant factor can be obtained for the L^2 and sup-norm; [15] used this method to sample from high-dimensional and multi-modal posterior distributions in Bayesian models, and compared that with Gibbs and Hamiltonian

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

Monte Carlo algorithms. In the physics literature, non-reversible chains are constructed from reversible chains without augmenting the state space (in contrast to lifting) by introducing *vorticity*, which is similar in spirit to the long cycles generated by lifting; see, for example, [4], which analyzes a non-reversible version of Metropolis–Hastings.

As mentioned above, there are other obvious methods to obtain a fast uniform sample from $\{0, 1\}^n$, in particular systematic scan, which generates an exact sample in precisely n steps! See [12] for a comparison of systematic and random scans on different finite groups.

3. Upper bound of Theorem 1

The proof is based on Fourier analysis on $\mathbb{Z}_2^n$. Let A be a matrix defined as

$$A := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}_{n\times n} . \tag{3}$$

Let $\{\epsilon_i\}$ be i.i.d. random vectors with the distribution

$$\epsilon_i = \begin{cases} \mathbf{0} & \text{with probability } 1/2, \\ e_1 & \text{with probability } 1/2n, \\ \vdots & \\ e_n & \text{with probability } 1/2n, \end{cases} \tag{4}$$

$$\text{where} \quad e_i = (0, \ldots, \underbrace{1}_{i\text{th place}}, \ldots, 0).$$

The random walk X_t with transition matrix Q_1 and $X_0 = x$ can be described as $X_t = A(X_{t-1} \oplus \epsilon_t)$. (The matrix arithmetic here is all modulo 2.) Induction shows that

$$X_t = \left(\sum_{j=1}^{t} A^{t-j+1}\epsilon_j\right) \oplus A^t x, \tag{5}$$

where again the matrix multiplication and vector sums are over the field $\mathbb{Z}_2$.

Lemma 1. *The matrix A in (3) satisfies $A^{n+1} = I_{n\times n}$.*

Proof. Note that

$$\begin{aligned} Ae_1 &= e_n, \\ A^2e_1 &= A(Ae_1) = e_n + e_{n-1}, \\ A^3e_1 &= A(A^2e_1) = e_{n-1} + e_{n-2}, \\ &\vdots \\ A^ne_1 &= e_2 + e_1. \end{aligned}$$

This implies that $A^{n+1}e_1 = A(A^ne_1) = A(e_2 + e_1) = e_1$. The reader can check similarly that $A^{n+1}e_j = e_j$ for $2 \leq j \leq n$. □

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

For $x, y \in \mathbb{Z}_2^n$, the Fourier transform of $Q_1^t(x, \cdot)$ at y is defined as

$$\begin{aligned}\widehat{Q_1^t}(x, y) := \sum_{z\in\mathbb{Z}_2^n} (-1)^{y\cdot z} Q_1^t(x, z) = \mathbb{E}\big[(-1)^{y\cdot X_t}\big] \\ = (-1)^{y\cdot A^t x} \prod_{j=1}^{t} \mathbb{E}\big[(-1)^{y\cdot A^{t-j+1}\epsilon_j}\big]. \qquad (6)\end{aligned}$$

The product $x \cdot y$ is the inner product $\sum_{i=1}^n x_i y_i$. The second equality follows from plugging (5) into the first equality and observing that the ϵ_j are independent. The following lemma bounds the total variation distance; this is proved in [9, Lemma 2.3].

Lemma 2. $\|Q_1^t(x, \cdot) - \pi\|_{\mathrm{TV}}^2 \le \frac{1}{4} \sum_{y\neq 0} \big(\widehat{Q_1^t}(x, y)\big)^2$.

We will need the following lemma to prove Theorem 1(i).

Lemma 3. *Let*

$$h(n, k) = \binom{n}{k}\left(1 - \frac{k}{n}\right)^{2n-2k}\left(\frac{k}{n}\right)^{2k}.$$

If $2 \le k \le n-2$ *and* $n > 5$*, then* $h(n, k) \le 1/n^2$ *and* $h(n, n-1) \le 1/n$.

Proof. We first prove this for $2 \le k \le n-2$. For $x, y \in \mathbb{Z}^+$, define

$$K(x, y) := \log \frac{\Gamma(x+1)}{\Gamma(y+1)\Gamma(x-y+1)} + (2x-2y)\log\left(1 - \frac{y}{x}\right) + 2y\log\left(\frac{y}{x}\right) + 2\log(x),$$

where Γ is the gamma function. We prove that, if $2 \le y \le x-2$ and $x > 5$, then $K(x, y) < 0$. Since $K(n, k) = \log(h(n, k)n^2)$, this establishes the lemma.

Let

$$\psi(x) := \frac{\mathrm{d}\log\Gamma(x)}{\mathrm{d}x}.$$

Then

$$\begin{aligned}\frac{\partial^2 K}{\partial y^2} &= -\psi'(y+1) - \psi'(x-y+1) + \frac{2}{x-y} + \frac{2}{y} \\ &> -\frac{1}{y+1} - \frac{1}{(y+1)^2} - \frac{1}{(x-y+1)} - \frac{1}{(x-y+1)^2} + \frac{2}{y} + \frac{2}{x-y} \\ &> 0. \qquad (7)\end{aligned}$$

The first inequality follows from [14, Lemma 1], which states that $\psi'(x) < 1/x + 1/x^2$ for all $x > 0$.

The second inequality follows since $2/y - (y+1)^{-1} - (y+1)^{-2} > 0$, and we can apply this again, substituting $x - y$ for y. Thus, $K(x, \cdot)$ is a convex function for all x. Also,

$$\begin{aligned}K(x, 2) = K(x, x-2) &= \log\left(\frac{x(x-1)}{2}\right) + 2(x-2)\log\left(1 - \frac{2}{x}\right) + 4\log\left(\frac{2}{x}\right) + 2\log x \\ &= \log\left(\frac{8x(x-1)}{x^2}\right) + 2(x-2)\log\left(1 - \frac{2}{x}\right) \\ &< \log(8) - \frac{4(x-2)}{x} \\ &< 0, \qquad (8)\end{aligned}$$

for $x > 5$. The first inequality follows from $\log(1-u) < -u$. Equations (8) and (7) prove this lemma for $2 \le k \le n-2$. Finally, $h(n, n-1) \le 1/n$ if and only if $nh(n, n-1) \le 1$, which is true because one can verify that $\frac{\mathrm{d}}{\mathrm{d}n} nh(n, n-1) < 0$ for $n \ge 5$ and that $nh(n, n-1) < 1$ for $n = 5$. □

Proof of Theorem 1(i). Let $y = (y_1, y_2, \dots, y_n) \in \mathbb{Z}_n^2$. First,

$$\widehat{Q_1^{n+1}}(x, y) = (-1)^{y\cdot A^{n+1}x} \prod_{j=1}^{n+1} \mathbb{E}\big[(-1)^{y\cdot A^{n-j+2}\epsilon_j}\big] = (-1)^{y\cdot Ix} \prod_{j=1}^{n+1} \mathbb{E}\big[(-1)^{y\cdot A^{n-j+2}\epsilon_j}\big],$$

which follows from (6) and Lemma 1. Note that the first factor in this product is $\big(\frac{1}{2} + (1/2n)\big[(-1)^{y_1} + (-1)^{y_2} + (-1)^{y_3} + \cdots + (-1)^{y_n}\big]\big)$, which follows from (4) and Lemma 1. We can similarly find the other factors in the product, which gives

$$\begin{aligned}
\widehat{Q_1^{n+1}}(x, y) = (-1)^{x\cdot y} \\
&\times \left(\frac{1}{2} + \frac{1}{2n}\big[(-1)^{y_1} + (-1)^{y_2} + (-1)^{y_3} + \cdots + (-1)^{y_n}\big]\right) \\
&\times \left(\frac{1}{2} + \frac{1}{2n}\big[(-1)^{y_1} + (-1)^{y_1+y_2} + (-1)^{y_1+y_3} + \cdots + (-1)^{y_1+y_n}\big]\right) \\
&\times \left(\frac{1}{2} + \frac{1}{2n}\big[(-1)^{y_2} + (-1)^{y_2+y_1} + (-1)^{y_2+y_3} + \cdots + (-1)^{y_2+y_n}\big]\right) \\
&\vdots \\
&\times \left(\frac{1}{2} + \frac{1}{2n}\big[(-1)^{y_n} + (-1)^{y_n+y_1} + (-1)^{y_n+y_2} + \cdots + (-1)^{y_n+y_{n-1}}\big]\right).
\end{aligned}$$

Observe that $\widehat{Q_1^{n+1}}(x, y) = 0$ for all $y \in \mathbb{Z}_2^n$ such that $W(y) \in \{1, n\}$, where $W(y)$ is the Hamming weight of $y \in \mathbb{Z}_2^n$. If $W(y) = 1$, then one of the factors displayed on the third through sixth lines above is zero. If $W(y) = n$, then the factor on the second line is zero. If we fix a $2 \le j \le n-1$ and look at all $y \in \mathbb{Z}_2^n$ with $W(y) = j$, then $\big[\widehat{Q_1^{n+1}}(x, y)\big]^2$ is the same for all such y, since the expression above is invariant over permutation of coordinates, once the first factor is squared.

If $y = (\underbrace{1, 1, \dots, 1}_{k \text{ ones}}, 0, \dots, 0)$ where $2 \le k \le n-1$, then

$$\widehat{Q_1^{n+1}}(x, y) = (-1)^{\left(\sum_{i=1}^{k} x_i\right)} \left(1 - \frac{k}{n}\right)^{n-k+1} \left(\frac{k-1}{n}\right)^{k}.$$

This holds because factors 3 through $(k+2)$ are equal to $(k-1)/n$, and all factors except the first and second are equal to $(n-k)/n$. To see this, note that factor 2 is equal to

$$\frac{1}{2} + \frac{1}{2n}\Bigg[\underbrace{(-1-1-\cdots-1)}_{k \text{ negative ones}} + \underbrace{(1+1+\cdots+1)}_{n-k \text{ positive ones}}\Bigg] = \frac{1}{2} + \frac{1}{2n}[-k + (n-k)] = \frac{n-k}{n};$$

factors 3 through $(k+2)$ are equal to

$$\frac{1}{2}+\frac{1}{2n}\Bigg[-1+(\underbrace{1+1+\cdots+1}_{k-1\text{ positive ones}})+(\underbrace{-1-1-\cdots-1}_{n-k\text{ negative ones}})\Bigg]$$
$$=\frac{1}{2}+\frac{1}{2n}[-1+k-1-(n-k)]=\frac{k-1}{n};$$

and factors $(k+3)$ through $(n+2)$ are equal to

$$\frac{1}{2}+\frac{1}{2n}\Bigg[1+(\underbrace{-1-1-\cdots-1}_{k\text{ negative ones}})+(\underbrace{1+1+\cdots+1}_{n-k-1\text{ positive ones}})\Bigg]$$
$$=\frac{1}{2}+\frac{1}{2n}[1-k+n-k-1]=\frac{n-k}{n}.$$

Thus,

$$\begin{aligned}\sum_{y\neq 0}\left(\widehat{Q_1^{n+1}}(x,y)\right)^2&=\sum_{k=2}^{n-1}\binom{n}{k}\left(1-\frac{k}{n}\right)^{2n-2k+2}\left(\frac{k-1}{n}\right)^{2k}\\&\leq\sum_{k=2}^{n-1}\binom{n}{k}\left(1-\frac{k}{n}\right)^{2n-2k}\left(\frac{k}{n}\right)^{2k}.\end{aligned}\tag{9}$$

We finally analyze the terms in the sum of (9). Note that

$$\binom{n}{k}\left(1-\frac{k}{n}\right)^{2n-2k}\left(\frac{k}{n}\right)^{2k}\leq\frac{1}{n^2}$$

for $2\leq k\leq n-2$ and $n>5$, and $h(n,n-1)\leq 1/n$ by Lemma 3. Thus,

$$\sum_{k=2}^{n-1}\binom{n}{k}\left(1-\frac{k}{n}\right)^{2n-2k}\left(\frac{k}{n}\right)^{2k}\leq\frac{n-3}{n^2}+\frac{1}{n}\leq\frac{2}{n}.\tag{10}$$

Lemma 2, with (9) and the bound in (10), establishes the upper bound in Theorem 1. □

4. Lower bound of Theorem 1

Let $\{U_t\}$ be the sequence of coordinates used to update the chain, and let $\{R_t\}$ be the sequence of random bits used to update. Thus, at time t, coordinate U_t is updated using bit R_t. Both sequences are i.i.d. Let $\mathcal{F}_t$ be the σ-algebra generated by $(U_1,\ldots,U_t)$ and $(R_1,\ldots,R_t)$. Let $X_t=\left(X_t^{(1)},\ldots,X_t^{(n)}\right)$ be the chain with transition matrix Q at time t. The proof is based on the distinguishing statistic $W_t=\sum_{i=1}^n X_t^{(i)}$, the Hamming weight at time t.

First, observe that $\mathbb{P}\left(X_{t+1}^{(n)}=1\mid\mathcal{F}_t\right)=\frac{1}{2}$, because if the state at time t is $x=(x_1,x_2,\ldots,x_n)$, then R_{t+1} is added at a uniformly chosen coordinate of x, and $X_{t+1}^{(n)}=\sum_{i=1}^n x_i\oplus R_{t+1}\in\{0,1\}$ with probability $\frac{1}{2}$ each, conditioned on $\mathcal{F}_t$. We now describe a recursive relation for W_t, the Hamming weight of X_t:

$$W_{t+1}=\sum_{j=2}^{n}\left(X_t^{(j)}\cdot\mathbf{1}_{(U_{t+1}\neq j)}+\left(X_t^{(j)}\oplus R_{t+1}\right)\cdot\mathbf{1}_{(U_{t+1}=j)}\right)+\mathbf{1}_{\left(X_{t+1}^{(n)}=1\right)}.\tag{11}$$

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

The first terms in (11) follow from the fact that, for $1 \le j \le n-1$,

$$X_{t+1}^{(j)} = \begin{cases} X_t^{(j+1)} & \text{if } U_{t+1} \neq j, \\ X_t^{(j+1)} \oplus R_{t+1} & \text{if } U_{t+1} = j. \end{cases}$$

Taking conditional expectation in (11), we get

$$\begin{aligned} \mathbb{E}[W_{t+1} \mid \mathcal{F}_t] &= \sum_{j=2}^{n} \left(X_t^{(j)} \left(\frac{n-1}{n} \right) + \frac{1}{2}[1 - X_t^{(j)} + X_t^{(j)}]\frac{1}{n} \right) + \frac{1}{2} \\ &= \left(1 - \frac{1}{n}\right) \sum_{j=2}^{n} X_t^{(j)} + \left(\frac{n-1}{2n} \right) + \frac{1}{2} \\ &= \left(1 - \frac{1}{n}\right) \sum_{j=1}^{n} X_t^{(j)} - \left(1 - \frac{1}{n}\right) X_t^{(1)} + \left(\frac{2n-1}{2n} \right) \\ &= \left(1 - \frac{1}{n}\right) W_t - \left(1 - \frac{1}{n}\right) X_t^{(1)} + \frac{2n-1}{2n}. \end{aligned}$$

Let $\mu_t := \mathbb{E}(W_t)$. Taking total expectation in the previous expression, we get

$$\mu_{t+1} = \left(1 - \frac{1}{n}\right)\mu_t - \left(1 - \frac{1}{n}\right)\mathbb{P}\big(X_t^{(1)} = 1\big) + \frac{2n-1}{2n}. \tag{12}$$

We now estimate the probability in (12). Since $X_0 = 0$, for $t \le n$,

$$\mathbb{P}\big(X_t^{(1)} = 1\big) = \left[1 - \left(1 - \frac{1}{n} \right)^t \right] \frac{1}{2}. \tag{13}$$

To obtain (13), follow the bit at coordinate 1 at time t backwards in time: at time $t-1$ it was at coordinate 2, at time $t-2$ it was at coordinate 3, etc. At time t it is at 0 unless it was updated at least once along this progression to the left, and the last time that it was updated, it was updated to a 1. The probability it was never updated along its trajectory is $(1-(1/n))^t$, as we require that coordinates $2, 3, \ldots, t, t+1$ at times $t, t-1, \ldots, 2, 1$ respectively have not been chosen for updates. The probability is thus $[1-(1-(1/n))^t]$ that at least one of these coordinates is chosen; the factor of $\frac{1}{2}$ appears because we need the last update to be to 1. Each update is independent of the chosen coordinate and the previous updates.

We now look at a recursive relation for μ_t,

$$\mu_{t+1} = C_1 \mu_t - \frac{C_1}{2}\big[1 - C_1^t\big] + C_2, \tag{14}$$

where (14) is obtained by plugging (13) into (12), and the constants are

$$C_1 := \left(1 - \frac{1}{n}\right), \qquad C_2 := \left(\frac{2n-1}{2n}\right).$$

Note that $\mu_0 = 0$. The following lemma obtains a solution of (14).

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

Lemma 4. *The solution of the recursive relation* (14) *is*

$$\mu_t = \left(\frac{t-n}{2}\right) \cdot C_1^t + \frac{n}{2}.$$

Proof. Clearly $\mu_0 = 0$, and the reader can check that μ_t obeys the recursion. □

Note that we can write $W_t = g_t(R_1, R_2, \ldots, R_t, U_1, U_2, \ldots, U_t)$ for some function g_t, and that the variables $R_1, \ldots, R_t, U_1, \ldots, U_t$ are independent. The following lemma is used in proving that W_t has variance of order n.

Lemma 5. *For* $1 \le i \le t$ *and* $1 \le t \le n$,

$$\max_{\substack{r_1,\ldots,r_t\\ u_1,\ldots,u_t}} \left| g_t(r_1, \ldots, r_i, \ldots, r_t, u_1, \ldots, u_t) - g_t(r_1, \ldots, r_i \oplus 1, \ldots, r_t, u_1, \ldots u_t) \right| \le 2.$$

Proof. Any sequence of coordinates $\{u_s\}_{s=1}^t$ in $\{1, \ldots, n\}$ and bits $\{r_s\}_{s=1}^t$ in $\{0, 1\}$ determine inductively a sequence $\{x_s\}_{s=0}^t$ in $\{0, 1\}^n$, by updating at time s the configuration x_s by adding the bit r_s at coordinate u_s followed by an application of the transformation f. We call a sequence of pairs $\{(u_s, r_s)\}_{s=1}^t$ a *driving sequence*, and the resulting walk in the hypercube $\{x_s\}_{s=0}^t$ the *configuration sequence*. We write $g_t = g_t(r_1, \ldots, r_t, u_1, \ldots, u_t)$, $g'_t = g_t(r'_1, \ldots, r'_t, u'_1, \ldots, u'_t)$.

Consider a specific driving sequence of locations and update bits, $\{(r_s, u_s)\}_{s=1}^t$, and a second such driving sequence $\{(r'_s, u'_s)\}_{s=1}^t$, which together satisfy

- $u'_s = u_s$ for $1 \le s \le t$,
- $r'_s = r_s$ for $1 \le s \le t$ and $s \ne s_0$,
- $r'_{s_0} = r_{s_0} \oplus 1$.

Thus, the two driving sequences agree everywhere except for at time s_0, where the update bits differ. We want to show that $|g_t - g'_t| \le 2$ for any $t \le n$.

Let $\{x_s\}_{1 \le s \le t}$ and $\{y_s\}_{1 \le s \le t}$ be the two configuration sequences in $\{0, 1\}^n$ obtained, respectively, from the two driving sequences. We will show inductively that the Hamming distance $d_s := d_s(x_s, y_s) := \sum_{j=1}^n \left|x_s^{(j)} - y_s^{(j)}\right|$ satisfies $d_s \le 2$ for $s \le t$, and hence the maximum weight difference $|g_t - g'_t|$ is bounded by 2.

Clearly $x_s = y_s$ for $s < s_0$, since the two driving sequences agree prior to time s_0, whence $d_s = 0$ for $s < s_0$.

We now consider d_{s_0}. Let $\ell = u_{s_0} = u'_{s_0}$ be the coordinate updated in both x_{s_0} and y_{s_0}, and as before let $x'_{s_0-1} = \left(x_{s_0-1}^{(1)}, \ldots, x_{s_0-1}^{(\ell)} \oplus r_{s_0}, \ldots, x_{s_0-1}^{(n)}\right)$, $y'_{s_0-1} = \left(y_{s_0-1}^{(1)}, \ldots, y_{s_0-1}^{(\ell)} \oplus r'_{s_0}, \ldots, y_{s_0-1}^{(n)}\right)$. Since $r_{s_0} \ne r'_{s_0}$ but $x_{s_0-1} = y_{s_0-1}$, the configurations x'_{s_0-1} and y'_{s_0-1} have different parities. Recalling that $x_{s_0} = f(x'_{s_0-1})$ and $y_{s_0} = f(y'_{s_0-1})$, we consequently have that $x_{s_0}^{(n)} \ne y_{s_0}^{(n)}$. Since x_{s_0} and y_{s_0} agree at all other coordinates except at $\ell - 1$, we have $d_{s_0} \le I\{\ell \ne 1\} + 1 \le 2$.

Next, suppose that $d_s = 1$ for some time $s \ge s_0$, so that for some $\ell \in \{1, \ldots, n\}$ we have $x_s^{(j)} = y_s^{(j)}$ for $j \ne \ell$ and $x_s^{(\ell)} \ne y_s^{(\ell)}$. Since $r_{s+1} = r'_{s+1}$ and $u_{s+1} = u'_{s+1}$, after adding the same update bit at the same coordinate in the configurations x_s and y_s, but before applying f, the

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

resulting configurations will still have a single disagreement at ℓ. Thus, after applying f to obtain the configurations at time $s+1$, we have $x_{s+1}^{(n)} \neq y_{s+1}^{(n)}$, but

$$d_{s+1} = \sum_{j=1}^{n-1} \left|x_{s+1}^{(j)} - y_{s+1}^{(j)}\right| + \left|x_{s+1}^{(n)} - y_{s+1}^{(n)}\right| = \sum_{j=2}^{n} \left|x_s^{(j)} - y_s^{(j)}\right| + 1 \leq d_s + 1 \leq 2.$$

(If $\ell = 1$, then $d_{s+1} = 1$.) Thus, $d_{s+1} \leq 2$.

Finally, consider the case that $d_s = 2$ for $s \geq s_0$. Again, $u_{s+1} = u'_{s+1}$ and $r_{s+1} = r'_{s+1}$. After updating x_s and y_s with the same bit at the same coordinate, but before applying f, the two configurations still differ at exactly these two coordinates. Thus, $x_{s+1}^{(n)} = y_{s+1}^{(n)}$, and

$$d_{s+1} = \sum_{j=1}^{n-1} \left|x_{s+1}^{(j)} - y_{s+1}^{(j)}\right| + 0 = \sum_{j=2}^{n-1} \left|x_s^{(j)} - y_s^{(j)}\right| \leq d_s \leq 2.$$

(Again, the sum is 1 if one of the two disagreements at time s is at coordinate 1.)

We now have that $d_s \leq 2$ for all $s \leq t$: for $s \leq s_0$, we have $d_s = 0$, and $d_{s_0} = 1$; for $s \geq s_0$, if $d_s \leq 2$ then $d_{s+1} \leq 2$. It then follows in particular that $d_t \leq 2$ and that $|g_t - g'_t| \leq 2$. □

Lemma 6.

$$\max_{\substack{r_1,\dots,r_t \\ u_1,\dots,u_t,u'_i}} \left|g_t(r_1, \dots, r_t, u_1, \dots, u_i, \dots, u_t) - g_t(r_1, \dots, r_t, u_1, \dots, u'_i, \dots, u_t)\right| \leq 2.$$

Proof. Again, if two trajectories differ only in the coordinate selected at time i, then the weight at time t can differ by at most 2. Fix the time $1 \leq t \leq n$, and consider the dynamics of the number of coordinates at which the two trajectories differ at time $k < t$.

The two trajectories agree with each other until time i because the same random bits and locations are used to define these trajectories. At time i, we add the same random bit r_i to update both trajectories, but use coordinate u_i for the first trajectory and coordinate u'_i in the second trajectory. If $r_i = 0$, then clearly the two trajectories continue to agree at time $k \geq i$.

Now suppose that $r_i = 1$. Let b_1, b_2 be the bits at coordinates u_i, u'_i in the *first* trajectory at time $i-1$, and b_3, b_4 be the bits at coordinates u_i, u'_i in the *second* trajectory at time $i-1$. Note that since the trajectories are identical for times less than i, $b_1 = b_3$ and $b_2 = b_4$. For all values of (b_1, b_2, b_3, b_4) satisfying $b_1 = b_3$ and $b_2 = b_4$, there are two disagreements between the trajectories at coordinates $u_i - 1$, $u'_i - 1$ at time i. (If $u_i - 1 < 0$ or $u'_i - 1 < 0$, then there is a single disagreement). The appended bit agrees, since $(b_1 \oplus 1) \oplus b_2 = b_1 \oplus (b_2 \oplus 1) = b_3 \oplus (b_4 \oplus 1)$. This takes care of what happens at time i when the single disagreement between update coordinates occurs: at time i the Hamming distance is bounded by 2.

Now we consider an induction on the Hamming distance, showing that at all times the Hamming distance is bounded by two.

Case A. Suppose that the trajectories differ at two coordinates, say ℓ_1, ℓ_2 at time $k > i$. Since the two trajectories only differ in the updated coordinate at time i with $i < k$, the chosen update coordinate and the chosen update bit are the same for both trajectories at time k. Let b_1, b_2 be the bits at coordinates ℓ_1, ℓ_2 in the *first* trajectory at time k, and b_3, b_4 the bits at coordinates ℓ_1, ℓ_2 in the *second* trajectory at time k. Necessarily, $b_1 \neq b_3$ and $b_2 \neq b_4$. There are two subcases to consider.

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

First, $(b_1, b_2, b_3, b_4) = (0, 1, 1, 0)$ or $(b_1, b_2, b_3, b_4) = (1, 0, 0, 1)$. If $u_k \notin \{\ell_1, \ell_2\}$, then the trajectories continue to have the same two disagreements at these coordinates shifted by one at time $k+1$, since the updated coordinate and the update bit is the same for both trajectories. Also, the new bit which is appended agrees, since $b_1 \oplus b_2 = 1 = b_3 \oplus b_4$, and all other bits in the mod 2 sum agree. So the Hamming distance remains bounded by two, allowing the possibility that the Hamming distance decreases if $\ell_1 \wedge \ell_2 = 1$.

Supposing that $u_k \in \{\ell_1, \ell_2\}$, without loss of generality, assume that $u_k = \ell_1$. These disagreements propagate to time $k+1$, allowing for the possibility of one being eliminated if it occurs at coordinate 1. For $r_k = 0$, the appended bit will agree since $(b_1 \oplus 0) \oplus b_2 = 1 = (b_3 \oplus 0) \oplus b_4$ and all other bits in the mod 2 sum agree. For $r_k = 1$, the appended bit will still agree since $(b_1 \oplus 1) \oplus b_2 = 1 = (b_3 \oplus 1) \oplus b_4$. This means that at time $k+1$ the Hamming distance is bounded by 2.

The second subcase has $(b_1, b_2, b_3, b_4) = (1, 1, 0, 0)$ or $(b_1, b_2, b_3, b_4) = (0, 0, 1, 1)$. If $u_k \notin \{\ell_1, \ell_2\}$, then the trajectories continue to have the same two disagreements (unless one of the disagreements is at coordinate 1), the appended bit agrees (since $1 \oplus 1 = 0 \oplus 0$), and the Hamming distance remains bounded by 2.

Suppose that $u_k = \ell_1$. If $r_k = 0$, then the two disagreements persist (or one is eliminated because it occurred at coordinate 1) and the appended bit agrees. If $r_k = 1$, then the two disagreements persist, and again the appended bit agrees, because now $0 \oplus 1 = 1 \oplus 0$.

Therefore, the Hamming distance remains bounded by 2 at time $k+1$.

Case B. Suppose that the trajectories differ at one coordinate, say ℓ, at time $k > i$. Again, there are two subcases to consider.

First, $u_k \neq \ell$. The disagreement persists unless $u_k = 1$, and the appended bit now disagrees. Thus the Hamming distance is bounded by 2 at time $k+1$.

Second, $u_k = \ell$. The disagreement persists at u_k (unless $u_k = 1$), and the appended bit now disagrees. Again, the Hamming distance is bounded by 2 at time $k+1$.

Thus, by induction, the Hamming distance between the two trajectories remains always bounded by 2. As a consequence, the difference in the Hamming *weight* remains never more than 2. □

Lemma 7. *If* $W_0 = 0$, *then* $\mathrm{Var}(W_t) \leq 4t$.

Proof. We use the following consequence of the Efron–Stein inequality. Suppose that $g : \mathcal{X}^n \to \mathbb{R}$ has the property that, for constants $c_1, \ldots, c_n > 0$, $\sup_{x_1,\ldots,x_n,x_i'} |g(x_1, \ldots, x_n) - g(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i$, and if $X_1, \ldots, X_n$ are independent variables, and $Z = g(X_1, \ldots, X_n)$ is square-integrable, then $\mathrm{Var}(Z) \leq 4^{-1} \sum_i c_i^2$. (See, for example, [6, Corollary 3.2].)

This inequality, together with Lemmas 5 and 6, show that $\mathrm{Var}(W_t) \leq \frac{1}{2} \sum_{i=1}^{2t} 2^2 = 4t \leq 4n$ for $t \leq n$. □

Proof of Theorem 1(ii). Plugging $t = n - n^\alpha$ into Lemma 4, where $\frac{1}{2} < \alpha < 1$, we get

$$\mathbb{E}(W_t) = \mu_t = \frac{n}{2} - \left(1 - \frac{1}{n}\right)^{n-n^\alpha} \frac{n^\alpha}{2} \leq \frac{n}{2} - \frac{1}{2e} n^\alpha. \tag{15}$$

For any real-valued function h on S and probability μ on S, write $E_\mu(h) := \sum_{x \in S} h(x)\mu(x)$. Similarly, $\mathrm{Var}_\mu(h)$ is the variance of h with respect to μ. As stated earlier, $W(x)$ is the Hamming weight of $x \in S$. The distribution of the random variable W under the stationary distribution

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

π (uniform on $\{0, 1\}^n$) is binomial with parameters n and $1/2$, whence $E_\pi(W) = n/2$ and $\mathrm{Var}_\pi(W) = n/4$.

Let $c > 0$ be a constant, and $A_c := (n/2 - c\sqrt{n}, \infty)$. Chebyshev's inequality yields $\pi\{W \in A_c\} \geq 1 - 1/4c^2$. Thus, we can pick c so that this is at least $1 - \eta$ for any $\eta > 0$.

Fix $\frac{1}{2} < \alpha < 1$. For $t_n = n - n^\alpha$, by (15), $\mathbb{P}_0(W_{t_n} \in A_c) = \mathbb{P}_0(W_{t_n} > n/2 - c\sqrt{n}) \leq \mathbb{P}_0\big(W_{t_n} - \mathbb{E}(W_{t_n}) \geq n^\alpha/2\mathrm{e} - c\sqrt{n}\big)$. Since

$$\frac{n^\alpha}{2\mathrm{e}} - cn^{1/2} = n^{1/2}\underbrace{\left(\frac{n^{\alpha-1/2}}{2\mathrm{e}} - c\right)}_{\delta_n(c)},$$

we again have, by Chebyshev's inequality,

$$\mathbb{P}_0(W_{t_n} \in A_c) \leq \frac{\mathrm{Var}(W_{t_n})}{n\delta_n(c)^2} \leq \frac{4t_n}{n\delta_n(c)^2} \leq \frac{4}{\delta_n(c)^2}.$$

The last inequality follows from Lemma 7, since $t_n \leq n$.

Finally, $\left\| Q_1^{t_n}(\mathbf{0}, \cdot) - \pi \right\|_{\mathrm{TV}} \geq |\pi(W \in A_c) - \mathbb{P}_0(W_{t_n} \in A_c)| \geq 1 - 1/4c^2 - 4/\delta_n(c)^2$. We can take, for example, $c_n = \log n$ so that $\delta_n(c_n) \to \infty$, in which case the bound above is $1 - o(1)$. □

5. A related chain

We now consider a Markov chain Q_2 on $\{0, 1\}^{2m}$ related to the the chain Q_1. One step of the Q_2 chain again consists of combining a stochastic move with f. Instead of updating a random coordinate, now the coordinate is always the 'middle' coordinate. Thus, when at x, first we have the random move $x \mapsto x' = (x_1, x_2, \ldots, x_m \oplus R, x_{m+1}, \ldots, x_{2m})$, where R is a an independent random bit. Afterwards, again the transformation f is applied to yield the new state $f(x')$.

Theorem 2. *For all x, $\left\| Q_2^{(n)}(x, \cdot) - \pi \right\|_{\mathrm{TV}} = 0$ for all $n = 2m$ where $m \geq 1$.*

Remark 3. Note that if the transformation is a circular shift instead of f, then this would be equivalent to systematic scan, which trivially yields an exact uniform sample in exactly $n = 2m$ steps. Thus, this chain can be viewed as a small perturbation of systematic scan which is Markovian and still yields an exact sample in $n = 2m$ steps.

Proof. We denote by $(R_1, R_2, \ldots)$ the sequence of bits used to update the chain. To demonstrate how the walk evolves with time by means of an example, Table 1 shows the coordinates of Y_t at different t for $2m = 6$, when starting from $\mathbf{0}$.

Let $n = 2m$ be an even integer, and let $Z_1, Z_2, \ldots, Z_m$ be the random variables occupying the n coordinates at time $t = n$. The following relationships hold for any starting state $x = (x_1, x_2, \ldots, x_n)$:

$$\begin{aligned} Z_1 &= R_1 \oplus R_{m+2} \oplus \bigoplus_{i=1}^{2m} x_i, \\ Z_2 &= R_2 \oplus R_{m+3} \oplus x_1, \\ &\vdots \\ Z_{m-1} &= R_{m-1} \oplus R_{2m} \oplus x_{m-2}, \end{aligned}$$

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

TABLE 1. Evolution of coordinates with time for $2m = 6$.

Coordinate	1	2	3	4	5	6
$t = 0$	0	0	0	0	0	0
$t = 1$	0	R_1	0	0	0	R_1
$t = 2$	R_1	R_2	0	0	R_1	R_2
$t = 3$	R_2	R_3	0	R_1	R_2	R_3
$t = 4$	R_3	R_4	R_1	R_2	R_3	$R_1 \oplus R_4$
$t = 5$	R_4	$R_1 \oplus R_5$	R_2	R_3	$R_1 \oplus R_4$	$R_2 \oplus R_5$
$t = 6$	$R_1 \oplus R_5$	$R_2 \oplus R_6$	R_3	$R_1 \oplus R_4$	$R_2 \oplus R_5$	$R_3 \oplus R_6$

$$\begin{aligned} Z_m &= R_n \oplus x_{m-1}, \\ Z_{m+1} &= R_1 \oplus R_{m+1} \oplus x_m, \\ Z_{m+2} &= R_2 \oplus R_{m+2} \oplus x_{m+1}, \\ &\vdots \\ Z_{2n} &= R_n \oplus R_{2n} \oplus x_{2n-1}. \end{aligned}$$

This is because at $t = 1$ the random variable at coordinate n is $R_1 + \bigoplus_{i=1}^n x_i$. At time $t = m + 1$, this random variable moves to coordinate m because of successive shift register operations. Because the coordinate updated at any time along this chain is m, we have that at time $t = m + 2$, the random variable at coordinate $m - 1$ is $R_1 + R_{m+2} \oplus \bigoplus_{i=1}^n x_i$. Again, because of successive shift register operations, the random variable $R_1 + R_{n+2} \oplus \bigoplus_{i=1}^n x_i$ moves to coordinate 1 at time n. The random variables at other coordinates can be worked out similarly. Thus, the above system of equations can be written in matrix form as $Z = BR + \vec{x}$, where $Z = (Z_1, \ldots, Z_n)^\top$, $R = (R_1, \ldots, Z_n)^\top$, and

$$B_{n\times n} = \begin{bmatrix} I_{m\times m} & C_{m\times m} \\ I_{m\times m} & I_{m\times m} \end{bmatrix}, \quad C_{m\times m} = \begin{bmatrix} 0_{(m-1)\times 1} & I_{(m-1)\times(m-1)} \\ 0_{1\times 1} & 0_{1\times(m-1)} \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} \oplus_{i=1}^n x_i \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix}.$$

Note that $\det(B) = \det(I) \times \det(I - II^{-1}C) = \det(I - C) = 1 \neq 0$. The last equality follows since $\det(I - C) = 1$ because $I - C$ is an upper triangular matrix with ones along the main diagonal. Hence, B is an invertible matrix and, if $z \in \{0, 1\}^n$, $\mathbb{P}(Z = z) = \mathbb{P}\big(R = B^{-1}(z - \vec{x})\big) = 1/2^n$, where the last equality follows from the fact that R is uniform over $S = \{0, 1\}^n$. Thus, the state along the Q_2 chain at $t = 2m = n$ is uniform over S, and $\big\| Q_2^{(n)}(\mathbf{0}, \cdot) - \pi \big\|_{\mathrm{TV}} = 0$, n even. □

6. Conclusion and open questions

We have shown that the ‘shift-register’ transformation speeds up the mixing of the walk on the hypercube for which the stationary distribution is uniform. The shift-register transformation is a good candidate for a deterministic mixing function, as shift registers were used for early pseudo-random number generation.

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press

One of our original motivations for analyzing this chain was in part that the uniform distribution corresponds to the infinite-temperature Ising measure. Indeed, of great interest are chains on product spaces having non-uniform stationary distributions, such as Gibbs measures. Finding deterministic transformations which speed up mixing for non-uniform distributions remains a challenging and intriguing open problem.

Acknowledgement

The authors thank the reviewer for helpful comments and pointing out relevant references.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Andersen, H. C. and Diaconis, P. (2007). Hit and run as a unifying device. *Journal de la société française de statistique* **148**, 5–28.

[2] Ben-Hamou, A. and Peres, Y. (2018). Cutoff for a stratified random walk on the hypercube. *Electron. Commun. Prob.* **23**, 1–10.

[3] Ben-Hamou, A. and Peres, Y. (2021). Cutoff for permuted Markov chains. Preprint, arXiv:2104.03568.

[4] Bierkens, J. (2016). Non-reversible metropolis-hastings. *Statist. Comput.* **26**, 1213–1228.

[5] Boardman, S., Rudolf, D. and Saloff-Coste, L. (2020). The hit-and-run version of top-to-random. Preprint, arXiv:2009.04977.

[6] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities*. Oxford University Press.

[7] Chatterjee, S. and Diaconis, P. (2020). Speeding up Markov chains with deterministic jumps. *Prob. Theory Relat. Fields* **178**, 1193–1214.

[8] Chen, F., Lovász, L. and Pak, I. (1999). Lifting Markov chains to speed up mixing. In *Proc. 31st Ann. ACM Symp. Theory of Computing*, pp. 275–281.

[9] Diaconis, P. and Graham, R. (1992). An affine walk on the hypercube. *J. Comput. Appl. Math.* **41**, 215–235.

[10] Diaconis, P., Graham, R. L. and Morrison, J. A. (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1**, 51–72.

[11] Diaconis, P., Holmes, S. and Neal, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Prob.* **10**, 726–752.

[12] Diaconis, P. and Ram, A. (2000). Analysis of systematic scan Metropolis algorithms using Iwahori–Hecke algebra techniques. *Michigan Math. J.* **48**, 157–190.

[13] Golomb, S. W. (2017). *Shift Register Sequences*, 3rd edn. World Scientific, Singapore.

[14] Guo, B.-N. and Qi, F. (2013). Refinements of lower bounds for polygamma functions. *Proc. Amer. Math. Soc.* **141**, 1007–1015.

[15] Jensen, S. and Foster, D. (2014). A level-set hit-and-run sampler for quasi-concave distributions. *Proc. Mach. Learn. Res.* **33**, 439–447.

[16] Levin, D. A. and Peres, Y. (2017). *Markov Chains and Mixing Times*, 2nd edn. American Mathematical Society, Providence, RI.

[17] Wilson, D. B. (1997). Random random walks on $\mathbb{Z}_2^d$. *Prob. Theory Relat. Fields* **108**, 441–457.

https://doi.org/10.1017/jpr.2022.37 Published online by Cambridge University Press