MIXING OF THE UPPER TRIANGULAR MATRIX WALK

YUVAL PERES AND ALLAN SLY

ABSTRACT. We study a natural random walk over the upper triangular matrices, with entries in the field $\mathbb{Z}_2$, generated by steps which add row $i + 1$ to row i. We show that the mixing time of the lazy random walk is $O(n^2)$ which is optimal up to constants. Our proof makes key use of the linear structure of the group and extends to walks on the upper triangular matrices over the fields $\mathbb{Z}_q$ for q prime.

1. INTRODUCTION

For $n \geq 2$ and q a prime let $G_n(q)$ be the group of $n \times n$-upper triangular matrices under multiplication with entries in the field $\mathbb{Z}_q$ and 1 along the diagonal. The upper triangular walk on $G_n(2)$ is the Markov chain where each step is given by choosing a random $1 \leq i \leq n-1$ and adding row $i + 1$ to row i. The walk is stationary with respect to the uniform distribution on $G_n(2)$ and our main result establishes the order of the mixing time answering a question of Stong [18] and of Arias-Castro, Diaconis and Stanley [3] and Problem 14 of [14].

Theorem 1. *There exists a constant $C > 0$ such that the mixing time of the lazy upper triangular walk satisfies $t_{\mathrm{MIX}}^{(n)} \leq Cn^2$.*

This bound is tight up to constants as it is known that the mixing time is at least order n^2 (see e.g. [18]). Indeed the projection onto the rightmost column of the matrix is itself a Markov chain given by an East model (see Section 2.1) which has mixing time of order n^2.

1.1. **Background.** The random walks on $G_n(q)$ have received significant attention. There are two natural extensions to fields $\mathbb{Z}_q$ for q prime. We will focus on the walk considered by Coppersmith and Pak [5, 16]. This walk is given by the set of generators $\{I + aE_{i,i+1} : 1 \leq i \leq n-1, a \in \mathbb{Z}_q\}$ where $E_{i,j}$ denotes the $n \times n$ matrix with 1 at position (i, j) and 0 everywhere else. We denote this walk as $\mathcal{W}_1$. An equivalent description is that each step entails choosing a uniformly random $1 \leq i \leq n-1$ and $a \in \mathbb{Z}_q$ and adding row $i+1$ multiplied by a to row i. This is an ergodic Markov chain reversible with respect to the uniform distribution on $G_n(q)$. When $q = 2$ this corresponds to the lazy version of our upper triangular walk of Theorem 1.

The other extension, which we will denote $\mathcal{W}_2$, is the walk on $G_n(q)$ given by generators $\{I \pm E_{i,i+1} : 1 \leq i \leq n-1\}$ which has been studied by Diaconis, Saloff-Coste, Stong and others [3, 7, 18]. In the case $q = 2$ it is exactly the upper triangular walk of Theorem 1. Random walks of the form $\mathcal{W}_2$ were first studied by Zack [19] in the case $n = 3$ as a random walk on the Heisenberg Group. In the setting of q growing and n fixed, Diaconis and Saloff-Coste [7–9] introduced a number of sophisticated techniques giving sharp rates of convergence.

In the case of n growing the first results follow from work of Ellenberg [11] which gives an upper bound of order n^7. This was subsequently improved by Stong who gave sharp bounds on the spectral gap of $\Theta(n^{-1}q^{-2})$ which translates into an upper bound on the mixing time of $n^3 q^2 \log q$.

The use of character theory has proved to be a key tool in analyzing random walks on groups (see e.g. [6, 10]). One reason random walks on $G_n(q)$ have resisted analysis is the lack of understanding of the character theory of $G_n(q)$ which is "considered unknowable" [3]. Arias-Castro, Diaconis and Stanley [3] approached the problem using the super-character theory of Andre, Carter and Yan and gave a sharp analysis of a related chain whose generators are the conjugacy classes of $I \pm E_{i,i+1}$. Using comparison, this approach yielded an upper bound on the mixing time of $O(n^4 q^2 \log(n) \log(q))$ for the chain $\mathcal{W}_2$.

Using a stopping time argument, Pak [16] showed that when $q > 2n^3$, the mixing time of the walk $\mathcal{W}_1$ is $O(n^{5/2})$. Under the same assumption on the growth of q, this was subsequently improved by Coppersmith and Pak [5] to a mixing time of $O(n^2)$ which is optimal. By taking q so large, they can effectively assume that once an entry is updated to a non-zero entry it remains non-zero.

Our work resolves the case of fixed q and also gives new bounds when q is growing with n.

Theorem 2. *There exists an absolute constant $C > 0$ such that the mixing time of the lazy upper triangular walk $\mathcal{W}_1$ on $G_n(q)$ satisfies*

$$t_{\mathrm{MIX}}^{(n)} \leq C n^2 \log q.$$

This yields new bounds when $q < 2n^3$ and so is outside of the regime studied by Coppersmith and Pak. As in the case $q = 2$ there is a natural lower bound of order n^2. We conjecture that the mixing time is order n^2 for all q which interpolates between our fixed q result and Coppersmith and Pak's large q result.

Our proof crucially makes use of the linear structure of $G_n(q)$. When multiplying by $I + aE_{n-1,n}$ we observe that the choice of a only affects the final column. Indeed we can write the off-diagonal portion of the final

column at time t as $b_0 + \sum_{k=1}^{J} a_k b_k$ for $b_k \in \mathbb{Z}_p^{n-1}$ where the $a_k \in \mathbb{Z}_p$ are uniformly chosen independent of the b_k and the remainder of the matrix. It follows that the final column will be uniform and independent of the rest of the matrix if the vectors $\{b_1, \ldots, b_J\}$ span $\mathbb{Z}_p^{n-1}$.

We prove that for large enough C and $t = Cn^2 \log q$ that the b_k do span with high probability by expressing the vectors as projections at random times of another matrix walk. This walk is a reflection of the original walk involving column rather than row operations. By the linear structure of the walk, the inner product of the a fixed vector and the projection of the Markov chain evolves as a site in the East model. Then using known spectral gap estimates for the East model and large deviation results for Markov chains we obtain the necessary bounds on when the b_k span $\mathbb{Z}_p^{n-1}$. Iterating this argument over all the columns completes the proof.

1.2. **The East model.** A key role is played by the one dimensional East model which is an extensively studied example of a kinetically constrained spin model (see [1, 4] and references therein). Formally the East model with parameter $0 < p < 1$ is a Markov chain whose state space is

$$\mathcal{H}_n = \{h = (h_1, \ldots, h_L) \in \{0, 1\}^n : h_1 = 1\}.$$

The dynamics are given as follows. In each step choose a uniformly random $1 \le i \le n-1$ and

- If $h_i = 1$ set h_{i+1} to 1 with probability p and 0 with probability $1 - p$.
- If $h_i = 0$ do nothing.

The dynamics are reversible with respect to the i.i.d. Bernoulli measure where the h_i are 1 with probability p for $2 \le i \le n$. When $q = 2$ the East Model with $p = 1/2$ describes the dynamics of a single column of the upper triangular matrix walk.

2. Preliminaries

It will be convenient for our proof to consider the continuous time version of $\mathcal{W}_1$ whose generator is given by

$$(\mathcal{L}f)(X) = \sum_{i=1}^{n-1} \frac{1}{q} \sum_{a \in \mathbb{Z}_q} f((I + aE_{i,i+1})X) - f(X) \qquad (2.1)$$

This version admits the following description: for each $1 \le i \le n-1$ we have a Poisson clock and when clock i rings we choose a uniform $a \in \mathbb{Z}_q$ and add row $i + 1$ multiplied by a to row i. Note that this speeds the chain up by a factor of $n - 1$ of the original. Let X_t denote this Markov chain. The continuous time representation has the useful property that for $1 < n' < n$

the process X_t projected onto the $n' \times n'$-submatrix of its first n' rows and columns is itself a Markov chain whose law is given by the walk on $G_{n'}(q)$.

In this paper we consider mixing measured in the total variation distance. For two probability measures ν_1, ν_2 on a countable space Ω the **total variation distance** is defined as

$$\|\nu_1 - \nu_2\|_{\text{TV}} = \max_{A \subset \Omega} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sum_{x \in \Omega} |\nu_1(x) - \nu_2(x)| \ .$$

For an ergodic Markov chain Y_t with stationary distribution ν we define

$$t_{\text{MIX}}(s) = \inf\{t : \max_{y \in \Omega} \|\mathbb{P}_y(Y_t \in \cdot) - \nu\|_{\text{TV}} < s\} \ ,$$

and the (worst case) **mixing time** is defined as $t_{\text{MIX}} = t_{\text{MIX}}(1/2e)$. We denote the walk $\mathcal{W}_1$ on $G_n(q)$ with generator given in (2.1) as $X_t = X_t^n$, its mixing time as $t_{\text{MIX}}^{(n)}$ and we let $d_n(t)$ denote $\|\mathbb{P}(X_t^n \in \cdot) - \pi^n\|_{\text{TV}}$ where π^n is the uniform distribution on $G_n(q)$ and the stationary distribution of X_t.

For an ergodic reversible Markov chain we denote the eigenvalues of the generator as $0 = \lambda_1 < \lambda_2 \leq \ldots$ and the spectral gap, the second eigenvalue, as $\lambda = \lambda_2$. For any ergodic reversible finite Markov chain $\lambda^{-1} \leq t_{\text{MIX}}$ (c.f., e.g. [2,14]). The mixing time of the lazy random walks and continuous time random walks are closely related. Theorem 20.3 of [14] implies that if the mixing time of the continuous time random walk $\mathcal{W}_1$ is $O(n \log q)$ then the mixing time of the discrete time lazy version of the walk is $O(n^2 \log q)$. As such for the rest of the paper we consider the continuous time version of the walk.

We will make use of Hoeffding type bounds for Markov chains. There are various versions including results of Gillman [12] and sharper exponential rates given by León and Perron [13]. We will make use of Theorem 3.4 of [15] by Lezaud which gives a continuous time Markov chain analogue. In particular it shows that for a reversible Markov chain Y_t with stationary distribution ν, and for any subset $A \subset \Omega$,

$$\mathbb{P}\left(\left| \int_0^t I(Y_t \in A)dt - t\nu(A) \right| > \varepsilon t \right) \leq \frac{2}{\sqrt{\nu_{\min}}} \exp(-t\varepsilon^2 \lambda / 12) \qquad (2.2)$$

where $\nu_{\min} = \min\{\nu(x) : x \in \Omega, \nu(x) > 0\}$.

2.1. **Continuous time East models.** We will also use the continuous time version of the East Model. Its dynamics on $\mathcal{H}_n$ are given as follows. For each $1 \leq i \leq n-1$ we have a rate 1 Poisson clock. When clock i rings:

- If $h_i = 1$ set h_{i+1} to 1 with probability p and 0 with probability $1 - p$.
- If $h_i = 0$ do nothing.

Again the dynamics is reversible with respect to the i.i.d. Bernoulli measure where the h_i are 1 with probability p for $2 \leq i \leq n$. Let $\theta_p(n)$ denote the spectral gap of the East model of length n with parameter p. Aldous and Diaconis [1] showed that $\theta_p := \inf_n \theta_p(n) > 0$ for all p, establishing a bounded spectral gap for the infinite dynamics, and also giving the asymptotics as p tends to 0.

We now consider a natural q-state extension of the East model. Define the state-space as

$$\mathcal{H}_n^q = \{h = (h_1, \ldots, h_n) \in \mathbb{Z}_q^n : h_1 = 1\}$$

and again for each $1 \leq i \leq n-1$ we have rate 1 Poisson clocks so that when clock i rings:

- If $h_i \neq 0$ set h_{i+1} to a uniform choice in $\mathbb{Z}_q$.
- If $h_i = 0$ do nothing.

Let $\lambda^q(n)$ denote the spectral gap of this process.

Proposition 2.1. *The spectral gaps satisfy $\lambda^\star := \inf_{q,n} \lambda^q(n) > 0$.*

Proof. Let Z_t denote the q-state East model and let $\psi : \mathcal{H}_n^q \to \mathcal{H}_n$ given by

$$\psi(h)_i = \begin{cases} 0 & \text{if } h_i = 0, \\ 1 & \text{otherwise.} \end{cases}$$

The process $Z_t' = \psi(Z_t)$ has the law of the standard East model with parameter $p = \frac{q-1}{q}$.

We now consider an update of Z_t where clock i rings and $Z_t(i) \neq 0$. The site $i+1$ is updated to 0 with probability $\frac{1}{q}$ and from a uniform choice on $\{1, \ldots, q-1\}$ with probability $\frac{q-1}{q}$. In particular it is uniformly distributed, conditional on the value $Z_t'(i+1)$. It is easily verified that this conditional independence property holds at all times after such an update and is independent of the process at other sites.

So let $\mathcal{B}(t)$ be the event that for each $1 \leq i \leq n-1$ there has been an update from clock i at a time s when $Z_s(i) \neq 0$ before time t. It follows that on the event $\mathcal{B}(t)$ the variable Z_t is uniformly distributed conditional on the value of Z_t'. Hence we have that

$$\|\mathbb{P}(Z_t \in \cdot) - \mathbb{P}(Z \in \cdot)\|_{\mathrm{TV}} \leq \|\mathbb{P}(Z_t' \in \cdot) - \mathbb{P}(\psi(Z) \in \cdot)\|_{\mathrm{TV}} + \mathbb{P}(\mathcal{B}^c(t)), \quad (2.3)$$

where Z is distributed according to the stationary distribution of Z_t. Since the spectral gap of Z_t' is $\theta_{\frac{q-1}{q}}(n)$ we have that

$$\|\mathbb{P}(Z_t' \in \cdot) - \mathbb{P}(\psi(Z) \in \cdot)\|_{\mathrm{TV}} \leq C(n,q) \exp(-t\theta_{\frac{q-1}{q}}(n)), \quad (2.4)$$

for some constant $C(n,q)$ as the spectral gap gives the asymptotic convergence rate to stationarity. Fix some $1 \leq i \leq n-1$. Then equation (2.2) implies that for some $c > 0$ and constant $C'(n,q)$

$$\mathbb{P}\left(\int_0^t I(Z_s(i) \neq 0)ds \leq \tfrac{1}{3}t\right) \leq C'(n,q)\exp(-c\theta_{\frac{q-1}{q}}(n)t). \qquad (2.5)$$

The times at which clock i rings form a rate 1 Poisson process on $[0,t]$ which is independent of the process $Z_t(i)$. Hence the chance that $\mathcal{B}$ fails is given by

$$\mathbb{P}\left(\mathcal{B}^c(t)\right) \leq \mathbb{P}\left(\exists 1 \leq i \leq n-1, \int_0^t I(Z_t(i) \neq 0)dt \leq \tfrac{1}{3}t\right) \qquad (2.6)$$

$$+ (n-1)\mathbb{P}\left(\text{Poisson}(\tfrac{1}{3}t) = 0\right)$$

$$\leq nC'(n,q)\exp(-c\theta_{\frac{q-1}{q}}(n)t) + n\exp(-\tfrac{1}{3}t). \qquad (2.7)$$

Combining equations (2.3), (2.4) and (2.6) we have

$$\|\mathbb{P}(Z_t \in \cdot) - \mathbb{P}(Z \in \cdot)\|_{\text{TV}} \leq C(n,q)\exp(-c\theta_{\frac{q-1}{q}}(n)t)$$

$$+ nC'(n,q)\exp(-c\theta_{\frac{q-1}{q}}(n)t) + n\exp(-\tfrac{1}{3}t)$$

and so the exponential rate of decay is at least $c\theta_{\frac{q-1}{q}} \wedge \frac{1}{3}$. Hence we have that

$$\lambda^q(n) \geq c\theta_{\frac{q-1}{q}}(n) \wedge \frac{1}{3}.$$

Theorem 4.2 of [4] establishes that for some $0 < p_0 < 1$, whenever $p_0 < p < 1$ we have that $\theta_p(n) > \frac{1}{2}$. Combining this with the fact that $\inf_n \theta_p(n) > 0$ for all p we have that

$$\lambda^\star = \inf_{q,n} \lambda^q(n) \geq \frac{1}{3} \wedge \inf_{q,n} c\theta_{\frac{q-1}{q}}(n) > 0$$

which completes the proof.   ∎

It was observed in [3] that the projection onto column j of the upper triangular walk on $G_n(q)$ is given by an East model, specifically the j length q-state East model. Our proof makes use of this fact and additionally that the observation also holds for linear combinations of columns.

3. Proof of Theorem 1

In this section we work exclusively with the continuous time version of the Markov chain. We begin with some notation. By symmetry we will assume that the Markov chain begins at $X_0 = I$. Fix some terminal time T. First we split the clocks and associated row operations into those with

$1 \leq i \leq n - 2$ and those with $i = n - 1$. Let $0 < t_1 < t_2 < \ldots$ denote the times at which the Poisson clocks ring for $1 \leq i \leq n - 2$ and let W_j denote the associated row operation matrices. We denote the σ-algebra generated by these updates as $\mathcal{F}$. Let $N(t) = \max\{j \geq 0 : t_j \leq t\}$ and define the backwards process on the interval $[0, T]$ by $Y_0 = I$ and

$$Y_t = \prod_{j=0}^{N(T)-N(T-t)-1} W_{N(T)-j} = W_{N(T)} W_{N(T)-1} \ldots W_{N(T-t)+1}$$

and for $0 \leq t < t' \leq T$ we let

$$Y_{t,t'} = Y_t^{-1} Y_{t'} = W_{N(T-t)} W_{N(T-t)-1} \ldots W_{N(T-t')+1}.$$

The process Y_t is a Markov process given by the following column dynamics description. For each $1 \leq i \leq n - 2$ we have a Poisson clock and when clock i rings we choose a uniform $a \in \mathbb{Z}_q$ and add column i multiplied by a to column $i + 1$. With this description column n is fixed while the walk on the top left $(n-1) \times (n-1)$ submatrix is equivalent to the walk $\mathcal{W}_1$ on $G_{n-1}(q)$ up to a reflection of the matrix.

Next consider the second type of row operations which have $i = n - 1$. Let $s_1 < s_2 < \ldots$ denote the times at which row n is added to row $n - 1$ and let $a_1, a_2, \ldots$ denote the associated scalers. Note that these are independent of $\mathcal{F}$. Denoting

$$J(t) = \max\{j \geq 0 : s_j \leq t\},$$

as the number of row operations with $i = n - 1$ up to time t. The following expansions of X_T allow us to track the contribution of operations of the second type, taking advantage of the linear nature of $G_n(q)$. By construction, with the definitions above, we have that

$$X_{s_\ell} = (I + a_\ell E_{n-1,n}) Y_{T-s_\ell, T-s_{\ell-1}} X_{s_{\ell-1}} \qquad (3.1)$$

for $1 \leq \ell \leq J(T)$ where $s_0 = 0$. We now show, by induction, that

$$X_{s_\ell} = Y_{T-s_\ell, T} + \sum_{k=1}^{\ell} a_k Y_{T-s_\ell, T-s_k} E_{n-1,n}. \qquad (3.2)$$

For $\ell = 0$ equation (3.2) is immediate and for any upper triangular matrix $Y \in G_n(q)$ we have that

$$E_{n-1,n} Y = E_{n-1,n} \text{ and } E_{n-1,n} E_{n-1,n} = 0. \qquad (3.3)$$

Then by equation (3.1)

$$X_{s_{\ell+1}} = (I + a_{\ell+1}E_{n-1,n})Y_{T-s_{\ell+1},T-s_\ell}X_{s_\ell}$$

$$= (I + a_{\ell+1}E_{n-1,n})Y_{T-s_{\ell+1},T-s_\ell}\left(Y_{T-s_\ell,T} + \sum_{k=1}^{\ell} a_k Y_{T-s_\ell,T-s_k}E_{n-1,n}\right)$$

$$= Y_{T-s_{\ell+1},T} + \sum_{k=1}^{\ell+1} a_k Y_{T-s_{\ell+1},T-s_k}E_{n-1,n}$$

where the final equality follows by expanding the sum, applying equation (3.3) and using the fact that $Y_{T-s_{\ell+1},T-s_{\ell+1}} = I$ completing the inductive step. Finally, since $X_T = Y_{0,J(T)}X_{J(T)}$, it follows from equation (3.2) that

$$X_T = Y_T + \sum_{k=1}^{J(T)} a_k Y_{T-s_k}E_{n-1,n}. \tag{3.4}$$

This expression allows us to separate the mixing in the first $n-1$ columns from that of the final column. Let $\mathcal{M}$ denote set of $n \times n$-matrices with entries in $\mathbb{Z}_p$ and let $\mathcal{M}^*$ denote the sub-space of matrices spanned by $E_{i,n}$ for $1 \le i \le n-1$. Each matrix $Y_{T-s_k}E_{n-1,n}$ is in $\mathcal{M}^*$. Let $\mathcal{A} = \mathcal{A}(T,n)$ denote the event that the collection $\{Y_{T-s_k}E_{n-1,n} : 1 \le k \le J(T)\}$ spans $\mathcal{M}^*$. Since the a_k are independent of the Y_{T-s_k} it follows that on $\mathcal{A}$ the sum $\sum_{k=1}^{J(T)} a_k Y_{T-s_k}E_{n-1,n}$ is uniformly distributed on $\mathcal{M}^*$ and so we have the following proposition.

Proposition 3.1. *The distances to stationarity satisfy*

$$d_n(T) \le d_{n-1}(T) + \mathbb{P}(\mathcal{A}(T,n)^c).$$

Proof. Let X be uniform on $G_n(q)$, the stationary distribution of X_t. As observed above, the Markov chain restricted to the first $n-1$ columns of X_t gives the walk $\mathcal{W}_1$ on $G_{n-1}(q)$. Hence we may couple the first $n-1$ columns of X_T and X except with probability $d_{n-1}(T)$. Moreover the walk on the first $n-1$ columns is $\mathcal{F}$ measurable and independent of the scalars $\{a_1, \ldots, a_{J(T)}\}$.

On the event $\mathcal{A}(T,n)$ we have that $\sum_{k=1}^{J(T)} a_k Y_{T-s_k}E_{n-1,n}$ is uniformly distributed on $\mathcal{M}^*$ and independent of $\mathcal{F}$. Hence we may couple the final column of X_T and X independent of $\mathcal{F}$ when $\mathcal{A}(T,n)$ holds. Hence we may couple X_T and X except with probability $d_{n-1}(T) + \mathbb{P}(\mathcal{A}(T,n)^c)$ which completes the proposition. ∎

It remains to bound the probability of $\mathcal{A}$. For this we use the fact that Y_t is a Markov chain and its first $n-1$ columns are given by the column version of the walk $\mathcal{W}_1$.

Proposition 3.2. *There exist absolute constants $0 < c_1, c_2 < \infty$ such that $\mathbb{P}(\mathcal{A}(T, n)) \geq 1 - e^{-c_1(T - c_2 n \log q)}$.*

Proof. The vectors $\{Y_{T-s_k} E_{n-1,n} : 1 \leq k \leq J(T)\}$ span $\mathcal{M}^*$ if and only if for each non-zero vector $b = (b_1, \ldots, b_{n-1}, 0) \in \mathbb{Z}_q^n \setminus \{(0, \ldots, 0)\}$ we have that for some $1 \leq k \leq J(T)$ that

$$b \cdot Y_{T-s_k} e_{n-1} \neq 0. \tag{3.5}$$

Let $l = \min_j b_j \neq 0$. We may assume without loss of generality that $b_l = 1$ since multiplying b by a non-zero scaler does not affect (3.5).

Now recall our observation that Y_t is the dynamics on upper triangular matrices where for each $1 \leq i \leq n - 2$ we add column i times a random scaler a to column $i + 1$ according to the times of rate 1 Poisson clocks. The process $Z_t = b \cdot Y_t$ is also a Markov chain on vectors of length n with the following description. For each $1 \leq i \leq n - 2$ at rate 1 we add position i times a random scaler a to position $i + 1$. Its initial condition is $Z_0 = b$ and so the value of its first $l - 1$ entries and the final entry are fixed as 0. The value in entry l is fixed at 1 and entries l to $n - 1$ perform the q-state East model of length $n - l$. By Proposition 2.1 this chain has spectral gap at least $\lambda^* > 0$. Its state space is the uniform distribution over q^{n-l-1} possible vectors and so applying equation (2.1) we have that

$$\mathbb{P}\left(\int_0^T I(Z_t e_{n-1} \neq 0) dt \leq \tfrac{1}{3} T\right) \leq q^{n/2} \exp(-c\lambda^* T) \tag{3.6}$$

where $I(\cdot)$ denotes the indicator and c is an absolute constant. The times $T - S_k$ form a rate 1 Poisson process on $[0, T]$ which is independent of the process Y_t. Hence

$$\mathbb{P}\left(\forall 1 \leq k \leq J(T), Z_{T-s_k} e_{n-1} = 0\right)$$
$$\leq \mathbb{P}\left(\int_0^T I(Z_t e_{n-1} \neq 0) dt \leq \tfrac{1}{3} T\right) + \mathbb{P}\left(\text{Poisson}(\tfrac{1}{3} T) = 0\right)$$
$$\leq q^{n/2} \exp(-c\lambda^* T) + \exp(-\tfrac{1}{3} T). \tag{3.7}$$

Taking a union bound over b completes the proof. ∎

We are now ready to prove Theorem 2 which in turn proves Theorem 1

Proof of Theorem 2. As noted above it is sufficient by Theorem 20.3 of [14] to prove an upper bound on the mixing time of the continuous time chain $\mathcal{W}_1$ is $O(n \log q)$. Combining Propositions 3.1 and 3.2 we have that

$$d_n(T) \leq d_{n-1}(T) + e^{-c_1(T - c_2 n \log q)}$$

and hence by induction

$$d_n(T) = \sum_{i=2}^{n} e^{-c_1(T - c_2 i \log q)}.$$

For $c > c_2$ we have that $d_n(cn \log q) = o(1)$ which completes the proof. $\blacksquare$

References

[1] D. Aldous and P. Diaconis, *The asymmetric one-dimensional constrained Ising model: rigorous results*, Journal of Statistical Physics **107** (2002), 945–975.

[2] D. Aldous and J. A. Fill, *Reversible Markov Chains and Random Walks on Graphs.* In preparation, `http://www.stat.berkeley.edu/~aldous/RWG/book.html`.

[3] E. Arias-Castro, P. Diaconis, and R. Stanley, *A super-class walk on upper-triangular matrices*, Journal of algebra **278** (2004), 739–765.

[4] N. Cancrini, F. Martinelli, C. Roberto, and C. Toninelli, *Kinetically constrained spin models*, Probability Theory and Related Fields **140** (2008), 459–504.

[5] D. Coppersmith and I. Pak, *Random walk on upper triangular matrices mixes rapidly*, Probability Theory and Related Fields **117** (2000), 407–417.

[6] P. Diaconis, *Random walks on groups: characters and geometry*, Groups Saint Andrews 2001 in Oxford, 2003, pp. 120–142.

[7] P. Diaconis and L. Saloff-Coste, *Moderate growth and random walk on finite groups*, Geometric and Functional Analysis **4** (1994), 1–36.

[8] P. Diaconis and L. Saloff-Coste, *An Application of Harnaek Inequalities to Random Walk on Nilpotent Quotients*, Proceedings of the Conference in Honor of Jean-Pierre Kahane, 1995, pp. 189.

[9] P. Diaconis and L. Saloff-Coste, *Nash inequalities for finite Markov chains*, Journal of Theoretical Probability **9** (1996), 459–510.

[10] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transpositions*, Probability Theory and Related Fields **57** (1981), 159–179.

[11] J. Ellenberg, *A Sharp Diameter Bound for Upper Triangular Matrices*, Senior Honors Thesis, Department of Mathematics, Harvard University, 1993.

[12] D. Gillman, *A Chernoff Bound for Random Walks on Expander Graphs*, SIAM Journal on Computing **27** (1998), 1203.

[13] C. A. León and F. Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains*, Annals of Applied Probability **14** (2004), 958–970.

[14] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, 2008.

[15] P. Lezaud, *Chernoff-type bound for finite Markov chains*, The Annals of Applied Probability **8** (1998), 849–867.

[16] I. Pak, *Two random walks on upper triangular matrices*, Journal of Theoretical Probability **13** (2000), 1083–1100.

[17] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.

[18] R. Stong, *Random walks on the groups of upper triangular matrices*, The Annals of Probability **23** (1995), 1939–1949.

[19] M. Zack, *A Random Walk on the Heisenberg Group*, University of California, San Diego, 1984.

YUVAL PERES

MICROSOFT RESEARCH, ONE MICROSOFT WAY, REDMOND, WA 98052-6399, USA.
E-mail address: peres@microsoft.com

ALLAN SLY

MICROSOFT RESEARCH, ONE MICROSOFT WAY, REDMOND, WA 98052-6399, USA.
E-mail address: allansly@microsoft.com