

# A BIRTHDAY PARADOX FOR MARKOV CHAINS WITH AN OPTIMAL BOUND FOR COLLISION IN THE POLLARD RHO ALGORITHM FOR DISCRETE LOGARITHM

BY JEONG HAN KIM<sup>1</sup>, RAVI MONTENEGRO,  
YUVAL PERES<sup>2</sup> AND PRASAD TETALI<sup>3</sup>

*Yonsei University, University of Massachusetts Lowell, Microsoft Research and  
Georgia Institute of Technology*

We show a Birthday Paradox for self-intersections of Markov chains with uniform stationary distribution. As an application, we analyze Pollard’s Rho algorithm for finding the discrete logarithm in a cyclic group  $G$  and find that if the partition in the algorithm is given by a random oracle, then with high probability a collision occurs in  $\Theta(\sqrt{|G|})$  steps. Moreover, for the parallelized distinguished points algorithm on  $J$  processors we find that  $\Theta(\sqrt{|G|}/J)$  steps suffices. These are the first proofs of the correct order bounds which do not assume that every step of the algorithm produces an i.i.d. sample from  $G$ .

**1. Introduction.** The Birthday Paradox states that if  $C\sqrt{N}$  items are sampled uniformly at random with replacement from a set of  $N$  items, then, for large  $C$  with high probability, some items will be chosen twice. This can be interpreted as a statement that with high probability, a Markov chain on the complete graph  $K_N$  with transitions  $P(i, j) = 1/N$  will intersect its past in  $C\sqrt{N}$  steps; we refer to such a self-intersection as a *collision* and say the “*collision time*” is  $O(\sqrt{N})$ . Miller and Venkatesan generalized this in [8] by showing that for a general Markov chain, the collision time is bounded by  $O(\sqrt{N}T_s(1/2))$  where  $T_s(\varepsilon) = \min\{n : \forall u, v \in V, P^n(u, v) \geq (1 - \varepsilon)\pi(v)\}$  measures the time required for the  $n$ -step distribution to assign every state a suitable multiple of its stationary probability. Kim, Montenegro and Tetali [6] further improved the bound on collision time to  $O(\sqrt{NT_s(1/2)})$ . In contrast, while this shows the average path to be quickly self-intersecting, Pak [11] has shown that undirected regular graphs of large degree have a nonintersecting path of length  $N/(32T_s(1/2))$ .

The motivation of [6, 8] was to study the collision time for a Markov chain involved in Pollard’s Rho algorithm for finding the discrete logarithm on a cyclic

---

Received April 2008; revised June 2009.

<sup>1</sup>Supported by Korea Science and Engineering Foundation (KOSEF) Grant funded by the Korean government (MOST) R16-2007-075-01000-0.

<sup>2</sup>Supported in part by NSF Grant DMS-06-05166.

<sup>3</sup>Supported in part by NSF Grants DMS-04-01239, DMS-07-01043.

*AMS 2000 subject classifications.* Primary 60J10; secondary 68Q25, 94A60.

*Key words and phrases.* Pollard’s Rho, discrete logarithm, Markov chain, mixing time.

group  $G$  of prime order  $N = |G| \neq 2$ . For this walk,  $T_s(1/2) = \Omega(\log N)$ , and so the results of [6, 8] are insufficient to show the widely believed  $\Theta(\sqrt{N})$  collision time for this walk. In this paper we improve upon these bounds and show that if a finite ergodic Markov chain has uniform stationary distribution over  $N$  states, then  $O(\sqrt{N})$  steps suffice for a collision to occur as long as the relative-pointwise distance ( $L_\infty$  of the densities of the current and the stationary distribution) drops steadily *early* in the random walk; it turns out that the precise mixing time is largely, although not entirely, unimportant. See Theorem 3.2 for a precise statement. This is then applied to the Rho walk to give the first proof of collision in  $\Theta(\sqrt{N})$  steps, matching Shoup's lower bound [16] on time required for any probabilistic generic algorithm to solve this problem, and to van Oorschot and Wiener's [19] parallel version of the algorithm on  $J$  processors to prove collision in  $\Theta(\sqrt{N}/J)$  steps.

We note here that it is also well known (see, e.g., [1], Section 4.1) that a random walk of length  $L$  contains roughly  $L\lambda$  samples from the stationary measure (of the Markov chain) where  $\lambda$  is the spectral gap of the chain. This yields another estimate on collision time for a Markov chain which is also of a multiplicative nature (namely,  $\sqrt{N}$  times a function of the mixing time) as in [6, 8]. A main point of the present work is to establish sufficient criteria under which the collision time has an *additive* bound:  $C\sqrt{N}$  plus an estimate on the mixing time. While the Rho algorithm provided the main motivation for the present work, we find the more general Birthday Paradox result to be of independent interest, and, as such, expect to have other applications in the future.

A bit of detail about the Pollard Rho algorithm is in order. The classical discrete logarithm problem on a cyclic group deals with computing the exponents, given the generator of the group; more precisely, given a generator  $g$  of a cyclic group  $G$  and an element  $h = g^x$ , one would like to compute  $x$  efficiently. Due to its presumed computational difficulty, the problem figures prominently in various cryptosystems, including the Diffie–Hellman key exchange, El Gamal system and elliptic curve cryptosystems. About 30 years ago, J. M. Pollard suggested algorithms to help solve both factoring large integers [13] and the discrete logarithm problem [14]. While the algorithms are of much interest in computational number theory and cryptography, there has been little work on rigorous analysis. We refer the reader to [8] and other existing literature (e.g., [3, 18]) for further cryptographic and number-theoretical motivation for the discrete logarithm problem.

A standard variant of the classical Pollard Rho algorithm for finding discrete logarithms can be described using a Markov chain on a cyclic group  $G$ . While there has been no rigorous proof of rapid mixing of this Markov chain of order  $O(\log^c |G|)$  until recently, Miller and Venkatesan [8] gave a proof of mixing of order  $O(\log^3 |G|)$  steps and collision time of  $O(\sqrt{|G|} \log^3 |G|)$ , and Kim, Montenegro and Tetali [6] showed mixing of order  $O(\log |G| \log \log |G|)$  and collision time of  $O(\sqrt{|G|} \log |G| \log \log |G|)$ . In this paper we give the first proof

of the correct  $\Theta(\sqrt{|G|})$  collision time. By recent results of Miller and Venkatesan [9] this collision will be nondegenerate and will solve the discrete logarithm problem with probability  $1 - o(1)$  for almost every prime order  $|G|$ , if the start point of the algorithm is chosen at random or if there is no collision in the first  $O(\log |G| \log \log |G|)$  steps.

The paper proceeds as follows. Section 2 contains some preliminaries, primarily an introduction to the Pollard Rho algorithm and a simple multiplicative bound on the collision time in terms of the mixing time. The more general Birthday Paradox for Markov chains with uniform stationary distribution is shown in Section 3. In Section 4 we bound the appropriate constants for the Rho walk and show the optimal collision time. We finish in Section 5 by proving similar results for the distinguished points method of parallelizing the algorithm.

**2. Preliminaries.** Our intent in generalizing the Birthday Paradox was to bound the collision time of the Pollard Rho algorithm for discrete logarithm. As such, we briefly introduce the algorithm here. Throughout the analysis in the following sections, we assume that the size  $N = |G|$  of the cyclic group on which the random walk is performed is odd. Indeed there is a standard reduction (see [15] for a very readable account and [12] for a classical reference) justifying the fact that it suffices to study the discrete logarithm problem on cyclic groups of *prime* order.

Suppose  $g$  is a generator of  $G$ , that is,  $G = \{g^i\}_{i=0}^{N-1}$ . Given  $h \in G$ , the discrete logarithm problem asks us to find  $x$  such that  $g^x = h$ . Pollard suggested an algorithm on  $\mathbb{Z}_N^\times$  based on a random walk and the Birthday Paradox. A common extension of his idea to groups of prime order is to start with a partition of  $G$  into sets  $S_1, S_2, S_3$  of roughly equal sizes, and define an iterating function  $F : G \rightarrow G$  by  $F(y) = gy$  if  $y \in S_1$ ,  $F(y) = hy = g^x y$  if  $y \in S_2$  and  $F(y) = y^2$  if  $y \in S_3$ . Then consider the walk  $y_{i+1} = F(y_i)$ . If this walk passes through the same state twice, say  $g^{a+xb} = g^{\alpha+x\beta}$ , then  $g^{a-\alpha} = g^{x(\beta-b)}$ , and so  $a - \alpha \equiv x(\beta - b) \pmod N$  and  $x \equiv (a - \alpha)(\beta - b)^{-1} \pmod N$  which determines  $x$  as long as  $(\beta - b, N) = 1$  (the nondegenerate case). Hence, if we define a *collision* to be the event that the walk passes over the same group element twice, then the first time there is a collision it might be possible to determine the discrete logarithm.

To estimate the running time until a collision, one heuristic is to treat  $F$  as if it outputs uniformly random group elements. By the Birthday Paradox, if  $O(\sqrt{|G|})$  group elements are chosen uniformly at random, then there is a high probability that two of these are the same. Teske [17] has given experimental evidence that the time until a collision is slower than what would be expected by an independent uniform random process. We analyze instead the actual Markov chain in which it is assumed only that each  $y \in G$  is assigned independently and at random to a partition  $S_1, S_2$  or  $S_3$ . In this case, although the iterating function  $F$  described earlier is deterministic, because the partition of  $G$  was randomly chosen, then the walk is equivalent to a Markov chain (i.e., a random walk), at least until the walk visits a previously visited state and a collision occurs. The problem is then one of

considering a walk on the exponent of  $g$ , that is, a walk  $P$  on the cycle  $\mathbb{Z}_N$  with transitions  $P(u, u + 1) = P(u, u + x) = P(u, 2u) = 1/3$ .

REMARK 2.1. By assuming each  $y \in G$  is assigned independently and at random to a partition, we have eliminated one of the key features of the Pollard Rho algorithm, space efficiency. However, if the partitions are given by a hash function,  $f : (G, N) \rightarrow \{1, 2, 3\}$ , which is sufficiently pseudo-random, then we might expect behavior similar to the model with random partitions.

REMARK 2.2. While we are studying the time until a collision occurs, there is no guarantee that the first collision will be nondegenerate. If the first collision is degenerate then so will be all collisions as the algorithm becomes deterministic after the first collision.

As mentioned in the introduction, we first recall a simple multiplicative bound on collision time from [6]. The following proposition relates  $T_s(1/2)$  to the time until a collision occurs for any Markov chain  $P$  with uniform distribution on  $G$  as the stationary distribution.

PROPOSITION 2.3. *With the above definitions, a collision occurs after*

$$T_s(1/2) + 2\sqrt{2c|G|T_s(1/2)}$$

*steps with probability at least  $1 - e^{-c}$ , for any  $c > 0$ .*

PROOF. Let  $S$  denote the first  $\lceil \sqrt{2c|G|T_s(1/2)} \rceil$  states visited by the walk. If two of these states are the same then a collision has occurred, so assume all states are distinct. Even if we only check for collisions every  $T_s(1/2)$  steps, the chance that no collision occurs in the next  $tT_s(1/2)$  steps (so consider  $t$  semi-random states) is then at most

$$\left(1 - \frac{1}{2} \frac{|S|}{|G|}\right)^t \leq \left(1 - \sqrt{\frac{cT_s(1/2)}{2|G|}}\right)^t \leq \exp\left(-t \sqrt{\frac{cT_s(1/2)}{2|G|}}\right).$$

When  $t = \lceil \sqrt{\frac{2c|G|}{T_s(1/2)}} \rceil$ , this is at most  $e^{-c}$ , as desired, and so at most

$$\lceil \sqrt{2c|G|T_s(1/2)} \rceil - 1 + \left\lceil \sqrt{\frac{2c|G|}{T_s(1/2)}} \right\rceil T_s(1/2)$$

steps are required for a collision to occur with probability at least  $1 - e^{-c}$ .  $\square$

Obtaining a more refined additive bound on collision time will be the focus of the next section. While the proof can be seen as another application of the well-known second moment method, it turns out that bounding the second moment of the number of collisions *before* the mixing time is somewhat subtle. To handle this, we use an idea from [7], who in turn credit their approach to [5].

**3. Collision time.** Consider a finite ergodic Markov chain  $P$  with uniform stationary distribution  $U$  (i.e., doubly stochastic), state space  $\Omega$  of cardinality  $N = |\Omega|$ , and let  $X_0, X_1, \dots$  denote a particular instance of the walk. In this section we determine the number of steps of the walk required to have a high probability that a “collision” has occurred, that is, a self-intersection  $X_i = X_j$  for some  $i \neq j$ .

A key notion when studying Markov chains is the mixing time, or the time required until the probability of being at each state is suitably close to its stationary probability.

**DEFINITION 3.1.** The *mixing time*  $\tau(\varepsilon)$  of a Markov chain  $P$  with stationary distribution  $U$  is given by

$$\tau(\varepsilon) = \min\{T : \forall u, v \in \Omega, (1 - \varepsilon)U(v) \leq P^T(u, v) \leq (1 + \varepsilon)U(v)\}.$$

Now some notation.

Fix some  $T \geq 0$  and integer  $\beta > 0$ . Let the indicator function  $\mathbf{1}_{\{X_i=X_j\}}$  equal one if  $X_i = X_j$ , and zero otherwise. Define

$$S = \sum_{i=0}^{\beta\sqrt{N}} \sum_{j=i+2T}^{\beta\sqrt{N}+2T} \mathbf{1}_{\{X_i=X_j\}}$$

to be the number of times the walk intersects itself in  $\beta\sqrt{N} + 2T$  steps where  $i$  and  $j$  are at least  $2T$  steps apart. Also, for  $u, v \in \Omega$ , let

$$G_T(u, v) = \sum_{i=0}^T P^i(u, v)$$

be the expected number of times a walk beginning at  $u$  hits state  $v$  in  $T$  steps. Finally, let

$$A_T = \max_u \sum_v G_T^2(u, v) \quad \text{and} \quad A_T^* = \max_u \sum_v G_T^2(v, u).$$

To see the connection between these and the collision time, observe that

$$\begin{aligned} \sum_v G_T^2(u, v) &= \sum_v \left( \sum_{i=0}^T \sum_{j=0}^T P^i(u, v) P^j(u, v) \right) \\ &= \sum_{i=0}^T \sum_{j=0}^T \sum_v P^i(u, v) P^j(u, v) \\ &= \sum_{i=0}^T \sum_{j=0}^T \Pr(X_i = Y_j) \\ &= \sum_{i=0}^T \sum_{j=0}^T E(\mathbf{1}_{\{X_i=Y_j\}}) = E \sum_{i,j=0}^T \mathbf{1}_{\{X_i=Y_j\}}, \end{aligned}$$

where  $\{X_i\}, \{Y_j\}$  are i.i.d. copies of the chain, both having started at  $u$  at time 0, and  $E$  denotes expectation. Hence  $A_T$  is the maximal expected number of collisions of two  $T$ -step i.i.d. walks of  $P$  starting at the same state  $u$ . Likewise,  $A_T^*$  is the same for the reversal  $P^*$  where  $P^*(u, v) = P(v, u)$  (recall that the stationary distribution was assumed to be uniform).

The main result of this section is the following.

**THEOREM 3.2** (Birthday Paradox for Markov chains). *Consider a finite ergodic Markov chain with uniform stationary distribution on a state space of size  $N$ . Let  $T$  be such that  $\frac{m}{N} \leq P^T(u, v) \leq \frac{M}{N}$  for some  $m \leq 1 \leq M$  and every pair of states  $u, v$ . After*

$$4c \left(\frac{M}{m}\right)^2 \left(\sqrt{\frac{2N}{M} \max\{A_T, A_T^*\}} + T\right)$$

*steps a collision occurs with a probability of at least  $1 - e^{-c}$ , for any  $c \geq 0$ .*

At the end of this section a slight strengthening of Theorem 3.2 is shown at the cost of a somewhat less intuitive bound.

In Example 3.5, near the end of this section, we present an example to illustrate the need for the pre-mixing term  $A_T$  in Theorem 3.2. In contrast, very recently Nazarov and Peres [10] proved a general bound for the birthday problem on any reversible Markov chain on  $N$  states: Suppose that the ratio of the stationary measures of any two states is at most  $A$ . Then they show that for any starting state, the expected time until the chain visits a previously visited state is at most  $C\sqrt{N} + \log(A)$  for some universal constant  $C$ . In particular, this implies an expected collision time of  $O(\sqrt{N})$  for the simple random walk on an undirected graph on  $N$  vertices, and so the pre-mixing term is not necessary when considering reversible walks.

Observe that if  $A_T, A_T^*, m, M = \Theta(1)$  and  $T = O(\sqrt{N})$ , then the collision time is  $O(\sqrt{N})$  as in the standard Birthday Paradox. By Lemma 3.3, for this to occur it suffices that  $P^T$  be sufficiently close to uniform after  $T = o(\sqrt{N})$  steps and that  $P^j(u, v) = o(T^{-2}) + d^j$  for all  $u, v$ , for  $j \leq T$  and some  $d < 1$ . More generally, to upper bound  $A_T$  and  $A_T^*$  it suffices to show that the maximum probability of being at a vertex decreases quickly.

**LEMMA 3.3.** *If a finite ergodic Markov chain has uniform stationary distribution, then*

$$A_T, A_T^* \leq 2 \sum_{j=0}^T (j+1) \max_{u,v} P^j(u, v).$$

PROOF. If  $u$  is such that equality occurs in the definition of  $A_T$ , then

$$\begin{aligned} A_T &= \sum_v G_T^2(u, v) = \sum_{i=0}^T \sum_{j=0}^T \sum_v P^i(u, v) P^j(u, v) \\ &\leq 2 \sum_{j=0}^T \sum_{i=0}^j \max_y P^j(u, y) \sum_v P^i(u, v) \\ &\leq 2 \sum_{j=0}^T (j+1) \max_y P^j(u, y). \end{aligned}$$

The quantity  $A_T^*$  plays the role of  $A_T$  for the reversed chain, and so the same bound holds for  $A_T^*$  but with  $\max_{u,v} (P^*)^j(u, v) = \max_{u,v} P^j(v, u) = \max_{u,v} P^j(u, v)$ .  $\square$

In particular, suppose  $P^j(u, v) \leq c + d^j$  for every  $u, v \in \Omega$  and some  $c, d \in [0, 1)$ . The sum

$$\begin{aligned} \sum_{j=0}^T (j+1)(c + d^j) &= c \frac{(T+1)(T+2)}{2} + \frac{1 - d^{T+1} - (T+1)d^{T+1}(1-d)}{(1-d)^2} \\ &\leq (1 + o(1)) \frac{cT^2}{2} + \frac{1}{(1-d)^2}, \end{aligned}$$

and so if  $P^j(u, v) \leq o(T^{-2}) + d^j$  for every  $u, v \in \Omega$ , then  $A_T, A_T^* = \frac{2+o(1)}{(1-d)^2}$ .

The proof of Theorem 3.2 relies largely on the following inequality which shows that the expected number of self-intersections is large with low variance:

LEMMA 3.4. *Under the conditions of Theorem 3.2,*

$$\begin{aligned} E[S] &\geq \frac{m}{N} \left( \frac{\beta\sqrt{N} + 2}{2} \right), \\ E[S^2] &\leq \frac{M^2}{N^2} \left( \frac{\beta\sqrt{N} + 2}{2} \right)^2 \left( 1 + \frac{8 \max\{A_T, A_T^*\}}{M\beta^2} \right). \end{aligned}$$

PROOF OF THEOREM 3.2. Recall the standard second moment bound: using Cauchy–Schwarz, we have that

$$E[S] = E[S\mathbf{1}_{\{S>0\}}] \leq E[S^2]^{1/2} E[\mathbf{1}_{\{S>0\}}]^{1/2},$$

and hence  $\Pr[S > 0] \geq E[S]^2/E[S^2]$ . If  $\beta = 2\sqrt{2 \max\{A_T, A_T^*\}/M}$ , then by Lemma 3.4,

$$\Pr[S > 0] \geq \frac{m^2/M^2}{1 + (8 \max\{A_T, A_T^*\})/(M\beta^2)} \geq \frac{m^2}{2M^2}$$

independent of the starting point. If no collision occurs in  $\beta\sqrt{N} + 2T$  steps then  $S = 0$  as well, and so  $\Pr[\text{no collision}] \leq \Pr[S = 0] \leq 1 - m^2/2M^2$ . Hence, in  $k(\beta\sqrt{N} + 2T)$  steps,

$$(3.1) \quad \Pr[\text{no collision}] \leq (1 - m^2/2M^2)^k \leq e^{-km^2/2M^2}.$$

Taking  $k = 2cM^2/m^2$  completes the proof.  $\square$

**PROOF OF LEMMA 3.4.** We will repeatedly use the relation that there are  $\binom{\beta\sqrt{N}+2}{2}$  choices for  $i, j$  appearing in the summation for  $S$ , that is,  $0 \leq i$  and  $i + 2T \leq j \leq \beta\sqrt{N} + 2T$ .

Now to the proof. The expectation  $E[S]$  satisfies

$$(3.2) \quad \begin{aligned} E[S] &= E \sum_{i=0}^{\beta\sqrt{N}} \sum_{j=i+2T}^{\beta\sqrt{N}+2T} \mathbf{1}_{\{X_i=X_j\}} \\ &= \sum_{i=0}^{\beta\sqrt{N}} \sum_{j=i+2T}^{\beta\sqrt{N}+2T} E[\mathbf{1}_{\{X_i=X_j\}}] \geq \binom{\beta\sqrt{N}+2}{2} \frac{m}{N} \end{aligned}$$

because if  $j \geq i + T$ , then

$$(3.3) \quad \Pr(X_j = X_i) = \sum_u \Pr(X_i = u) P^{j-i}(u, u) \geq \sum_u \Pr(X_i = u) \frac{m}{N} = \frac{m}{N}.$$

Similarly,  $\Pr(X_j = X_i) \leq \frac{M}{N}$  when  $j \geq i + T$ .

Now for  $E[S^2]$ . Note that

$$(3.4) \quad \begin{aligned} E[S^2] &= E \left( \sum_{i=0}^{\beta\sqrt{N}} \sum_{j=i+2T}^{\beta\sqrt{N}+2T} \mathbf{1}_{\{X_i=X_j\}} \right) \left( \sum_{k=0}^{\beta\sqrt{N}} \sum_{l=k+2T}^{\beta\sqrt{N}+2T} \mathbf{1}_{\{X_k=X_l\}} \right) \\ &= \sum_{i=0}^{\beta\sqrt{N}} \sum_{k=0}^{\beta\sqrt{N}} \sum_{j=i+2T}^{\beta\sqrt{N}+2T} \sum_{l=k+2T}^{\beta\sqrt{N}+2T} \text{Prob}(X_i = X_j, X_k = X_l). \end{aligned}$$

To evaluate this quadruple sum we break it into 3 cases.

**CASE 1.** Suppose  $|j - l| \geq T$ . Without loss, assume  $l \geq j$  so that, in particular,  $l \geq \max\{i, j, k\} + T$ . Then

$$(3.4) \quad \begin{aligned} &\text{Prob}(X_i = X_j, X_k = X_l) \\ &= \text{Prob}(X_i = X_j) \text{Prob}(X_l = X_k \mid X_i = X_j) \\ &\leq \text{Prob}(X_i = X_j) \max_{u,v} \text{Prob}(X_l = v \mid X_{\max\{i,j,k\}} = u) \\ &\leq \text{Prob}(X_i = X_j) \frac{M}{N} \leq \left(\frac{M}{N}\right)^2. \end{aligned}$$

The first inequality holds because  $\{X_t\}$  is a Markov chain and so given  $X_i, X_j, X_k$  the walk at any time  $t \geq \max\{i, j, k\}$  depends only on the state  $X_{\max\{i, j, k\}}$ .

CASE 2. Suppose  $|i - k| \geq T$  and  $|j - l| < T$ . Without loss, assume  $i \leq k$ . If  $j \leq l$ , then

$$\begin{aligned}
 & \text{Prob}(X_i = X_j, X_k = X_l) \\
 (3.5) \quad &= \sum_{u,v} \text{Prob}(X_i = u) P^{k-i}(u, v) P^{j-k}(v, u) P^{l-j}(u, v) \\
 &\leq \sum_u \text{Prob}(X_i = u) \frac{M}{N} \frac{M}{N} \sum_v P^{l-j}(u, v) = \left(\frac{M}{N}\right)^2
 \end{aligned}$$

because  $k \geq i + T, j \geq k + T$  and  $\sum_v P^t(u, v) = 1$  for any  $t$  because  $P$  and hence also  $P^t$  are stochastic matrices. If, instead,  $l < j$  then essentially the same argument works, but with  $\sum_v P^t(v, u) = 1$  because  $P$  and hence also  $P^t$  are doubly-stochastic.

CASE 3. Finally, consider those terms with  $|j - l| < T$  and  $|i - k| < T$ . Without loss, assume  $i \leq k$ . If  $l \leq j$ , then

$$\begin{aligned}
 & \text{Prob}(X_i = X_j, X_k = X_l) \\
 (3.6) \quad &= \sum_{u,v} \text{Prob}(X_i = u) P^{k-i}(u, v) P^{l-k}(v, v) P^{j-l}(v, u) \\
 &\leq \sum_u \text{Prob}(X_i = u) \sum_v P^{k-i}(u, v) \frac{M}{N} P^{j-l}(v, u).
 \end{aligned}$$

The sum over elements with  $i \leq k < i + T$  and  $l \leq j < l + T$  is upper bounded as follows:

$$\begin{aligned}
 & \sum_{i=0}^{\beta\sqrt{N}} \sum_{k=i}^{i+T} \sum_{l=k+2T}^{\beta\sqrt{N}+2T} \sum_{j=l}^{l+T} \text{Prob}(X_i = X_j, X_k = X_l) \\
 &\leq \frac{M}{N} \sum_{i=0}^{\beta\sqrt{N}} \sum_{l=i+2T}^{\beta\sqrt{N}+2T} \max_u \sum_v \sum_{k \in [i, i+T)} P^{k-i}(u, v) \sum_{j \in [l, l+T)} P^{j-l}(v, u) \\
 (3.7) \quad &\leq \frac{M}{N} \sum_{i=0}^{\beta\sqrt{N}} \sum_{l=i+2T}^{\beta\sqrt{N}+2T} \max_u \sum_v G_T(u, v) G_T(v, u) \\
 &\leq \frac{M}{N} \sum_{i=0}^{\beta\sqrt{N}} \sum_{l=i+2T}^{\beta\sqrt{N}+2T} \max_u \sqrt{\sum_v G_T^2(u, v) \sum_v G_T^2(v, u)} \\
 &\leq \frac{M}{N} \left(\frac{\beta\sqrt{N} + 2}{2}\right) \sqrt{A_T A_T^*}.
 \end{aligned}$$

The case when  $j < l$  gives the same bound but with the observation that  $j \geq k + T$  and with  $A_T$  instead of  $\sqrt{A_T A_T^*}$ .

Putting together these various cases we get that

$$E[S^2] \leq \binom{\beta\sqrt{N} + 2}{2} \left(\frac{M}{N}\right)^2 + 2 \binom{\beta\sqrt{N} + 2}{2} \frac{M}{N} A_T + 2 \binom{\beta\sqrt{N} + 2}{2} \frac{M}{N} \sqrt{A_T A_T^*}.$$

The  $\binom{\beta\sqrt{N} + 2}{2}$  term is the total number of values of  $i, j, k, l$  appearing in the sum for  $E[S^2]$ , and hence also an upper bound on the number of values in Cases 1 and 2. Along with the relation  $\binom{\beta\sqrt{N} + 2}{2} \geq \frac{\beta^2 N}{2}$  this simplifies to complete the proof.  $\square$

As promised earlier, we now present an example that illustrates the need for the pre-mixing term  $A_T$  in Theorem 3.2.

**EXAMPLE 3.5.** Consider the random walk on  $\mathbb{Z}_N$  which transitions from  $u \rightarrow u + 1$  with probability  $1 - 1/\sqrt{N}$ , and with probability  $1/\sqrt{N}$  transitions  $u \rightarrow v$  for a uniformly random choice of  $v$ .

Heuristically the walk proceeds as  $u \rightarrow u + 1$  for  $\approx \sqrt{N}$  steps, then randomizes, and then proceeds as  $u \rightarrow u + 1$  for another  $\sqrt{N}$  steps. This effectively splits the state space into  $\sqrt{N}$  blocks of size about  $\sqrt{N}$  each, so by the standard Birthday Paradox it should require about  $\sqrt{N^{1/2}}$  of these randomizations before a collision will occur, in short, about  $N^{3/4}$  steps in total.

To see the need for the pre-mixing term, observe that  $T_s \approx \sqrt{N} \log 2$  while if  $T = T_\infty \approx \sqrt{N} \log(2(N - 1))$ , then we may take  $m = 1/2$  and  $M = 3/2$  in Theorem 3.2. So, whether  $T_s$  or  $T_\infty$  are considered, it will be insufficient to take  $O(T + \sqrt{N})$  steps. However, the number  $A_T$  of collisions between two independent copies of this walk is about  $\sqrt{N}$  since once a randomization step occurs then the two independent walks are unlikely to collide anytime soon. Our collision time bound says that  $O(N^{3/4})$  steps will suffice which is the correct bound.

A proper analysis shows that  $\frac{1-o(1)}{\sqrt{2}} N^{3/4}$  steps are necessary to have a collision with a probability of  $1/2$ . Conversely, when  $T = \sqrt{N} \log^2 N$  then  $m = 1 - o(1)$ ,  $M = 1 + o(1)$  and  $A_T, A_T^* \leq \frac{1+o(1)}{2} \sqrt{N}$ , so by equation (3.1),  $(2 + o(1))N^{3/4}$  steps are sufficient to have a collision with a probability of at least  $1/2$ . Our upper bound is thus off by at most a factor of  $2\sqrt{2} \approx 2.8$ .

We finish the section with a slight sharpening of Theorem 3.2. This will be used to improve the lead constant in our upcoming bound on collision time for the Pollard Rho walk.

**THEOREM 3.6** (Improved Birthday Paradox). *Consider a finite ergodic Markov chain with uniform stationary distribution on a state space of size  $N$ . Let  $T$  be such that  $\frac{m}{N} \leq P^T(u, v) \leq \frac{M}{N}$  for some  $m \leq 1 \leq M$  and every pair of states  $u, v$ . After*

$$2c \left( \sqrt{\left( 1 + \sum_{j=1}^{2T} 3j \max_{u,v} P^j(u, v) \right) \frac{N}{M} + T} \right)$$

*steps a collision occurs with a probability of at least  $1 - (1 - \frac{m^2}{2M^2})^c$ , independent of the starting state.*

**PROOF.** We give only the steps that differ from before.

First, in equation (3.7), note that the triple sum after  $\max_u$  can be re-written as

$$\sum_{\alpha \in [0, T)} \sum_{\beta \in [0, T)} \sum_v P^\alpha(u, v) P^\beta(v, u) \leq \sum_{\gamma=0}^{2(T-1)} (\gamma + 1) P^\gamma(u, u).$$

The original quadruple sum then reduces to

$$\frac{M}{N} \left( \frac{\beta\sqrt{N} + 2}{2} \right) \max_u \sum_{\gamma=0}^{2(T-1)} (\gamma + 1) P^\gamma(u, u).$$

For the case when  $i < k$  and  $j < l$  proceed similarly, then reduce, as in Lemma 3.3, to obtain the upper bound,

$$\begin{aligned} & \frac{M}{N} \left( \frac{\beta\sqrt{N} + 2}{2} \right) \sum_{\alpha=1}^{T-1} \sum_{\beta=1}^{T-1} \sum_v P^\alpha(u, v) P^\beta(u, v) \\ & \leq \frac{M}{N} \left( \frac{\beta\sqrt{N} + 2}{2} \right) \sum_{\gamma=1}^{T-1} (2\gamma - 1) \max_v P^\gamma(u, v). \end{aligned}$$

Adding these two expressions gives an expression of at most

$$\frac{M}{N} \left( \frac{\beta\sqrt{N} + 2}{2} \right) \left( 1 + \sum_{\gamma=1}^{2T} 3\gamma \max_v P^\gamma(u, v) \right).$$

The remaining two cases will add to the same bound, so effectively this substitutes the expression  $2(1 + \max_u \sum_{\gamma=1}^{2T} 3\gamma \max_v P^\gamma(u, v))$  in place of a  $4 \max\{A_T, A_T^*\}$  in the original theorem.  $\square$

To simplify, note that if  $\max_{u,v} P^j(u, v) \leq c + d^j$  for  $c, d \in [0, 1)$ , then

$$(3.8) \quad \sum_{j=1}^{2T} 3j(c + d^j) = 3cT(2T + 1) + 3d \frac{1 - d^{2T} - 2Td^{2T}(1 - d)}{(1 - d)^2} \leq (1 + o(1))6cT^2 + \frac{3d}{(1 - d)^2}.$$

**4. Convergence of the Rho walk.** Let us now turn our attention to the Pollard Rho walk for discrete logarithm. To apply the collision time result we will first show that  $\max_{u,v \in \mathbb{Z}_N} P^s(u, v)$  decreases quickly in  $s$  so that Lemma 3.3 may be used. We then find  $T$  such that  $P^T(u, v) \approx 1/N$  for every  $u, v \in \mathbb{Z}_N$ . However, instead of studying the Rho walk directly, most of the work will instead involve a “block walk” in which only a certain subset of the states visited by the Rho walk are considered.

**DEFINITION 4.1.** Let us refer to the three types of moves that the Pollard Rho random walk makes, namely  $(u, u + 1)$ ,  $(u, u + x)$  and  $(u, 2u)$  as moves of Type 1, Type 2 and Type 3, respectively. In general, let the random walk be denoted by  $Y_0, Y_1, Y_2, \dots$  with  $Y_t$  indicating the position of the walk (modulo  $N$ ) at time  $t \geq 0$ . Let  $T_1$  be the first time that the walk makes a move of Type 3. Let  $b_1 = Y_{T_1-1} - Y_{T_0}$  (i.e., the ground covered, modulo  $N$ , only using consecutive moves of Types 1 and 2). More generally, let  $T_i$  be the first time, since  $T_{i-1}$ , that a move of Type 3 happens, and set  $b_i = Y_{T_i-1} - Y_{T_{i-1}}$ . Then the *block walk*  $\mathbf{B}$  is the walk  $X_s = Y_{T_s} = 2^s Y_{T_0} + 2 \sum_{i=1}^s 2^{s-i} b_i$ .

By combining our Birthday Paradox for Markov chains with several lemmas to be shown in this section, we obtain the main result of the paper:

**THEOREM 4.2.** *For every choice of starting state, the expected number of steps required for the Pollard Rho algorithm for discrete logarithm on a group  $G$  to have a collision is at most*

$$(1 + o(1))12\sqrt{19}\sqrt{|G|} < (1 + o(1))52.5\sqrt{|G|}.$$

In order to prove this it is necessary to show that  $\mathbf{B}^s(u, v)$  decreases quickly for the block walk.

**LEMMA 4.3.** *If  $s \leq \lfloor \log_2 N \rfloor$ , then for every  $u, v \in \mathbb{Z}_N$  the block walk satisfies*

$$\mathbf{B}^s(u, v) \leq (2/3)^s.$$

*If  $s > \lfloor \log_2 N \rfloor$  then  $\mathbf{B}^s(u, v) \leq \frac{3/2}{N^{\log_2 3 - 1}} \leq \frac{3/2}{\sqrt{N}}$ .*

A bound on the asymptotic rate of convergence is also required:

**THEOREM 4.4.** *If  $s \geq \lceil m \log \frac{2(m-1)}{\varepsilon} \rceil$  where  $m = \lceil \log_2 N \rceil$ , then for every  $u, v \in \mathbb{Z}_N$  the block walk satisfies*

$$\frac{1 - \varepsilon}{N} \leq B^{2s}(u, v) \leq \frac{1 + \varepsilon}{N}.$$

This is all that is needed to prove the main result:

**PROOF OF THEOREM 4.2.** The proof will use Theorem 3.6 because this gives a somewhat sharper bound. Alternatively, Theorem 3.2 and Lemma 3.3 can be applied nearly identically to get the slightly weaker  $(1 + o(1))72\sqrt{|G|}$ .

First consider steps of the block walk. Lemma 4.3 implies that  $B^s(u, v) \leq \frac{3/2}{\sqrt{N}} + (\frac{2}{3})^s$ , for  $s \geq 0$  and for all  $u, v$ . Hence, by equation (3.8), if  $T = o(\sqrt[4]{N})$ , then  $1 + \sum_{j=1}^{2T} 3jB^j(u, v) \leq 19 + o(1)$ . By Theorem 4.4,  $M \leq 1 + \varepsilon$  and  $m \geq 1 - \varepsilon$  after  $2(\log_2 N)(\log \log N + \log \frac{3}{\varepsilon})$  steps. Hence, if  $\varepsilon = 1/N^2$ , then  $T = (4 + o(1))(\log_2 N)^2 = o(\sqrt[4]{N})$ , and  $m = 1 - o(1/N)$ , and  $M = 1 + o(1/N)$ . Plugging this into Theorem 3.6, a collision fails to occur in

$$k \left( 2 \sqrt{\left( 1 + \sum_{j=1}^{2T} 3j \max_{u,v} B^j(u, v) \right) \frac{N}{M} + 2T} \right) = (1 + o(1))2\sqrt{19}k\sqrt{N}$$

steps with a probability of, at most,  $(1 - \delta)^k$  where  $\delta = m^2/2M^2 = (1 - o(1))/2$ .

Now return to the Rho walk. Recall that  $T_i$  denotes the number of Rho steps required for  $i$  block steps. The difference  $T_{i+1} - T_i$  is an i.i.d. random variable with the same distribution as  $T_1 - T_0$ . Hence, if  $i \geq j$ , then  $E[T_i - T_j] = (i - j)E[T_1 - T_0] = 3(i - j)$ . In particular, if we let  $r = (1 + o(1))2\sqrt{19N}$ , let  $R$  denote the number of Rho steps before a collision, and let  $B$  denote the number of block steps before a collision, then

$$\begin{aligned} E[R] &\leq \sum_{k=0}^{\infty} \Pr[B > kr] E[T_{(k+1)r} - T_{kr} \mid B > kr] \\ &= \sum_{k=0}^{\infty} \Pr[B > kr] E[T_{(k+1)r} - T_{kr}] \\ &\leq \sum_{k=0}^{\infty} \left( \frac{1 + o(1)}{2} \right)^k 3r = (1 + o(1))12\sqrt{19}\sqrt{N}. \end{aligned}$$

□

**PROOF OF LEMMA 4.3.** We start with a weaker, but somewhat more intuitive, proof of a bound on  $B^s(u, v)$  and then improve it to obtain the result of the lemma. The key idea here will be to separate out a portion of the Markov chain which is tree-like with some large depth  $L$ , namely the moves induced solely by  $b_i = 0$  and

$b_i = 1$  moves. Because of the high depth of the tree, the walk spreads out for the first  $L$  steps, and hence the probability of being at a vertex also decreases quickly.

Let  $S = \{i \in [1, \dots, s] : b_i \in \{0, 1\}\}$  and  $z = \sum_{i \notin S} 2^{s-i} b_i$  be random variables whose values are determined by the first  $T_s$  steps of the random walk. Then  $Y_{T_s} = 2^s Y_{T_0} + 2z + 2 \sum_{i \in S} 2^{s-i} b_i$ . Hence, choosing  $Y_{T_0} = u$ ,  $Y_{T_s} = v$ , we may write

$$\begin{aligned} B^s(u, v) &= \sum_S \text{Prob}(S) \sum_{z \in \mathbb{Z}_N} \text{Prob}(z \mid S) \text{Prob}\left(\sum_{i \in S} 2^{s-i} b_i = \frac{v}{2} - 2^{s-1}u - z \mid z, S\right) \\ &\leq \sum_S \text{Prob}(S) \max_{w \in \mathbb{Z}_N} \text{Prob}\left(\sum_{i \in S} 2^{s-i} b_i = w \mid S\right), \end{aligned}$$

and so for a fixed choice of  $S$ , we can ignore what happens on  $S^c$ .

Each  $w \in [0, \dots, N - 1]$  has a unique binary expansion, and so if  $s \leq \lfloor \log_2 N \rfloor$ , then modulo  $N$  each  $w$  can still be written in, at most, one way as an  $s$  bit string. For the block walk,  $\text{Prob}(b_i = 0) \geq 1/3$  and  $\text{Prob}(b_i = 1) \geq 1/9$ , and so  $\max\{\text{Prob}(b_i = 0 \mid i \in S), \text{Prob}(b_i = 1 \mid i \in S)\} \leq \frac{8}{9}$ . It follows that

$$(4.1) \quad \max_{w \in \mathbb{Z}_N} \text{Prob}\left(\sum_{i \in S} 2^{s-i} b_i = w \mid S\right) \leq (8/9)^{|S|}$$

using independence of the  $b_i$ s. Hence,

$$\begin{aligned} B^s(u, v) &\leq \sum_S \text{Prob}(S) (8/9)^{|S|} = \sum_{r=0}^s \text{Prob}(|S| = r) (8/9)^r \\ &\leq \sum_{r=0}^s \binom{s}{r} \left(\frac{4}{9}\right)^r \left(1 - \frac{4}{9}\right)^{s-r} \left(\frac{8}{9}\right)^r = \left(\frac{4}{9} + \frac{5}{9}\right)^s = \left(\frac{77}{81}\right)^s. \end{aligned}$$

The second inequality was because  $(8/9)^{|S|}$  is decreasing in  $|S|$  and so underestimating  $|S|$  by assuming  $\text{Prob}(i \in S) = 4/9$  will only increase the upper bound on  $B^s(u, v)$ .

In order to improve on this, we will shortly redefine  $S$  (namely, events  $\{i \in S\}, \{i \notin S\}$ ) and auxiliary variables  $c_i$ , using the steps of the Rho walk. Also note that the block walk is induced by a Rho walk, so we may assume that the  $b_i$  were constructed by a series of steps of the Rho walk. With probability  $1/4$  set  $i \in S$  and  $c_i = 0$ , otherwise if the first step is of Type 1, then set  $i \in S$  and  $c_i = 1$  while if the first step is of Type 3 then put  $i \notin S$  and  $c_i = 0$ , and finally if the first step is of Type 2, then again repeat the above decision making process, using the subsequent steps of the walk. Note that the above construction can be summarized as consisting of one of four equally likely outcomes (at each time) where the last three outcomes depend on the type of the step that the Rho walk takes; indeed, each

of these three outcomes happens with probability  $\frac{3}{4} \times \frac{1}{3} = 1/4$ ; finally, a Type 2 step forces us to reiterate the four-way decision-making process.

In summary,  $\Pr(i \in S) = \sum_{l=0}^{\infty} (1/4)^l (1/2) = 2/3$ . Also observe that  $\Pr(c_i = 0 \mid i \in S) = \Pr(c_i = 1 \mid i \in S)$  and that  $\Pr(b_i - c_i = x \mid i \in S, c_i = 0) = \Pr(b_i - c_i = x \mid i \in S, c_i = 1)$ . Hence the steps done earlier (leading to the weaker bound) carry through with  $z = \sum_i 2^{s-i} (b_i - c_i)$  and with  $\sum_{i \in S} 2^{s-i} b_i$  replaced by  $\sum_{i \in S} 2^{s-i} c_i$ . In (4.1) replace  $(8/9)^{|S|}$  by  $(1/2)^{|S|}$ , and in showing the final upper bound on  $B^s(u, v)$ , replace  $4/9$  by  $2/3$ . This leads to the bound  $B^s(u, v) \leq (2/3)^s$ .

Finally, when  $s > \lceil \log_2 N \rceil$ , simply apply the preceding argument to  $S' = S \cap [1, \dots, \lfloor \log_2 N \rfloor]$ . Alternately, note that when  $s \geq \lfloor \log_2 N \rfloor$ , then

$$B^s(u, v) \leq \max_w B^{\lfloor \log_2 N \rfloor}(u, w) \leq (2/3)^{\log_2 N - 1}$$

for every doubly-stochastic Markov chain  $B$ .  $\square$

In [6, 8] sufficiently strong bounds on the asymptotics of  $B^{2^s}(u, v)$  are shown in several ways, including the use of characters and quadratic forms, canonical paths and Fourier analysis. We give here the Fourier approach, as it establishes the sharpest mixing bounds. To bound mixing time of the block walk, it suffices to show that for large enough  $s$ , the distribution  $\nu_s$  of

$$Z_s = 2^{s-1}b_1 + 2^{s-2}b_2 + \dots + b_s$$

is close to the uniform distribution  $U = 1/N$  because then the distribution of  $X_s = 2^s Y_{T_0} + 2Z_s$  will be close to uniform as well. More precisely, convergence in chi-square distance will be shown that

LEMMA 4.5. *If  $\nu_s(j) = \Pr[Z_s = j]$ ,  $\xi = 1 - \frac{4-\sqrt{10}}{9}$ , and  $m$  satisfies  $2^{m-1} < N < 2^m$ , then*

$$N \sum_{j=0}^{N-1} (\nu_s(j) - U(j))^2 \leq 2((1 + \xi^{2\lfloor s/m \rfloor})^{m-1} - 1).$$

PROOF OF THEOREM 4.4. By Cauchy–Schwarz,

$$\begin{aligned} & \left| \frac{B^{2^s}(u, v) - U(v)}{U(v)} \right|^2 \\ &= \left| \frac{\sum_w (B^s(u, w) - U(w))(B^s(w, v) - U(v))}{U(v)} \right|^2 \\ (4.2) \quad &= \left| \sum_w U(w) \left( \frac{B^s(u, w)}{U(w)} - 1 \right) \left( \frac{B^{*s}(v, w)}{U(w)} - 1 \right) \right|^2 \\ &\leq \sum_w U(w) \left| \frac{B^s(u, w)}{U(w)} - 1 \right|^2 \sum_x U(x) \left| \frac{B^{*s}(v, x)}{U(x)} - 1 \right|^2. \end{aligned}$$

Lemma 4.5 bounds the first sum of (4.2). The second sum is the same quantity but for the time-reversed walk  $\mathbf{B}^*(y, x) = \mathbf{B}(x, y)$ . To examine the reversed walk let  $b_i^*$  denote the sum of steps taken by  $\mathbf{B}^*$  between the  $(i - 1)$ st and  $i$ th time that a  $u \rightarrow u/2$  transition is chosen (i.e., consider block steps for the reversed walk), and let  $Z_s^* = 2^{-s+1}b_1^* + \dots + b_s^*$ . If we define  $b_i = -b_i^*$ , then the  $b_i$  are independent random variables from the same distribution as the blocks of  $\mathbf{B}$ , and so

$$\begin{aligned} \Pr[-2^{s-1}Z_s^* = j] &= \Pr[b_1 + 2b_2 + \dots + 2^{s-1}b_s = j] \\ &= \Pr[Z_s = j]. \end{aligned}$$

Lemma 4.5 thus bounds the second sum of (4.2) as well, and the theorem follows.  $\square$

Before proving Lemma 4.5 let us review the standard Fourier transform and the Plancherel identity. For any complex-valued function  $f$  on  $\mathbb{Z}_N$  and  $\omega = e^{2\pi i/N}$ , recall that the Fourier transform  $\hat{f} : \mathbb{Z}_N \rightarrow \mathbb{C}$  is given by  $\hat{f}(\ell) = \sum_{j=0}^{N-1} \omega^{\ell j} f(j)$ , and the Plancherel identity asserts that

$$N \sum_{j=0}^{N-1} |f(j)|^2 = \sum_{j=0}^{N-1} |\hat{f}(j)|^2.$$

For the distribution  $\mu$  of a  $\mathbb{Z}_N$ -valued random variable  $X$ , its Fourier transform is

$$\hat{\mu}(\ell) = \sum_{j=0}^{N-1} \omega^{\ell j} \mu(j) = E[\omega^{\ell X}].$$

Thus, given distributions  $\mu_1, \mu_2$  of two independent random variables  $Y_1, Y_2$ , the distribution  $\nu$  of  $X := Y_1 + Y_2$  has the Fourier transform  $\hat{\nu} = \hat{\mu}_1 \hat{\mu}_2$ , since

$$\begin{aligned} \hat{\nu}(\ell) &= E[\omega^{\ell X}] = E[\omega^{\ell(Y_1+Y_2)}] \\ (4.3) \quad &= E[\omega^{\ell Y_1}]E[\omega^{\ell Y_2}] = \hat{\mu}_1(\ell)\hat{\mu}_2(\ell). \end{aligned}$$

Generally, the distribution  $\nu$  of  $X := Y_1 + \dots + Y_s$  with independent  $Y_i$ s has the Fourier transform  $\hat{\nu} = \prod_{r=1}^s \hat{\mu}_r$ . Moreover, for the uniform distribution  $U$ , it is easy to check that

$$\hat{U}(\ell) = \begin{cases} 1, & \text{if } \ell = 0, \\ 0, & \text{otherwise.} \end{cases}$$

As the random variables  $2^r b_{s-r}$ s are independent,  $\hat{\nu}_s = \prod_{r=0}^{s-1} \hat{\mu}_r$  where  $\mu_r$  are the distributions of  $2^r b_{s-r}$ . The linearity of the Fourier transform and  $\hat{\nu}_s(0) = E[1] = 1$  yield

$$\widehat{\nu_s - U}(\ell) = \hat{\nu}_s(\ell) - \hat{U}(\ell) = \begin{cases} 0, & \text{if } \ell = 0, \\ \prod_{r=0}^{s-1} \hat{\mu}_r(\ell), & \text{otherwise.} \end{cases}$$

PROOF OF LEMMA 4.5. By Plancherel’s identity, it is enough to show that

$$\sum_{\ell=1}^{N-1} \left| \prod_{r=0}^{s-1} \hat{\mu}_r(\ell) \right|^2 \leq 2((1 + \xi^{2\lfloor s/m \rfloor})^{m-1} - 1).$$

Let  $A_r$  be the event that  $b_{s-r} = 0$  or  $1$ . Then,

$$\begin{aligned} \hat{\mu}_r(\ell) &= E[\omega^{\ell 2^r b_{s-r}}] \\ &= \Pr[b_{s-r} = 0] + \Pr[b_{s-r} = 1] \omega^{\ell 2^r} \\ &\quad + \Pr[\bar{A}_r] E[\omega^{\ell 2^r b_{s-r}} | \bar{A}_r], \end{aligned}$$

and, for  $x := \Pr[b_{s-r} = 0]$  and  $y := \Pr[b_{s-r} = 1]$ ,

$$\begin{aligned} |\hat{\mu}_r(\ell)| &\leq |x + y \omega^{\ell 2^r}| + (1 - x - y) |E[\omega^{\ell 2^r b_{s-r}} | \bar{A}_r]| \\ &\leq |x + y \omega^{\ell 2^r}| + 1 - x - y. \end{aligned}$$

Notice that

$$\begin{aligned} |x + y \omega^{\ell 2^r}|^2 &= \left( x + y \cos \frac{2\pi \ell 2^r}{N} \right)^2 + y^2 \sin^2 \frac{2\pi \ell 2^r}{N} \\ &= x^2 + y^2 + 2xy \cos \frac{2\pi \ell 2^r}{N}. \end{aligned}$$

If  $\cos \frac{2\pi \ell 2^r}{N} \leq 0$ , then

$$\begin{aligned} |\hat{\mu}_r(\ell)| &\leq (x^2 + y^2)^{1/2} + 1 - x - y \\ &= 1 - (x + y - (x^2 + y^2)^{1/2}). \end{aligned}$$

Since  $x = \Pr[b_{s-r} = 0] \geq 1/3$ , and  $y = \Pr[b_{s-r} = 1] \geq 1/9$ , it is easy to see that  $x + y - (x^2 + y^2)^{1/2}$  has its minimum when  $x = 1/3$  and  $y = 1/9$ . (For both partial derivatives are positive.) Hence,

$$|\hat{\mu}_r(\ell)| \leq \xi = 1 - \frac{4 - \sqrt{10}}{9} \quad \text{provided } \cos \frac{2\pi \ell 2^r}{N} \leq 0.$$

If  $\cos \frac{2\pi \ell 2^r}{N} > 0$ , we use the trivial bound  $\hat{\mu}_r(\ell) = E[\omega^{\ell 2^r b_{s-r}}] \leq 1$ .

For  $\ell = 1, \dots, N - 1$ , let  $\phi_s(\ell)$  be the number of  $r = 0, \dots, s - 1$  such that  $\cos \frac{2\pi \ell 2^r}{N} \leq 0$ . Then

$$(4.4) \quad \prod_{r=0}^{s-1} |\hat{\mu}_r(\ell)| \leq \xi^{\phi_s(\ell)}.$$

To estimate  $\phi_s(\ell)$ , we consider the binary expansion of

$$\ell/N = 0.\alpha_{\ell,1}\alpha_{\ell,2} \cdots \alpha_{\ell,s} \cdots,$$

$\alpha_{\ell,r} \in \{0, 1\}$  with  $\alpha_{\ell,r} = 0$  infinitely often. Hence,  $\ell/N = \sum_{r=1}^{\infty} 2^{-r} \alpha_{\ell,r}$ . The fractional part of  $\ell 2^r / N$  may be written as

$$\{\ell 2^r / N\} = 0.\alpha_{\ell,r+1}\alpha_{\ell,r+2} \cdots \alpha_{\ell,s} \cdots$$

Notice that  $\cos \frac{2\pi \ell 2^r}{N} \leq 0$  if the fractional part of  $\ell 2^r / N$  is (inclusively) between  $1/4$  and  $3/4$  which follows if  $\alpha_{r+1} \neq \alpha_{r+2}$ . Thus  $\phi_s(\ell)$  is at least as large as the number of alterations in the sequence  $(\alpha_{\ell,1}, \alpha_{\ell,2}, \dots, \alpha_{\ell,s+1})$ .

We now take  $m$  such that  $2^{m-1} < N < 2^m$ . Observe that, for  $\ell = 1, \dots, N - 1$ , the subsequences  $\alpha(\ell) := (\alpha_{\ell,1}, \alpha_{\ell,2}, \dots, \alpha_{\ell,m})$  of length  $m$  are pairwise distinct: If  $\alpha(\ell) = \alpha(\ell')$  for some  $\ell < \ell'$  then  $\frac{\ell' - \ell}{N}$  is less than  $\sum_{r \geq m+1} 2^{-r} \leq 2^{-m}$  which is impossible as  $N < 2^m$ . Similarly, for fixed  $r$  and  $\ell = 1, \dots, N - 1$ , all subsequences  $\alpha(\ell; r) := (\alpha_{\ell,r+1}, \alpha_{\ell,r+2}, \dots, \alpha_{\ell,r+m})$  are pairwise distinct. In particular, for fixed  $r$  with  $r = 0, \dots, \lfloor s/m \rfloor - 1$ , all subsequences  $\alpha(\ell; rm), \ell = 1, \dots, N - 1$ , are pairwise distinct. Since the fractional part  $\{\frac{2^{rm}\ell}{N}\} = 0.\alpha_{\ell,rm+1}\alpha_{\ell,rm+2} \cdots$  must be the same as  $\frac{\ell'}{N}$  for some  $\ell'$  in the range  $1 \leq \ell' \leq N - 1$ , there is a unique permutation  $\sigma_r$  of  $1, \dots, N - 1$  such that  $\alpha(\ell; rm) = \alpha(\sigma_r(\ell))$ . Writing  $|\alpha(\sigma_r(\ell))|_A$  for the number of alterations in  $\alpha(\sigma_r(\ell))$ , we have

$$\phi_s(\ell) \geq \sum_{r=0}^{\lfloor s/m \rfloor - 1} |\alpha(\sigma_r(\ell))|_A,$$

where  $\sigma_0$  is the identity. Therefore, (4.4) gives

$$\sum_{\ell=1}^{N-1} \left| \prod_{r=0}^{s-1} \hat{\mu}_r(\ell) \right|^2 \leq \sum_{\ell=1}^{N-1} \xi^{2 \sum_{r=0}^{\lfloor s/m \rfloor - 1} |\alpha(\sigma_r(\ell))|_A}.$$

Using

$$\begin{aligned} &\xi^{x+y} + \xi^{x'+y'} \\ &\leq \xi^{\min\{x,x'\} + \min\{y,y'\}} + \xi^{\max\{x,x'\} + \max\{y,y'\}} \end{aligned}$$

inductively, the above upper bound may be maximized when all  $\sigma_r$ s are the identity, that is,

$$\sum_{\ell=1}^{N-1} \left| \prod_{r=0}^{s-1} \hat{\mu}_r(\ell) \right|^2 \leq \sum_{\ell=1}^{N-1} \xi^{2 \lfloor s/m \rfloor |\alpha(\ell)|_A}.$$

Note that  $1/N \leq \ell/N \leq 1 - 1/N$  implies that  $\alpha(\ell)$  is neither  $(0, \dots, 0)$  nor  $(1, \dots, 1)$  (both are of length  $m$ ). This means that all  $\alpha(\ell)$  have at least one alteration. Since  $\alpha(\ell)$ s are pairwise distinct,

$$\sum_{\ell=1}^{N-1} \xi^{2 \lfloor s/m \rfloor |\alpha(\ell)|_A} \leq \sum_{\alpha: |\alpha|_A > 0} \xi^{2 \lfloor s/m \rfloor |\alpha|_A},$$

where the sum is taken over all sequences  $\alpha \in \{0, 1\}^m$  with  $|\alpha|_A > 0$ .

Let  $H(z)$  be the number of  $\alpha$ s with exactly  $z$  alterations. Then

$$H(z) = 2 \binom{m-1}{z},$$

and hence

$$\begin{aligned} \sum_{\alpha: |\alpha|_A > 0} \xi^{2\lfloor s/m \rfloor |\alpha|_A} &= 2 \sum_{z=1}^{m-1} \binom{m-1}{z} \xi^{2\lfloor s/m \rfloor z} \\ &= 2((1 + \xi^{2\lfloor s/m \rfloor})^{m-1} - 1). \end{aligned} \quad \square$$

REMARK 4.6. For the reader interested in applying these methods to show a Birthday-type result for other problems, it is worth noting that a Fourier approach can also be used to show that  $B^s(u, v)$  decreases quickly, and so  $A_T, A_T^* = O(1)$ .

For the distribution  $\nu_s$  of  $X_s$  the Plancherel identity gives

$$\begin{aligned} \max_v \Pr[X_s = v] &= \max_v \nu_s(v)^2 \leq \sum_{w=0}^{N-1} \nu_s(w)^2 \\ &= \frac{1}{N} \sum_{\ell=0}^{N-1} |\hat{\nu}_s(\ell)|^2 = \frac{1}{N} \sum_{\ell=0}^{N-1} \left| \prod_{r=0}^{s-1} \hat{\mu}_r(\ell) \right|^2. \end{aligned}$$

For  $\ell = 0, 1, \dots, N-1$ , let  $\phi_s(\ell)$  be the number of  $r = 0, \dots, s-1$  such that  $\cos \frac{2\pi \ell 2^r}{N} \leq 0$ . Then

$$\prod_{r=0}^{s-1} |\hat{\mu}_r(\ell)| \leq \xi^{\phi_s(\ell)}.$$

Take  $m$  such that  $2^{m-1} < N < 2^m$ . Then, for  $s \leq m-1$  and any (fixed) binary sequence  $\alpha_1, \dots, \alpha_s$  (that is,  $\alpha_j \in \{0, 1\}$ ), there are at most  $\lceil 2^{-s} N \rceil$   $\ell$ s such that the binary expansion of  $\ell/N$  up to  $s$  digits is  $0.\alpha_1 \dots \alpha_s$ . Since there are at most  $2e^{-\Omega(s)} 2^s$  binary sequences with fewer than  $(s-1)/3$  alterations,

$$\prod_{r=0}^{s-1} |\hat{\mu}_r(\ell)| = 2e^{-\Omega(s)}$$

except for at most  $2e^{-\Omega(s)} 2^s \lceil 2^{-s} N \rceil = 2e^{-\Omega(s)} N$  values of  $\ell$ . Using a trivial bound  $\prod_{r=0}^{s-1} |\hat{\mu}_r(\ell)| \leq 1$  for such  $\ell$ 's, we have

$$\max_v \Pr[X_s = v] = 2e^{-\Omega(s)} + 2e^{-\Omega(s)} = 2e^{-\Omega(s)}.$$

If  $s > m-1$ , then  $\prod_{r=0}^{s-1} |\hat{\mu}_r(\ell)| \leq \prod_{r=0}^{m-2} |\hat{\mu}_r(\ell)|$  implies that

$$\max_v \Pr[X_s = v] = 2e^{-\Omega(m-1)} = O(N^{-\Omega(1)}).$$

One might expect that the correct order of the mixing time of the Block walk  $X_s$  is indeed  $\Theta(\log p \log \log p)$ . This is in fact the case, at least for certain values of  $p$  and  $x$ .

**THEOREM 4.7.** *If  $p = 2^t - 1$  and  $x = p - 1$ , then the block walk has mixing time  $\tau(1/4) = \Theta(\log p \log \log p)$ .*

**PROOF.** The upper bound on mixing time,  $O(\log p \log \log p)$ , was shown in Theorem 4.4 via a Fourier argument.

The proof of a lower bound of  $\Omega(\log p \log \log p)$  is modeled after an arguments of Hildebrand [4] which in turn closely follows a proof of Chung, Diaconis and Graham [2]. The basic idea is by now fairly standard: choose a function and show that its expectation under the stationary distribution and under the  $n$ -step distribution  $P^n$  are far apart with sufficiently small variance to conclude that the two distributions ( $P^n$  and  $\pi$ ) must differ significantly. Theorem 4.7 is not used in main results of this paper, and the proof is fairly long, and so it is left for the [Appendix](#).  $\square$

**5. Distinguished point methods.** The Rho algorithm can be parallelized to  $J$  processors via the Distinguished Points method of van Oorschot and Wiener [19]. To do this, start with a global choice of (random) partition  $S_1 \amalg S_2 \amalg S_3$  (i.e., a common iterating function  $F$ ), and choose  $J$  initial values  $\{y_0^j\}_{j=1}^J$  from  $\mathbb{Z}_N$ , one per processor. Then run the Rho walk on processor  $j$  starting from initial state  $g^{(y_0^j)}$ , until a collision occurs between either two walks or a walk and itself. To detect a collision let  $\varphi: G \rightarrow \{0, 1\}$  be an easily computed hash function with support  $\{x \in G: \varphi(x) = 1\}$  to be called the *distinguished points*. Each time a distinguished point is reached by a processor, it is sent to a central repository and compared against previously received states. Once a distinguished point is reached twice then a collision has occurred, and the discrete logarithm can likely be found while, conversely, once a collision occurs, the collision will be detected the next time a distinguished point is reached.

The proofs in previous sections immediately imply a factor of  $J$  speed up when parallelizing. To see this, suppose the initial values  $\{y_0^j\}_{j=1}^J$  are chosen uniformly at random. Run a Rho walk for some  $\mathcal{T}$  steps per processor, and then define  $\{X_i\}$  by starting with the Rho walk of processor #1, then appending that from processor #2, etc.; that is, if  $Y_i^j$  denotes the  $i$ th state of copy  $j$  of the walk, for  $i \in \{0, 1, \dots, \mathcal{T}\}$  and  $j \in \{0, 1, 2, \dots, J-1\}$ , then  $X_i = Y_{i \bmod (\mathcal{T}+1)}^{i \operatorname{div}(\mathcal{T}+1)}$  for  $i \in \{0, 1, \dots, J(\mathcal{T}+1)-1\}$ . This is a time-dependent random walk which follows the Rho walk, except at multiples of time  $\mathcal{T} + 1$  where it instead jumps to a uniformly random state. Since our proofs involve pessimistic estimates on the distance of a distribution from uniform, and these jumps result in uniform samples, then they can only improve the result. Hence this effectively leads to a Rho walk with  $J(\mathcal{T} + 1) - 1$  steps, and a

factor  $J$  speed-up per processor is achieved. If the initial values were not uniform then discard the first  $O(\log^2 N)$  steps per processor and treat the next state as the initial value which by Theorem 4.4 will give a nearly uniform start state.

APPENDIX

PROOF OF THEOREM 4.7. Our approach follows that taken in Section 4 of Hildebrand [4], “A proof of Case 2.” Recall that we will show that  $E_U(f)$  and  $E_{P_n}(f)$  are far apart for some function  $f$  with sufficiently small variance to conclude that the two distributions ( $P^n$  and the uniform distribution  $U$  on  $\mathbb{Z}_p$ ) must differ significantly.

More precisely let  $P_n$  be the distribution of the block walk on  $\mathbb{Z}_p$  starting at state  $u = 0$  and proceeding for  $n$  steps. For some  $\alpha > 0$ , let

$$A = \{y : |f(y) - E_U(f)| \geq \alpha \sqrt{\text{Var}_U(f)}\}.$$

By Chebyshev’s inequality,  $U(A) \leq 1/\alpha^2$ . Also, for some  $\beta > 0$  let

$$B = \{y : |f(y) - E_{P_n}(f)| \geq \beta \sqrt{\text{Var}_{P_n}(f)}\}.$$

By Chebyshev’s inequality,  $P_n(B) \leq 1/\beta^2$ . If  $A^c \cap B^c = \emptyset$ , then  $P_n(A^c) \leq P_n(B) \leq 1/\beta^2$ , and so

$$\min_{v \in A^c} \frac{P_n(v)}{U(v)} \leq \frac{P_n(A^c)}{U(A^c)} \leq \frac{1/\beta^2}{1 - 1/\alpha^2}.$$

If  $\sqrt{\text{Var}_U(f)}, \sqrt{\text{Var}_{P_n}(f)} = o(|E_U(f) - E_{P_n}(f)|)$  for a sequence  $n(p) = \Omega(\log p \log \log p)$ , then as  $p \rightarrow \infty$  it is possible to choose  $\alpha, \beta \rightarrow \infty$  such that  $A^c \cap B^c = \emptyset$ . The theorem then follows as  $\min_{v \in \Omega} \frac{P_n(v)}{U(v)} \xrightarrow{p \rightarrow \infty} 0$ .

The “separating” function  $f : \mathbb{Z}_p \rightarrow \mathbb{C}$  to be used here is

$$f(k) = \sum_{j=0}^{t-1} q^{k2^j} \quad \text{where } q = e^{2\pi i/p}.$$

Then  $E_U(f) = \frac{1}{p} \sum_{j=0}^{t-1} \sum_{k=0}^{p-1} (q^{2^j})^k = \frac{1}{p} \sum_{j=0}^{t-1} 0 = 0$  since  $q^\alpha$  is a primitive root of unity when  $p$  is prime and  $1 \leq \alpha < p$ . Likewise,  $E_U(f\bar{f}) = \frac{1}{p} \sum_{j,j'=0}^{t-1} \sum_{k=0}^{p-1} q^{k*2^j} \overline{q^{k*2^{j'}}} = \frac{1}{p} \sum_{j=0}^{t-1} p = t$  by the orthogonality relationship of roots of unity. It follows that  $\text{Var}_U(f) = t$ .

The Block walk on  $\mathbb{Z}_p$  with  $n = rt$  steps will be considered where  $r \in \mathbb{N}$  will be chosen later. Let  $P_n(\cdot)$  denote the distribution of  $Z_n = 2^{n-1}b_1 + 2^{n-2}b_2 + \dots + b_n$  induced by  $n$  steps of the block walk starting at state  $u = 0$ . A generic increment will be denoted by  $b$ , since the  $b_i$  are independent random variables from the same distribution.

It will be useful to introduce a bit of notation.

If  $\alpha \in \mathbb{Z}$  and  $y \in \mathbb{Z}_p$ , then define

$$\mu_\alpha(y) = \Pr[2^\alpha b = y] = \Pr[b = y2^{-\alpha}].$$

Recall the Fourier transform of a distribution  $\nu$  on  $\mathbb{Z}_p$  is given by  $\hat{\nu}(\ell) = \sum_{k=0}^{p-1} q^{\ell k} \nu(k) = E_\nu(q^{\ell k})$ . Properties of certain Fourier transforms are required in our work. First, since  $p = 2^t - 1$  then  $\mu_\alpha(y) = \mu_{\alpha+ct}(y)$  for  $c \in \mathbb{N}$ , and so  $\hat{\mu}_\alpha(y) = \hat{\mu}_{(\alpha \bmod t)}(y)$ . By this and (4.3), if  $y \in \mathbb{Z}_p$ , then

$$\hat{P}_n(y) = \prod_{\alpha=0}^{n-1} \hat{\mu}_\alpha(y) = \left( \prod_{\alpha=0}^{t-1} \hat{\mu}_\alpha(y) \right)^r = \hat{P}_t(y)^r.$$

Also,  $\mu_\alpha(2^j y) = \mu_{(\alpha-j)}(y)$ , and so

$$\forall j \in \mathbb{N} : \hat{P}_t(2^j y) = \prod_{\alpha=0}^{t-1} \hat{\mu}_\alpha(2^j y) = \prod_{\alpha=0}^{t-1} \hat{\mu}_\alpha(y) = \hat{P}_t(y).$$

Finally, for  $0 \leq j \leq t - 1$ , define

$$\Pi_j = \hat{P}_t(2^j - 1) = \prod_{\alpha=0}^{t-1} \hat{\mu}_\alpha(2^j - 1).$$

Note that  $\hat{P}_t(1 - 2^j) = \hat{P}_t(2^j - 1)$  because  $\mu_\alpha(y) = \mu_\alpha(-y)$  when the step sizes are  $\{1, x\} = \{1, -1\}$ . Also,  $\Pi_{t-j} = \Pi_j$  because, modulo  $p$ ,  $2^{t-j} - 1 = 2^{-j} - 1$ , and so  $\hat{P}_t(2^{t-j} - 1) = \hat{P}_t(2^{-j} - 1) = \hat{P}_t(2^j(2^{-j} - 1)) = \hat{P}_t(2^j - 1)$ .

Now turn to mean and variance:

$$\begin{aligned} E_{P_n}(f) &= \sum_k P_n(k) \sum_{j=0}^{t-1} q^{k2^j} = \sum_{j=0}^{t-1} \hat{P}_n(2^j) \\ &= \sum_{j=0}^{t-1} \hat{P}_t(2^j)^r = \sum_{j=0}^{t-1} \hat{P}_t(1)^r = t \Pi_1^r. \end{aligned}$$

Likewise,

$$\begin{aligned} E_{P_n}(f \bar{f}) &= \sum_k P_n(k) \sum_{j,j'=0}^{t-1} q^{k(2^j - 2^{j'})} = \sum_{j,j'=0}^{t-1} \hat{P}_n(2^j - 2^{j'}) \\ &= \sum_{j,j'=0}^{t-1} \hat{P}_t(2^j - 2^{j'})^r = \sum_{j=0}^{t-1} \sum_{\beta=0}^{t-1} \hat{P}_t(2^j(1 - 2^\beta))^r \\ &= t \sum_{\beta=0}^{t-1} \hat{P}_t(2^\beta - 1)^r = t \sum_{\beta=0}^{t-1} \Pi_\beta^r. \end{aligned}$$

It follows that

$$\text{Var}_{P_n}(f) = E_{P_n}(f\bar{f}) - E_{P_n}(f)E_{P_n}(\bar{f}) = t \sum_{j=0}^{t-1} \Pi_j^r - t^2 |\Pi_1|^{2r}.$$

To apply these relations in Chebyshev’s inequality the quantities  $\Pi_j$  need to be examined further. Let  $b$  denote the increment taken but with arithmetic NOT done modulo  $p$ , so that  $b = p + 1$  is possible, that is, repeatedly decide with probability  $1/3$  whether to terminate and if not then flip a coin and decide whether to add  $+1$  or  $-1$ . Let  $a_k = \Pr[b = k]$ , and note that also  $a_k = \Pr[b = -k]$  since the nondoubling steps are symmetric, that is,  $u \rightarrow u + 1$  and  $u \rightarrow u + x = u - 1$ . Then  $a_k$  satisfies the recurrence relation

$$a_k = \frac{1}{3}(a_{k-1} + a_{k+1}), \quad a_0 = \frac{1}{3} + \frac{2}{3}a_1, \quad a_\infty = 0$$

which has solution  $a_k = \frac{1}{\sqrt{5}} \left(\frac{3-\sqrt{5}}{2}\right)^{|k|}$ . For  $y \in \mathbb{R}$ , let

$$G(y) = \sum_{k=-\infty}^{\infty} a_k e^{2\pi iky} = \frac{1}{\sqrt{5}} \frac{1 - ((3 - \sqrt{5})/2)^2}{1 + ((3 - \sqrt{5})/2)^2 - (3 - \sqrt{5}) \cos(2\pi y)}.$$

Since

$$\mu_\alpha(k) = \Pr[b = k2^{-\alpha}] = \Pr[b2^\alpha \equiv k \pmod p] = \sum_{\{b: b2^\alpha \equiv k \pmod p\}} a_b,$$

then

$$\begin{aligned} \Pi_j &= \prod_{\alpha=0}^{t-1} \hat{\mu}_\alpha(2^j - 1) = \prod_{\alpha=0}^{t-1} \sum_{k=-\infty}^{\infty} q^{k(2^j-1)} \mu_\alpha(k) \\ &= \prod_{\alpha=0}^{t-1} \sum_{b=-\infty}^{\infty} a_b e^{2\pi i b 2^\alpha (2^j-1)/p} \\ &= \prod_{\alpha=0}^{t-1} G\left(\frac{2^\alpha (2^j - 1)}{p}\right). \end{aligned}$$

We can now show the necessary bounds. Recall that  $n = rt$  for some sequence  $r = r(t) \in \mathbb{N}$  to be defined. Let  $\lambda = \lambda(t) \in \mathbb{R}$  be a sequence such that  $\lambda \xrightarrow{t \rightarrow \infty} \infty$ ,  $\lambda = o(\log t)$ , and

$$r = \frac{\log t}{2 \log(1/|\Pi_1|)} - \lambda$$

is an integer. Such a sequence will exist if  $|\Pi_1|$  is bounded away from 0.

CLAIM 1.  $|\Pi_1|$  is bounded away from 0 and 1 as  $p = 2^t - 1 \rightarrow \infty$ .

PROOF. First, a few preliminary calculations are necessary. If  $y \in [0, 1/2]$ , then

$$G(y) \geq \frac{1}{\sqrt{5}} \frac{1 - ((3 - \sqrt{5})/2)^2}{1 + ((3 - \sqrt{5})/2)^2 + (3 - \sqrt{5})} = \frac{1}{5}.$$

Also, since  $y \in \mathbb{R}$ , then  $|G(y)| \leq \sum a_k |e^{2\pi iky}| = \sum a_k = 1$  with equality at  $y = 0$ . And, since the first derivative satisfies the relation  $0 \geq G'(y) \geq -2\pi(3 - \sqrt{5}) > -5$ , then  $G(\varepsilon) \geq G(0) - 5\varepsilon = 1 - 5\varepsilon$  for  $\varepsilon \geq 0$ .

To show that  $\Pi_1$  is bounded away from 1 observe that

$$\Pi_1 \leq G(2^{t-1}/p) * 1^{t-1} \xrightarrow{t \rightarrow \infty} G(1/2) = \frac{1}{5}.$$

To bound  $\Pi_1$  away from 0 note that since  $G(y)$  is decreasing for  $y \in [0, 1/2]$ , then

$$\begin{aligned} \Pi_1 &= \prod_{\alpha=0}^{t-1} G(2^\alpha/p) \\ &= \prod_{\alpha=t-5}^{t-1} G(2^\alpha/p) \prod_{\alpha=0}^{t-6} G(2^\alpha/p) \\ &\geq \prod_{\beta=1}^5 \frac{1}{5} \prod_{\beta=6}^t (1 - 5 \cdot 2^{-\beta} - 5 \cdot 2^{-t}) \\ &\geq 5^{-5} e^{-5/(2^5(1-5/2^6-5/2^t))}. \end{aligned}$$

The final inequality used the relation is  $\ln(1 - x) \geq \frac{-x}{1-x}$ .  $\square$

CLAIM 2.  $(\text{Var}_U(f))^{1/2} = o(|E_{P_n}(f) - E_U(f)|)$ .

PROOF. Since  $|\Pi_1|$  is bounded away from 0 and 1, then

$$|E_{P_n}(f) - E_U(f)| = t|\Pi_1|^r = \sqrt{t}|\Pi_1|^{-\lambda}.$$

The claim follows as  $(\text{Var}_U(f))^{1/2} = \sqrt{t}$  and  $|\Pi_1|^\lambda \rightarrow 0$ .  $\square$

CLAIM 3.  $(\text{Var}_{P_n}(f))^{1/2} = o(|E_{P_n}(f) - E_U(f)|)$ .

PROOF. Assume

$$\frac{1}{t} \sum_{j=0}^{t-1} \left( \frac{\Pi_j}{\Pi_1^2} \right)^r \xrightarrow{t \rightarrow \infty} 1.$$

Then the claim follows from

$$(\text{Var}_{P_n}(f))^{1/2} = t|\Pi_1|^r \sqrt{\frac{1}{t} \sum_{j=0}^{t-1} \left(\frac{\Pi_j}{\Pi_1^2}\right)^r - 1} = o(t|\Pi_1|^r).$$

Hildebrand [4] requires 4 pages (pages 351–354) to prove the assumption, albeit for a different function  $G(y)$ . Fortunately we do not need to rework his proof as it does not make explicit use of  $G(y)$  but instead depends on only a few properties of it. There are two facts required in his proof. First, he shows the following:

FACT 1. There is some  $t_0$  such that if  $t \geq t_0$ , then  $\Pi_j \leq \Pi_1$  for all  $j \geq 1$ .

Hildebrand’s proof of this utilizes the following properties:  $G(y) = G(1 - y) = G(-y)$ ,  $G(y)$  is decreasing when  $y \in [0, 1/4]$ , and  $G(\frac{1}{2} - 2^{-i})$  is decreasing in  $i \geq 3$  [this corrects an error where he should have claimed

$$\lim_{p=2^t-1 \rightarrow \infty} G\left(\frac{2^{t-1} - 2^{t-j-1}}{p}\right) \leq G(3/8)$$

instead of  $G(1/2)$ ]. These conditions apply to our choice of  $G(y)$  as is shown in the proof of our Claim 1.

The second necessary tool is the following:

FACT 2. There exists constants  $c_0, t_1$  such that for  $t \geq t_1$  and  $t^{1/3} \leq j \leq t/2$ , then

$$\frac{\Pi_j}{\Pi_1^2} \leq 1 + \frac{c_0}{2^j}.$$

Hildebrand’s proof utilizes the following properties:  $G'(0) = 0$ ,  $|G'(y)| \leq A$ , and  $G(y) \geq B$  for all  $y$  and some  $A, B > 0$ . All three conditions were shown in the proof of our Claim 1, and so his argument will carry through.

The two facts can now be combined to finish the proof. Let  $t$  be sufficiently large that  $\frac{c_0}{2^{(t^{1/3})}} \leq 1/r$  which is possible as  $r = O(\log t)$ . Recalling that  $\Pi_{t-j} = \Pi_j$ , then

$$\begin{aligned} \sum_{j=1}^{t-1} \left| \left(\frac{\Pi_j}{\Pi_1^2}\right)^r - 1 \right| &\leq 2 \sum_{1 \leq j < t^{1/3}} \frac{1}{\Pi_1^r} + 2 \sum_{t^{1/3} \leq j \leq t/2} r^2 \frac{c_0}{2^j} \\ &\leq 2t^{1/3} t^{1/2} \Pi_1^\lambda + 2 \frac{t}{2} \frac{r^2 c_0}{2^{(t^{1/3})}} \\ &= o(t). \end{aligned}$$

Since  $\Pi_0 = 1$ , then

$$\frac{1}{t} \sum_{j=0}^{t-1} \left( \frac{\Pi_j}{\Pi_1^2} \right)^r \leq \frac{1}{t} \left( \frac{1}{\Pi_1^{2r}} + (t-1) + o(t) \right) = 1 + o(1).$$

In the other direction, since  $\text{Var}_{P_n}(f) \geq 0$ , then  $\frac{1}{t} \sum_{j=0}^{t-1} \left( \frac{\Pi_j}{\Pi_1^2} \right)^r \geq 1$ .  $\square$

**Acknowledgments.** The authors thank S. Kijima, S. Miller, I. Mironov, R. Venkatesan and D. Wilson for several helpful discussions.

## REFERENCES

- [1] ALDOUS, D. and FILL, J. (2009). *Reversible Markov Chains and Random Walks on Graphs*. Available at <http://www.stat.berkeley.edu/users/aldous/RWG/book.html>.
- [2] CHUNG, F. R. K., DIACONIS, P. and GRAHAM, R. L. (1987). Random walks arising in random number generation. *Ann. Probab.* **15** 1148–1165. [MR893921](#)
- [3] CRANDALL, R. and POMERANCE, C. (2005). *Prime Numbers*, 2nd ed. Springer, New York. [MR2156291](#)
- [4] HILDEBRAND, M. (2006). On the Chung–Diaconis–Graham random process. *Electron. Comm. Probab.* **11** 347–356 (electronic). [MR2274529](#)
- [5] LE GALL, J.-F. and ROSEN, J. (1991). The range of stable random walks. *Ann. Probab.* **19** 650–705. [MR1106281](#)
- [6] KIM, J.-H. MONTENEGRO, R. and TETALI, P. (2007). Near optimal bounds for collision in Pollard Rho for discrete log. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science* 215–223. IEEE Computer Society Press, Los Alamitos, CA.
- [7] LYONS, R., PERES, Y. and SCHRAMM, O. (2003). Markov chain intersections and the loop-erased walk. *Ann. Inst. H. Poincaré Probab. Statist.* **39** 779–791. [MR1997212](#)
- [8] MILLER, S. D. and VENKATESAN, R. (2006). Spectral analysis of Pollard Rho collisions. In *Algorithmic Number Theory. Lecture Notes in Computer Science* **4076** 573–581. Springer, Berlin. [MR2282950](#)
- [9] MILLER, S. D. and VENKATESAN, R. (2009). Non-degeneracy of Pollard Rho collisions. *Int. Math. Res. Not. IMRN* **1** 1–10. [MR2471294](#)
- [10] NAZAROV, F. and PERES, Y. (2010). The birthday problem for finite reversible Markov chains: A uniform bound. Preprint.
- [11] PAK, I. (2002). Mixing time and long paths in graphs. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms* 321–328. ACM, New York.
- [12] POHLIG, S. C. and HELLMAN, M. E. (1978). An improved algorithm for computing logarithms over  $\text{GF}(p)$  and its cryptographic significance. *IEEE Trans. Information Theory* **IT-24** 106–110. [MR0484737](#)
- [13] POLLARD, J. M. (1975). A Monte Carlo method for factorization. *Nordisk Tidskr. Informationsbehandling (BIT)* **15** 331–334. [MR0392798](#)
- [14] POLLARD, J. M. (1978). Monte Carlo methods for index computation (mod  $p$ ). *Math. Comp.* **32** 918–924. [MR0491431](#)
- [15] POMERANCE, C. (2008). Elementary thoughts on discrete logarithms. In *Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography. Mathematical Sciences Research Institute Publications* **44** 385–396. Cambridge Univ. Press, Cambridge. [MR2467551](#)

- [16] SHOUP, V. (1997). Lower bounds for discrete logarithms and related problems. In *Advances in Cryptology—EUROCRYPT'97 (Konstanz)*. *Lecture Notes in Computer Science* **1233** 256–266. Springer, Berlin. [MR1603068](#)
- [17] TESKE, E. (1998). Speeding up Pollard's Rho method for computing discrete logarithms. In *Algorithmic Number Theory (Portland, OR, 1998)*. *Lecture Notes in Computer Science* **1423** 541–554. Springer, Berlin. [MR1726100](#)
- [18] TESKE, E. (2001). Square-root algorithms for the discrete logarithm problem (a survey). In *Public-key Cryptography and Computational Number Theory (Warsaw, 2000)* 283–301. de Gruyter, Berlin. [MR1881641](#)
- [19] VAN OORSCHOT, P. C. and WIENER, M. J. (1999). Parallel collision search with cryptanalytic applications. *J. Cryptology* **12** 1–28. [MR1664774](#)

J. H. KIM  
DEPARTMENT OF MATHEMATICS  
YONSEI UNIVERSITY  
SEOUL, 120-749  
KOREA  
AND  
NATIONAL INSTITUTE FOR  
MATHEMATICAL SCIENCES  
628 DAEDUK BOULEVARD  
DAEJEON, 305-340  
KOREA  
E-MAIL: [jehkim@yonsei.ac.kr](mailto:jehkim@yonsei.ac.kr)  
[jehkim@nims.re.kr](mailto:jehkim@nims.re.kr)

Y. PERES  
MICROSOFT RESEARCH  
ONE MICROSOFT WAY  
REDMOND, WASHINGTON 98052  
USA  
E-MAIL: [peres@microsoft.com](mailto:peres@microsoft.com)

R. MONTENEGRO  
DEPARTMENT OF MATHEMATICAL SCIENCES  
UNIVERSITY OF MASSACHUSETTS LOWELL  
LOWELL, MASSACHUSETTS 01854  
USA  
E-MAIL: [ravi\\_montenegro@uml.edu](mailto:ravi_montenegro@uml.edu)

P. TETALI  
SCHOOL OF MATHEMATICS  
GEORGIA INSTITUTE OF TECHNOLOGY  
ATLANTA, GEORGIA 30332  
USA  
E-MAIL: [tetali@math.gatech.edu](mailto:tetali@math.gatech.edu)