Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE @ DIRECT®

Journal of Algorithms 50 (2004) 215–231

Journal of Algorithms

www.elsevier.com/locate/jalgor

Strong bias of group generators: an obstacle to the "product replacement algorithm"

László Babai [a] and Igor Pak [b,*]

[a] Department of Computer Science, University of Chicago, Chicago, IL 60637, USA
[b] Department of Mathematics, MIT, Cambridge, MA 02139, USA

Received 21 July 2000

Abstract

Let G be a finite group. Efficient generation of *nearly uniformly distributed* random elements in G, starting from a given set of generators of G, is a central problem in computational group theory. In this paper we demonstrate a weakness in the popular "product replacement algorithm," widely used for this purpose. The main results are the following. Let $\mathcal{N}_k(G)$ be the set of generating k-tuples of elements of G. Consider the distribution of the first components of the k-tuples in $\mathcal{N}_k(G)$ induced by the uniform distribution over $\mathcal{N}_k(G)$. We show that there exist infinite sequences of groups G such that this distribution is very far from uniform in two different senses: (1) its variation distance from uniform is $> 1 - \epsilon$; and (2) there exists a short word (of length $(\log\log|G|)^{O(k)}$) which separates the two distributions with probability $1-\epsilon$. The class of groups we analyze is direct powers of alternating groups. The methods used include statistical analysis of permutation groups, the theory of random walks, the AKS sorting network, and a randomized simulation of monotone Boolean operations by group operations, inspired by Barrington's work on bounded-width branching programs. The problem is motivated by the *product replacement algorithm* which was introduced in [Comm. Algebra 23 (1995) 4931–4948] and is widely used. Our results show that for certain groups the probability distribution obtained by the product replacement algorithm has a bias which can be detected by a short straight line program.

[*] Corresponding author.
E-mail addresses: laci@cs.chicago.edu (L. Babai), pak@math.mit.edu (I. Pak).

0196-6774/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0196-6774(03)00091-9

1. Introduction

Let G be a finite group. A sequence of k group elements $(g_1, \dots, g_k)$ $(g_i \in G)$ is called a *generating k-tuple* of G if the g_i generate G. Let $\mathcal{N}_k(G)$ be the set of all generating k-tuples of G, and let $N_k(G) = |\mathcal{N}_k(G)|$. Let $\varphi_k(G)$ denote the probability that k uniformly distributed independent random group elements generate G:

$$\varphi_k(G) = \frac{N_k(G)}{|G|^k}.$$

Let Q^k denote the probability distribution on G of the first components of k-tuples chosen uniformly from $\mathcal{N}_k(G)$. This distribution appears as the limiting distribution obtained by the "product replacement algorithm," a widely used heuristic intended to rapidly generate nearly uniformly distributed random elements in G (see the next section). While the question of mixing rate for this algorithm remains open for small values of k (see Section 3), we show that even the limiting distribution Q^k can be *very far from uniform* in this case.

2. The main results

The groups on which we demonstrate this anomaly are the direct powers

$$G = A_n^m = A_n \times A_n \times \cdots \times A_n \quad (m \text{ times}),$$

where A_n is the alternating group of degree n (the group of even permutations of $n \geqslant 5$ elements).

A *probability distribution* over G is a function $\mathrm{R} : G \to \mathbb{R}$ such that $(\forall g \in G)$ $(\mathrm{R}(g) \geqslant 0)$ and $\sum_{g \in G} \mathrm{R}(g) = 1$.

For a function $\mathrm{T} : G \to \mathbb{R}$ and a subset $B \subseteq G$ we write $\mathrm{T}(B) = \sum_{g \in B} \mathrm{T}(g)$. The *total variation* of T is half of its ℓ_1-norm: $\|\mathrm{T}\|_{tv} = \max_{B \subset G} |T(B)| = (1/2) \sum_{g \in G} |\mathrm{T}(g)|$. The *variation distance* of the probability distributions R and S over G is defined as $\|\mathrm{R} - \mathrm{S}\|_{tv}$. This quantity is between 0 and 1.

Let U denote the uniform distribution over G. The *bias* of the distribution R is the variation distance between R and U. In other words, the bias of R is

$$\|\mathrm{R} - \mathrm{U}\|_{tv} = \max_{B \subset G} \bigl|\mathrm{R}(B) - \mathrm{U}(B)\bigr| = \frac{1}{2} \sum_{g \in G} \left|\mathrm{R}(g) - \frac{1}{|G|}\right|.$$

The *bias of a subset* $B \subset G$ under R is the quantity $|\mathrm{R}(B) - \mathrm{U}(B)|$.

Theorem 2.1. *Let $G = A_n^m$, where $m = n!/8$. Then*

$$\bigl\|\mathrm{Q}^k - \mathrm{U}\bigr\|_{tv} \to 1 \quad \text{as } n \to \infty$$

assuming $k \geqslant 4$ and $k = o(n)$.

We note that for $m = n!/8$, the group G is generated by 2 elements, but a uniform random pair of elements (or even a random k-tuple of elements for $k = o(n)$) is unlikely to generate it. The intuition behind the proof builds on this discrepancy.

To prove the theorem we find an explicit set B such that $\mathrm{Q}^k(B) \to 0$ while $\mathrm{U}(B) \to 1$. The set B can be chosen to be a union of conjugacy classes in G and therefore has direct significance to applications in computational group theory.

Let $\mathbf{w}$ be a word over the alphabet $\{x_i^{\pm 1}, i = 1, 2, \ldots\}$. Substituting elements of G for the x_i assigns $\mathbf{w}$ a value in G. Assume that the x_i are chosen independently from the probability distribution P over G. We denote by $\mathbf{w}[P]$ the probability distribution of the value of $\mathbf{w}$.

Terminology. We say that an event is *factorially unlikely* if its probability is $O(n^{-cn})$ for some constant $c > 0$; and it is *factorially likely* if its probability is $1 - O(n^{-cn})$. (The letter c will be used to denote different positive constants at each occurrence in this paper. The expression "factorially (un)likely" will always refer to the parameter n regardless of the other parameters such as k and m involved in the definition of the groups in question.)

Theorem 2.2. *There exists a family of words $\mathbf{w}_{n,k}$ with the following properties. The length of $\mathbf{w}_{n,k}$ is $n^{O(k)}$. Let $\omega(n) \to \infty$, $\omega(n) = o(n)$. Also, let $k = k(n) \geqslant 4$ and $k = o(n)$. Set $m = n^{k\omega(n)}$. Let $G = A_n^m$. Then $\mathbf{w}[\mathrm{Q}^k] = 1$ is factorially likely (has probability $1 - O(n^{-cn})$), while $\mathbf{w}[\mathrm{U}] = 1$ is factorially unlikely.*

Remark 2.3. Note that if we choose $\omega(n)$ to be $\sqrt{n}$, the length of $\mathbf{w}$ *becomes* $(\log\log|G|)^{O(k)}$. This is at most polylogarithmic compared to the bit-length of the input: the names of most group elements require $\Omega(\log|G|)$ bits (in any encoding of the group elements).

Intuitively, the theorem implies that the probability distribution Q^k is so far from uniform that even the evaluation of a polylog-length (compared to $\log(|G|)$) word will show extreme bias if we use Q^k as a substitute for the uniform distribution. (Assuming that the group elements are encoded by strings of uniform length, this length must be $\Omega(\log|G|)$.) The proof of Theorem 2.2 is based on a probabilistic simulation of the monotone Boolean operations AND and OR by group operations. We also employ some known results in probabilistic group theory and the theory of random walks.

3. The "product replacement algorithm"

It is known that nearly uniformly distributed random elements of a finite group can be constructed using a polynomial number of group operations, starting from any given set of generators [Bbl]. However, the number of operations proven in [Bbl] to guarantee near-uniformity is rather large $O((\log|G|)^5)$, not suitable in practice. Therefore, heuristic algorithms are used.

One such heuristic, the *product replacement algorithm*, is an important recent advancement in symbolic algebra [CLMNO] (see also [Bb3,Ka,Pa2,PaB]). It was designed by Leedham-Green and Soicher to generate efficiently nearly uniform group elements. It is by far the most popular practical generator of random group elements,[1] implemented in the two symbolic algebra packages most frequently used in computational group theory, GAP [Sc] and Magma [BoC].

The product replacement algorithm works as follows [CLMNO]. We construct a Markov chain $\mathbf{M} = \mathbf{M}(G, k) = \{X_t\}$ on $\mathcal{N}_k(G)$. Let $X_t = (g_1, \ldots, g_k) \in \mathcal{N}_k(G)$. Define

$$X_{t+1} = (g_1, \ldots, h_j, \ldots, g_k),$$

where $h_j = g_j g_i^{\pm 1}$ or $h_j = g_i^{\pm 1} g_j$, where the pair (i, j), $1 \leqslant i, j \leqslant k$, $i \neq j$ is chosen uniformly; the order of multiplication and the exponent ±1 are determined by independent flips of a fair coin.

The algorithm runs the Markov chain $\mathbf{M}$ for T steps, starting from a given set of generators. Then it outputs a random component $g = g_i$ of the generating k-tuple X_T.

The Markov chain $\mathbf{M}$ is reversible and aperiodic, and the uniform distribution is stationary. $\mathbf{M}$ is irreducible, and therefore ergodic, if and only if it is connected. Therefore, if the chain $\mathbf{M}$ is connected, it can be used for approximate sampling from $\mathcal{N}_k(G)$.

Let $\varkappa(G)$ and $\tilde{\varkappa}(G)$ denote the smallest and the largest size, respectively, of a minimal generating set. It is conjectured that for $k \geqslant \varkappa(G) + 1$, the chain $\mathbf{M}(G, k)$ is always connected. However, this problem remains wide open.

It was shown in [CLMNO] (cf. [DsS2]) that $k \geqslant \varkappa + \tilde{\varkappa}$ suffices for connectivity of $\mathbf{M}(G, k)$. This, however, is a rather weak result because $\tilde{\varkappa}(G)$ tends to be close to $\log|G|$. It was shown in [Pa2] that if G is simple then for $k \geqslant 3$, the chain $\mathbf{M}(G, k)$ has a "giant component," comprising a $1 - o(1)$ fraction of the configuration space.

Empirical tests seem to indicate that for $k \geqslant \varkappa(G) + 1$, the chain $\mathbf{M}$ mixes rapidly [CLMNO,Le], but no results have been proved in this direction.

Observe that there can be two types of error when we try to generate a nearly uniform group element by this procedure. First, we may stop too soon (the distribution of X_T is not close to the stationary distribution on $\mathcal{N}_k(G)$); second, even the stationary distribution on $\mathcal{N}_k(G)$ may not yield (nearly) uniformly distributed elements of G.

The former problem (a problem of mixing rate) has been studied by several sets of authors (see [ChG,DsSl,DsS2,Pa2,PaB]). The breakthrough came in [Pa3], where the second named author showed a polynomial mixing time in the case $k = \Omega^*(\log|G|)$. We shall note the absence of the second type of error in this case (see [Pa2]).

While the presence of the second type of error was observed in [CLMNO], the present paper seems to be the first one to address the magnitude of this problem.

Let G be a finite group and let Q_*^k be the probability distribution of the product of all elements in a uniformly chosen generating k-tuple $(g_1, \ldots, g_k) \in \mathcal{N}_k(G)$. Let $\mathrm{Q}_\circ^k$ denote the probability distribution of the random component in a uniformly chosen element of $\mathcal{N}_k(G)$. This is the limit distribution of the algorithm output when $T \to \infty$ if $\mathbf{M}$ is ergodic.

Note that $(g_1, g_2, \ldots, g_k) \in \mathcal{N}_k(G)$ if and only if $(g_1 g_2 \cdots g_k, g_2, \ldots, g_k) \in \mathcal{N}_k(G)$. The following is now immediate and does not depend on the ergodicity of $\mathbf{M}$.

[1] Partly in reaction to the present work, Charles Leedham-Green has proposed new variants of the algorithm which avoid the nonuniform asymptotic behavior discussed in this paper [Le].

Proposition 3.1. *Let G be a finite group and $k \geqslant 1$. Then, for all $g \in G$,*

$$\mathrm{Q}_\circ^k(g) = \mathrm{Q}_*^k(g) = \mathrm{Q}^k(g).$$

Now Theorem 2.1 shows that the product replacement algorithm will not produce (nearly) uniform group elements for small values of k even if $\mathbf{M}(G,k)$ is ergodic. Indeed, take $G = A_n^{n!/8}$, $n \geqslant 5$. For $k = o(n)$ we obtain

$$\left\|\mathrm{Q}^k - \mathrm{U}\right\|_{tv} \to 1 \quad \text{as } n \to \infty.$$

It is known that G can be generated by two elements (see [Ha,KaL]). Thus, taking $k = \max\{10, \varkappa(G)\}$ as suggested in [CLMNO] will not give (nearly) uniform elements of G. Note also that it is not even clear whether the underlying graph of the Markov chain $\mathbf{M}$ is connected in this case (cf. [DsG,Pa2]).

Furthermore, Theorem 2.2 implies that the bias can be detected by a very short word (length $(\log\log|G|)^{O(k)}$). Thus, at least in theory, Monte Carlo algorithms which call a product replacement subroutine may be unreliable. Independent computer experiments by Leedham-Green and Niemeyer [Le] tend to confirm this point.

4. Statistics of element orders

A particularly important question in computational group theory is to sample the orders of elements faithfully. The authors of [CLMNO] performed a χ^2 test on the order distribution (and other characteristics) of the group elements for certain important classes of matrix groups and found no significant bias in the output of the product replacement algorithm for rather small values of T ($T < 100$ in all their examples).

Such bias, in fact, does exist. As pointed out in [CLMNO], it is obvious that the identity element is always underrepresented in generating k-tuples, and therefore Q^k can never be exactly uniform and it cannot even faithfully represent the element orders. In fact, if both $|G|$ and k are bounded, then this observation gives a constant bias against the identity, the only element of order 1. For instance, for $G = \mathbb{Z}_2 = \{0,1\}$ (0 is the "identity") it is clear that $\mathrm{Q}^k(0) = 1/2 - \delta_k$ and $\mathrm{Q}^k(1) = 1/2 + \delta_k$ where $\delta_k = 1/2(2^k - 1)$, hence the bias in this case is $\|\mathrm{Q}^k - \mathrm{U}\|_{tv} = \delta_k$.

This bias is then inherited by groups of arbitrarily large size. Indeed, let now $G = \mathbb{Z}_{2p}$, the cyclic group of order $2p$ where p is a large prime. Then G has a (unique) subgroup H of index 2. An easy calculation shows that the bias of H under Q^k is $|\mathrm{Q}^k(H) - \mathrm{U}(H)| = \delta_k + O(1/p^{k-1})$ and therefore the bias of the distribution Q^k is

$$\left\|\mathrm{Q}^k - \mathrm{U}\right\|_{tv} \geqslant \delta_k + O\bigl(1/p^{k-1}\bigr)$$

for this class of groups G.

All elements of H have odd orders and all elements of $G \setminus H$ have even orders. So the bias described is a bias in the parity of the orders of elements sampled.

We note that this bias is not due to the extreme simplicity of the structure of the groups chosen (cyclic groups); a large class of groups with an odd-order subgroup of index 2 will behave similarly. More generally, groups with a small quotient group often inherit the bias of the quotient.

While this means a constant bias for every fixed k for large classes of arbitrarily large groups, this bias is exponentially small as a function of k, and even for $k = 2$, it is only a bias of $\delta_2 = 1/6$.

In contrast, we shall show that for the groups $G = A_n^m$, the bias approaches 1 even when $k \to \infty$ not too fast $(k = o(n))$. We believe, however, that for these groups, this large bias cannot be detected by sampling element orders alone. This suggests that other statistics on groups should be tested as well.

It remains an *open problem* to decide whether or not the distribution Q^k produces a similarly overwhelming bias in the statistics of element orders for some class of groups (see [Pa2]). Let us note here that the sequence of powers of A_n^m we consider in this paper cannot be used for this purpose. Indeed, a random element $G = A_n^m$, with $m = \Omega(n^2)$, is exponentially likely to have an order $N = N(m,n)$. Thus, almost all elements in G have the same order N in this case (see [Pa2, Proposition 1.4.1].

5. Direct product of groups

Let H be a simple nonabelian group, and let $G = H^m$. Denote by $d_k(H)$ the maximal power m such that H^m is generated by k elements. In [Ha] Hall showed that

$$d_k(H) = \frac{N_k(H)}{|\mathrm{Aut}(H)|}$$

(see also [KaL]). The right-hand side can be interpreted as the number of orbits of the diagonal action of $\mathrm{Aut}(H)$ on $\mathcal{N}_k(H)$. For A_5, Hall found that $d_2(A_5) = 19$.

Now let us take a close look at the structure of $\mathcal{N}_k(G)$, where $G = H^m$. Denote by $(g_1, \ldots, g_k)$ the elements of $\mathcal{N}_k(G)$, and let $g_i = (h_1^{(i)}, \ldots, h_m^{(i)})$, where $h_j^{(i)} \in A_n$, $1 \leqslant i \leqslant k$, $1 \leqslant j \leqslant m$. Observe that in order for the elements $g_1, \ldots, g_k$ to generate G, the elements $h_j^{(1)}, \ldots, h_j^{(k)}$ must generate H for all j. Note, however, that these m generating k-tuples cannot be fully independent. Indeed, these k-tuples correspond to a generating set if and only if they lie in different orbits of the diagonal action of $\mathrm{Aut}(H)$ on $\mathcal{N}_k(H)$ (see [Ha,KaL]). If the number of orbits is very large, the probability that two generating k-tuples lie in the same orbit becomes negligible and we can treat them as independent. Below we give a formal meaning to this observation.

Observe that a birthday paradox type of argument gives us the following formula for the proportion of generating k-tuples of G (see [KaL]):

$$\varphi_k(G) = \big(\varphi_k(H)\big)^m \prod_{i=1}^{m-1} \left(1 - \frac{i}{d_k(H)}\right).$$

For simple groups H it is known that

$$\varphi_2(H) \to 1 \quad \text{as } |H| \to \infty.$$

For the family of alternating groups A_n this is a celebrated result of Dixon [Dx], for classical simple groups of Lie type this is due to Kantor and Lubotzky [KaL], and in full generality it was recently proved by Liebeck and Shalev [LiSl].

When $H = A_n$ we conclude that $\varphi_2 \geqslant 1/2$ when n is large enough. Therefore,

$$d_k(A_n) = \frac{N_k(A_n)}{|S_n|} \geqslant \frac{(n!/2)^k/2}{n!} = \frac{(n!/2)^{k-1}}{4},$$

where $k \geqslant 2$. In particular, $d_2(A_n) \geqslant n!/8$.

Now let $m \leqslant n!/8$, $G = A_n^m$. Then

$$\prod_{i=1}^{m-1} \left(1 - \frac{i}{d_k(A_n)}\right) \geqslant \left(1 - \frac{m}{d_k(A_n)}\right)^m = 1 + O\left(\frac{1}{(n!/2)^{k-3}}\right).$$

We conclude:

Lemma 5.1. *If $k \geqslant 4$ and $m \leqslant n!/8$ then*

$$\varphi_k(G) = \bigl(\varphi_k(A_n)\bigr)^m \cdot \bigl(1 + O(1/n!)\bigr). \tag{1}$$

Equation (1) says that the relative size of the difference of these two sets is $O(1/n!)$ and therefore it can be ignored in most calculations. The following is immediate:

Corollary 5.2. *Let $\mathcal{E} \subseteq (\mathcal{N}_k(A_n))^m$, and $G = A_n^m$. Then*

$$\bigl|\mathbf{P}\bigl(\mathcal{E}, \mathcal{N}_k(G)\bigr) - \mathbf{P}(\mathcal{E})\bigr| = O(1/n!).$$

(*The probabilities refer to uniform choice from $(\mathcal{N}_k(A_n))^m$.*)

Remark 5.3. Let $\sigma^{(j)} = (\sigma_1^{(j)}, \ldots, \sigma_k^{(j)})$ denote elements of $\mathcal{N}_k(A_n)$ $(1 \leqslant j \leqslant m)$. Corollary 5.2 means that for most calculations, we can treat the components $\sigma^{(j)}$ of a uniform random element $(\sigma^{(1)}, \ldots, \sigma^{(m)}) \in \mathcal{N}_k(G)$ as independent; for $k \geqslant 4$ and $m \leqslant n!/8$, the error will be $O(1/n!)$.

We remark that our work on the alternating group A_n can be extended to other classes of finite simple groups, see Section 11 (cf. [KaL,LiS2,Sh1,Sh2]).

6. Distribution of generating k-tuples in A_n

In this section we obtain rather accurate asymptotic estimates on the probability that generating k-tuples in A_n satisfy certain conditions.

First, we obtain bounds on the asymptotic behavior of $N_k(A_n)$ as $n \to \infty$.

Denote by $x = (\sigma_1, \ldots, \sigma_k)$ a uniformly distributed element in A_n^k. Let $\mathcal{A}$ denote the event that $x \in \mathcal{N}_k(A_n)$.

Proposition 6.1. *For $k \geqslant 2$ we have*

$$\mathbf{P}(\mathcal{A}) = 1 - \frac{1}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right).$$

Proof. As above, let $x = (\sigma_1, \dots, \sigma_k) \in A_k^n$ be chosen uniformly. The idea is to show that the most frequent reason for $\sigma_1, \dots, \sigma_k$ not to generate A_n is that all σ_i share a common fixed point.

The probability that each σ_i lies in a maximal subgroup which is not of the form $(S_r \times S_{n-r}) \cap A_n$ is at most c^n where $c = 2^{-1/4} + o(1)$ (so $c < 0.841$ for large n) ([Bb2, Dx], cf. [Sh1,Sh2]). Thus, for $k \geqslant 2$ we have

$$\begin{aligned}\mathbf{P}(\mathcal{A}) &= 1 - n \cdot \bigl(\mathbf{P}\bigl(\sigma(1) = 1\bigr)\bigr)^k + \binom{n}{2} \cdot \bigl(\mathbf{P}\bigl(\sigma(1) = 1, \sigma(2) = 2\bigr)\bigr)^k - \cdots \\ &\quad - \binom{n}{2} \cdot \bigl(\mathbf{P}\bigl(\sigma(1) = 2, \sigma(2) = 1\bigr)\bigr)^k + \cdots \\ &= 1 - n \cdot \frac{1}{n^k} + \binom{n}{2} \cdot \frac{1}{(n(n-1))^k} - \cdots - \binom{n}{2} \cdot \frac{1}{(n(n-1))^k} + \cdots \\ &= 1 - \frac{1}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right). \qquad \square\end{aligned}$$

Let $\mathcal{B}$ denote the event that $\sigma_k(1) = 1$. Clearly, $\mathbf{P}(\mathcal{B}) = 1/n$.

Proposition 6.2. *For $k \geqslant 2$ we have*

$$\mathbf{P}(\mathcal{B} \mid \mathcal{A}) = \frac{1}{n} - \frac{1}{n^k} + O\left(\frac{1}{n^{2k-1}}\right).$$

Proof. For an illustration, we include the proof in some detail. We have

$$\mathbf{P}(\mathcal{B} \mid \mathcal{A}) = \frac{\mathbf{P}(\mathcal{A} \mid \mathcal{B})\mathbf{P}(\mathcal{B})}{\mathbf{P}(\mathcal{A})} = \frac{1}{n} \cdot \frac{\mathbf{P}(\mathcal{A} \mid \mathcal{B})}{\mathbf{P}(\mathcal{A})}.$$

We estimate the conditional probability $\mathbf{P}(\mathcal{A} \mid \mathcal{B})$ similarly to the estimation of $\mathbf{P}(\mathcal{A})$. Again, we only need to worry about maximal subgroups of the form $S_r \times S_{n-r} \cap A_n$. We obtain

$$\begin{aligned}\mathbf{P}(\mathcal{A} \mid \mathcal{B}) &= 1 - \bigl(\mathbf{P}\bigl(\sigma(1) = 1\bigr)\bigr)^{k-1} - (n-1) \\ &\qquad \times \bigl(\mathbf{P}\bigl(\sigma(2) = 2\bigr)\bigr)^{k-1} \cdot \mathbf{P}\bigl(\sigma(2) = 2 \mid \sigma(1) = 1\bigr) + \cdots \\ &= 1 - \frac{1}{n^{k-1}} - (n-1) \cdot \frac{1}{(n-1)n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right) \\ &= 1 - \frac{2}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right).\end{aligned}$$

We conclude that

$$\mathbf{P}(\mathcal{B} \mid \mathcal{A}) = \frac{1}{n} \cdot \frac{1 - 2/n^{k-1} + O(1/n^{2k-2})}{1 - 1/n^{k-1} + O(1/n^{2k-2})} = \frac{1}{n} - \frac{1}{n^k} + O\left(\frac{1}{n^{2k-1}}\right).$$

This completes the proof. □

Let $\mathcal{C}$ denote the event that $\sigma_k(i) \neq i$ for all i (σ_k is a "derangement"). It is well known that $\mathbf{P}(\mathcal{C}) = 1/e + o(1/n!)$.

Proposition 6.3. *For $k \geqslant 2$ we have*

$$\mathbf{P}(\mathcal{C} \mid \mathcal{A}) = \frac{1}{e} + \frac{1}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right).$$

Proof. By analogy with the proof of Proposition 6.2, it suffices to prove that

$$\mathbf{P}\bigl(\bar{\mathcal{A}} \mid \mathcal{C}\bigr) = O\bigl(n^{-2k+2}\bigr),$$

where $\bar{\mathcal{A}}$ is the complement of the event $\mathcal{A}$. Observe that in this case the smallest index maximal subgroup which might include σ_k is $(S_2 \times S_{n-2}) \cap A_n$. Therefore,

$$\mathbf{P}\bigl(\bar{\mathcal{A}} \mid \mathcal{C}\bigr) \leqslant \mathbf{P}\bigl(\bar{\mathcal{A}} \wedge \mathcal{C}\bigr) \leqslant \binom{n}{2} \cdot \bigl(\mathbf{P}\bigl(\sigma(1) = 2, \sigma(2) = 1\bigr)\bigr)^k + \cdots = O\left(\frac{1}{n^{2k-2}}\right). \qquad \square$$

Let $\mathcal{D}$ denote the event that σ_k is *a long cycle*, i.e., a cycle of length n. Clearly, $\mathbf{P}(\mathcal{D}) = 1/n$.

Proposition 6.4. *For $k \geqslant 2$ we have*

$$\mathbf{P}(\mathcal{D} \mid \mathcal{A}) = \frac{1}{n} + \frac{1}{n^k} + O\left(\frac{1}{n^{2k-1}}\right).$$

Proof. The proof is analogous to the preceding one except that in this case the probability $\mathbf{P}(\mathcal{A} \mid \mathcal{D}) = 1 - O(c^n)$ by the observations above. We omit the details. □

7. Proof of Theorem 2.1

Let us recall some standard probability estimates for large deviations. Let $\zeta_1, \ldots, \zeta_m$ be independent random $(0,1)$-variables (Bernoulli trials) with $\mathbf{P}(\zeta_i = 1) = p$, $\mathbf{P}(\zeta_i = 0) = 1 - p$. Let $\xi = \zeta_1 + \cdots + \zeta_m$. We have $E(\xi) = p \cdot m$. For $p \leqslant 1/2$ and $a > 0$, the Chernoff bounds, as stated by Alon and Spencer [AS, Theorems A.11 and A.13, pp. 237–238], give us

$$\mathbf{P}(\xi > pm + a) < \exp\bigl(-2a^2/pm + 4a^3/(pm)^2\bigr), \tag{2}$$

and

$$\mathbf{P}(\xi < pm - a) < \exp\bigl(-2a^2/pm\bigr). \tag{3}$$

Let $B \subset A_n^m$ be the set of all elements $g = (\sigma_1, \ldots, \sigma_m)$ such that

$$\#\bigl\{j \mid \sigma_j(1) = 1,\ 1 \leqslant j \leqslant m\bigr\} > m \cdot \left(\frac{1}{n} - \frac{1}{2n^k}\right). \tag{4}$$

Lemma 7.1. *Under the conditions of Theorem* 2.1,

$$\mathrm{Q}^k(B) \to 0 \quad \textit{and} \quad \mathrm{U}(B) \to 1 \quad \textit{as } n \to \infty,$$

assuming $k = o(n)$.

Proof. For the proof we note that the quantity on the right-hand side of inequality (4) is halfway between the expected value of the left hand side under uniform distribution (m/n) and under $\mathrm{Q}^k(m(1/n - 1/n^k))$. Now both parts of the claim follow from Chernoff's bounds. Indeed, using the Chernoff bound (3) with $p = 1/n$, $a = m/(2n^{k-1})$, we obtain

$$\begin{aligned}\mathrm{U}(B) &= \mathbf{P}\left(\#\{j \mid \sigma_j(1) = 1,\ 1 \leqslant j \leqslant m\} > \frac{m}{n} \cdot \left(1 - \frac{1}{2n^{k-1}}\right)\right) \\ &> 1 - \exp\left(\frac{-m}{2n(2n^{k-1})^2}\right) = 1 - \exp\bigl(-m/\bigl(8n^{2k-1}\bigr)\bigr).\end{aligned}$$

By definition $m = n!/8$. Thus, when $k = o(n)$ we have $\mathrm{U}(B) \to 1$ as $n \to \infty$. This proves the second part of the lemma.

The first part goes analogously, except that in this case by Proposition 6.2 we have $\mathbf{P}(\sigma(1) = 1) = 1/n - 1/n^k + O(n^{1-2k})$. By Corollary 5.2 and Remark 5.3 we may assume that the events $\sigma_j(1) = 1$ are independent; the error thus made is negligible $(O(1/n!))$. Now take a as above and use the Chernoff bound (2). □

Observe that Lemma 7.1 immediately implies Theorem 2.1. Indeed,

$$\bigl\|\mathrm{Q}^k - \mathrm{U}\bigr\|_{tv} \leqslant \bigl|\mathrm{Q}^k(B) - \mathrm{U}(B)\bigr| \to 1 \quad \text{as } n \to \infty.$$

Remark 7.2. We could have used either of the events $\mathcal{C}$ or $\mathcal{D}$ instead of $\mathcal{B}$ in the lemma. The event $\mathcal{C}$ has the advantage that it gives a smaller bound on m. The event $\mathcal{D}$ gives the group theoretically significant additional feature that B becomes a union of conjugacy classes. We shall exploit this subtle difference in the next section.

8. Biased events

Let $z_1, \ldots, z_s$ be Boolean variables. Let $\mathrm{Th}_{s,t}(z_1, \ldots, z_s)$ denote the *threshold function* which takes value 1 if $\sum_{i=1}^{s} z_i \geqslant t$ and 0 otherwise. We use this function to separate statistically the distributions Q^k and U over G.

Let n be a prime number. Consider uniform samples of $x = (\sigma_1, \ldots, \sigma_k) \in A_n^k$. Using the notation of Section 6, let $\mathcal{D}$ be the event that the permutation $\sigma_k \in A_n$ is a long cycle. By $\mathcal{D}'$ we denote the event that $(\sigma_k)^n = 1$. We have

$$\mathbf{P}\bigl(\mathcal{D}'\bigr) = \frac{1}{n} + \frac{1}{n!}.$$

On the other hand, when $k = o(n)$, Proposition 6.4 gives us

$$\mathbf{P}\bigl(\mathcal{D}' \mid \mathcal{A}\bigr) = \frac{1}{n} + \frac{1}{n^k} + O\left(\frac{1}{n^{2k-1}}\right).$$

Let us now take s independent samples $x_1, \ldots, x_s \in A_n^k$, and let $\mathcal{D}_i$ denote the event $\mathcal{D}'$ with respect to x_i. We view the $\mathcal{D}_i$ as random $(0,1)$-variables. Then

$$E\left(\sum_{i=1}^{s} \mathcal{D}_i\right) = \frac{s}{n} + \frac{s}{n!};$$

whereas

$$E\left(\sum_{i=1}^{s}\mathcal{D}_i \,\middle|\, \mathcal{A}\right)=\frac{s}{n}+\frac{s}{n^k}+O\left(\frac{s}{n^{2k-1}}\right).$$

We set $s = 2n^{4k}$ and apply Chernoff's bounds with threshold $t = s/n + s/2n^k$. This value of t is halfway between the two expected values above for the given s. It follows that

$$\mathbf{P}\big(\mathrm{Th}_{s,t}(\mathcal{D}_1,\dots,\mathcal{D}_s)\big) < \exp\big(-n^k\big(1-1/4n^k\big)\big) < n^{-cn},$$

while

$$\mathbf{P}\big(\mathrm{Th}_{s,t}(\mathcal{D}_1,\dots,\mathcal{D}_s) \mid \mathcal{A}\big) > 1-\exp\big(-n^k\big)+O(1/n!) > 1-n^{-cn}.$$

Now we express the threshold function by a monotone Boolean circuit with suitable parameters. From the various options [Al,A2,Va], our choice is to use the Ajtai–Komlós–Szemerédi sorting network [AKS]. It is immediate that the AKS sorting network [AKS] can be turned into a monotone Boolean circuit with fan-in 2 gates for the threshold function $\mathrm{Th}_{s,t}$; the circuit will have depth $O(\log s)$ and width (maximum number of nodes per level) s.

Thus, we have proved the following result.

Proposition 8.1. *Give n, k there exists an explicit monotone fan-in 2 Boolean circuit $\mathbf{F}_{n,k}$ of size $\leqslant n^{O(k)}$ and depth $O(k\log n)$ with $s = 2n^{4k}$ input variables such that, assuming $k = k(n) = o(n)$, $k \geqslant 4$, we have $\mathbf{P}(\mathbf{F}) = 1 - O(n^{-cn})$ and $\mathbf{P}(\mathbf{F} \mid \mathcal{A}) = O(n^{-cn})$, where $\mathbf{F} = \mathbf{F}_{n,k}(\mathcal{D}_1,\dots,\mathcal{D}_s)$.*

9. Simulation of monotone Boolean operations

In this section we turn the Boolean circuit of the preceding section into a short word in the group G. The basic idea was inspired by Barrington's simulation of Boolean operations by group operations [Bar], although the actual details and the scope are quite different. In particular, in our context, negation cannot be simulated; and our simulation is (necessarily) randomized.

Let H be a group and $g \in H$. We consider the predicate $\mathcal{E}(g)$ meaning "$g = 1$." We wish to construct words w_1 and w_2 corresponding to the predicates $\mathcal{E}_1(g,h) = \mathcal{E}(g) \wedge \mathcal{E}(h)$ and $\mathcal{E}_2(g,h) = \mathcal{E}(g) \vee \mathcal{E}(h)$, respectively. Clearly, there is no word which would be 1 exactly if $\mathcal{E}_1$ holds, nor is there one for $\mathcal{E}_2$. But the product $w_1 = gh$ and the commutator $w_2 = [g,h] = g^{-1}h^{-1}gh$ go part of the way: $\mathcal{E}_1(g,h)$ implies $w_1 = 1$ and $\mathcal{E}_2$ implies $w_2 = 1$; and the converse holds often enough in each case. We shall formalize this last observation.

Lemma 9.1. *Given $n \geqslant 5$, there exist words w_1 and w_2 of length $O(n^2\log n)$ in $O(n\log n)$ variables $g, h, u_1, u_2, \dots$ such that for every $g, h \in A_n$,*

(a1) *if $g = h = 1$ then $w_1 = 1$ (regardless of the values u_i);*

(a2) *if $g = 1$ or $h = 1$ then $w_2 = 1$ (regardless of the values u_i);*
(b1) *if $g \neq 1$ or $h \neq 1$ and if the u_i are independent uniformly distributed random elements of A_n then the event $w_1 = 1$ is factorially unlikely;*
(b2) *if $g \neq 1$ and $h \neq 1$ and if the u_i are independent uniformly distributed random elements of A_n then the event $w_2 = 1$ is factorially unlikely.*

First, let us consider the word

$$z = \bigl(u_1^{-1} g u_1\bigr) \cdot \cdots \cdot \bigl(u_N^{-1} g u_N\bigr) \tag{5}$$

over a finite group H. For a fixed $g \in A_n$ and randomly chosen u_i one can think of z as the Nth state of a random walk on H generated by the conjugates of g.

Lemma 9.2. *Fix $g \in A_n$, $g \neq 1$. Let $N = \Omega(n^2 \log^2 n)$ and define z by Eq.* (5). *If the u_i are independent, uniformly distributed elements from A_n then*

$$\left|\mathbf{P}(z = h) - \frac{1}{|A_n|}\right| < \frac{1}{2|A_n|} \quad \textit{for all } h \in A_n.$$

Proof. Let R^N be the probability distribution of the element $z \in A_n$. This is the result of N steps of the random walk on a Cayley graph defined by a conjugacy class as the set of generators. This situation was considered by Roichman [Ro]; it follows from his results that

$$\bigl\|\mathrm{R}^N - \mathrm{U}\bigr\|_{tv} < c_1,$$

where $1 > c_1 > 0$, $N = cn \log n$, c, c_1 are universal constants.[2] Now use a standard bound which relates mixing in relative pointwise distance (or ℓ_∞ distance) to mixing in variation distance (see, e.g., [AF,LoW]). This implies that after $N' = \Omega(N \log |A_n|) = \Omega(n^2 \log^2 n)$ steps we obtain the inequality stated. □

Now we turn to the proof of Lemma 9.1. For $g \in H$, consider the word $z(g) = z(g, u_1, \ldots, u_n)$. For $h \in H$, consider the word $z(h) = z(h, u_{N+1}, \ldots, u_{2N})$.

Let now $w_1(g,h) = z(g) \cdot z(h)$ and $w_2(g,h) = [z(g), z(h)]$. It is obvious that these choices satisfy parts (a1) and (a2) of Lemma 9.1.

For the proof of (b1), there are two more cases to consider. If exactly one of g, h is 1, then $z(g) \cdot z(h)$ is nearly uniform over A_n and therefore factorially unlikely to be 1. If neither g, nor h is 1 then

$$\mathbf{P}\bigl(z(g) \cdot z(h) = 1\bigr) = \sum_{f \in A_n} \mathbf{P}\bigl(z(g) = f\bigr) \cdot \mathbf{P}\bigl(z(h) = f^{-1}\bigr) \leqslant |A_n| \cdot \left(\frac{3/2}{|A_n|}\right)^2.$$

We conclude that w_1 is factorially unlikely to be 1 when $g, h \neq 1$.

For (b2), the only case to consider is when $g, h \neq 1$. In this case,

[2] This result seems to have been known before [Ro]; it follows from the character bounds in an unpublished manuscript [CaH].

$$\mathbf{P}\big(\big[z(g), z(h)\big] = 1\big) = \sum_{v_1, v_2 \in A_n, [v_1, v_2]=1} \mathbf{P}\big(z(g) = v_1\big)\mathbf{P}\big(z(h) = v_2\big)$$
$$\leqslant r(A_n) \cdot \left(\frac{3/2}{|A_n|}\right)^2,$$

where $r(H)$ is the number of solutions of the equation $[v_1, v_2] = 1$ in the group H. Denote by $\eta(H)$ the number of conjugacy classes in H. Frobenius observed [Fr] that $r(H) = |H| \cdot \eta(H)$. The number of conjugacy classes of A_n is bounded by 2 times the number of partitions of the integer n and therefore $\eta(A_n) = O(e^{c\sqrt{n}})$. We conclude that

$$\mathbf{P}\big(\big[z(g), z(h)\big] = 1\big) = O\big(n^{-cn}\big). \qquad \square$$

10. Proof of Theorem 2.2

Now we can put together the results of the previous sections.

Let $k > 4$ be any large constant or any function of n such that $k(n) = o(n)$. Now let $G = A_n^m$, where n is a sufficiently large prime and let $m = m(n)$ grow faster than n^{ck} but slower than n^{cn} for all $c > 0$. Therefore, $m(n) \sim e^{n \cdot \omega(n)}$ as in Theorem 2.2 will work.

Now fix n. Consider independent samples from Q^k, i.e., samples obtained by projection of $\mathcal{N}_k(G)$ onto the first components g_i in generating k-tuples. Consider the Boolean circuit $\mathbf{F}$ given in Proposition 8.1. Substitute the expression x_i^n for the ith Boolean input variable. Substitute the words w_1, w_2 given in Section 9 for the Boolean operations to evaluate the circuit. Let $\mathbf{w}$ be the resulting output word. This is the word we will use to prove Theorem 2.2.

We claim that $\mathbf{P}_{\mathrm{U}}(\mathbf{w} = 1) = O(n^{-cn})$ (we substitute independent, uniformly distributed random members of G for the variables in $\mathbf{w}$). Indeed, Lemma 9.1 implies that the error in the Boolean operations in factorially small. The number of Boolean operations in $\mathbf{F}$ is $n^{O(k)} = n^{o(n)}$ so even the total error probability is factorially small. This and Proposition 8.1 imply that it is factorially likely that none of the components $\sigma_i \in A_n$ of $\mathbf{w} = (\sigma_1, \ldots, \sigma_m)$ is the identity (which is far more than what we need).

On the other hand, we claim that $\mathbf{P}_{\mathrm{Q}^k}(\mathbf{w} = 1) = 1 - O(n^{-cn})$. Again, let $\mathbf{w} = (\sigma_1, \ldots, \sigma_m)$ $(\sigma_i \in A_n)$. As before, we make only a factorially small error by assuming that the σ_i are independently chosen from the distribution $\mathrm{Q}^k(A_n)$.

Under this assumption, it is factorially likely that $\sigma_i = 1$ for any fixed i. This is immediate from Proposition 8.1 and the observation that the error made in the group-theoretic simulation of monotone circuits is one-way: if a gate outputs 1 then necessarily the simulating group element is the identity. This follows from properties (a1) and (a2) listed in Lemma 9.1.

Finally, $m = n^{o(n)}$; therefore, it is factorially likely that *all* components of $\mathbf{w}$ are 1.

It is easy to see that the length of the word $\mathbf{w}$ is $n^{O(k)}$. First of all, this is obvious if we allow the commutator to be an operation. Now the increase due to expanding the commutators is a factor of 4^d where d is the depth of the circuit. Since $d = O(k \log n)$, the bound on the length of $\mathbf{w}$ follows. □

11. Direct products of simple groups of Lie type

It turns out that virtually all the results (including Theorems 2.1, 2.2) have analogs for products of large simple groups of Lie type. In this section we state the main results and sketch the main steps of the proof, while omitting most details.

Let $H_n(q)$ be a family of simple groups of Lie type, where q is fixed, $n \to \infty$. Everywhere below $q = p^r$ stands for the size of the finite field, and n for the Lie rank of the group.

Theorem 11.1. *Let $G_n = H_n^m$ be the family of powers of simple groups, where $m = m(n) = |H_n|/2$. Let Q_n^k be the probability distribution of the first component of $\mathcal{N}_k(G_n)$. Then*

$$\left\|\mathrm{Q}_n^k - \mathrm{U}\right\|_{tv} \to 1 \quad \text{as } |H_n| \to \infty$$

given $k = o(n)$.

It is known, on the other hand, that the groups G_n can be generated by only two elements when $|H_n|$ is large enough (see [Ha,KaL,Pal]). Therefore, we again obtain a strong bias in the probability distribution of the output of the product replacement algorithm.

Now recall classification of finite simple nonabelian groups. There are only six series where n grows: $A_n(q)$, ${}^2A_n(q)$, $B_n(q)$, $C_n(q)$, $D_n(q)$, and ${}^2D_n(q)$ (see, e.g., [Go, CCNPW]). Much is known about these series, including a number of probabilistic results (see [Sh1,Sh2]). Nevertheless, some additional group theoretic work has to be done, in order to obtain the analogs of the results in the previous chapters.

Rather than give a number of known technical details about the structure of maximal subgroups in the above series, we will present a somewhat shortened proof only for a series $A_n(q) = \mathrm{PSL}(n, q)$, while omitting details in other cases. However, we will stress the key points in full generality, so that the interested reader can reconstruct the whole proof.

Sketch of proof. First, we need analogs for $\mathrm{PSL}(n, q)$ (and other simple groups of Lie type) of the results which were already established for alternating groups.

First, recall that a random pair of elements generate a simple group $H = H_n(q)$ of Lie type with probability $\to 1$ as $|H| \to \infty$ [KaL,LiSl] (see [Sh1]). It was shown there that

$$\varphi_2(H) = 1 - O\left(\frac{n^3 \log^2(q)}{q^n}\right).$$

Further, if $H_n = \mathrm{PSL}_n(q)$, we have

$$\varphi_k(H) = 1 - \left(\frac{1}{(n)_q^{k-1}}\right) + O\left(\frac{1}{q^{(n-1)(2k-1)}}\right),$$

where $(n)_q = (q^n - 1)/(q - 1)$. By abuse of notation, here and later in the probabilistic estimates, we assume that the constant implied by $O(\cdot)$ notation depends polynomially of n and $\log q$. In these cases the denominator will always grow exponentially, so this will make no difference to the final results.

The above estimate, while stated explicitly in the literature only for $k = 2$, follows immediately from the analysis in [KaL,LiSl,LiS2]. The proof idea is as follows. From

Aschbacher classification [As] of maximal subgroups, one knows that the subgroups of the smallest index (equal to $(n)_q$) are isomorphic to $\mathrm{PSL}(n-1,q)$, there are $(n)_q$ of them, while the remaining maximal subgroups have a much larger index (and there are not too many of them). Now proceed as in the proof of Proposition 6.1.

Let us note here that for other series one has to replace denominator $(n)_q$ by the smallest index of the proper subgroup (ibid).

The following analog of Lemma 5.1 was obtained in [Pal] for any series H_n of all simple groups of Lie type.

Proposition 11.2. *Let $G = H_n^m$, $|H_n| \to \infty$ as $n \to \infty$. If $k \geqslant 4$ and $m \leqslant |H_n|$ then*

$$\varphi_k(G) = \big(\varphi_k(H_n)\big)^m \cdot \big(1 + O\big(1/|H_n|\big)\big). \tag{6}$$

Now, let $H_n = \mathrm{PSL}(n,q)$, where q is a fixed prime. Let $x = (h_1,\dots,h_k)$ be chosen uniformly in H_n^k. As in Section 6, let $\mathcal{A}$ denote the event that $x \in \mathcal{N}_k(H_n)$. Note that there is a natural embedding $H_{n-1} \hookrightarrow H_n$. Let $\mathcal{B}$ denote the event that $h_1^{(1)} \in H_{n-1}$. For $k \geqslant 2$ we have

$$\mathbf{P}(\mathcal{B} \mid \mathcal{A}) = \frac{1}{(n)_q} - \left(\frac{1}{(n)_q}\right)^k + O\left(\frac{1}{q^{(n-1)(2k)}}\right).$$

This is a direct analog of Proposition 6.2, and the proof follows verbatim. Indeed, in this case the subspace stabilizing subgroups of $\mathrm{PSL}(n,q)$ play a role of the point stabilizers in A_n. The rest is the same, with a substitution of all (n)'s by their q-analogs $(n)_q$.

Finally, as we remarked earlier, H_n^m is 2-generated for large n, given that $m \leqslant |H_n|/2$ (this follows from Hall's theorem and $\varphi_2(H_n) \to 1$ as $n \to \infty$). The bias becomes significant then given the denominator $(n)_q^k = o(|H_n|)$, which is implied by $k = o(n)$ (for $|H_n| = q^{\theta(n^2)}$). One uses Chernoff bound as in the proof of Lemma 9.2 and obtains the result. We omit the easy details. □

Let us finish by presenting an analog of Theorem 2.2 for all simple groups of Lie type. Indeed, let $H_n(q)$ be as above, a series of simple groups of Lie type with a fixed q and $n \to \infty$. Let Q_n^k be as above.

Theorem 11.3. *There exists a family of words $\mathbf{w}_{n,k}$ with the following properties. The length of $\mathbf{w}_{n,k}$ is $q^{O(nk)}$. Let $\omega(n) \to \infty$, $\omega(n) = o(n)$. Also, let $k = k(n) \geqslant 4$ and $k = o(n)$. Set $m = q^{k\omega(n)}$. Let $G = A_n^m$. Then $\mathbf{w}[\mathrm{Q}^k] = 1$ has probability $1 - O(q^{-cn^2})$, while $\mathbf{w}[\mathrm{U}] = 1$ has probability $O(q^{-cn^2})$.*

Sketch of proof. Again we restrict ourselves to the case $H_n = \mathrm{PSL}(n,q)$. The remaining series are largely similar. Let n be a prime number.

First, we need describe the analog of the event $\mathcal{D}'$ in Section 8. Consider elements belong to large conjugacy class, known as Singer cycle. They have the order $q^n - 1$. While the order does not completely characterize these elements, the presence of Zsigmondy primes (and n, q being prime) ensures that there are many Singer cycles (namely, $\theta(1/n)$) while only $O(1/q^n)$ other elements have order dividing $q^n - 1$. This approach has been extensively used in the literature on recognition of black box simple groups of Lie type, notably in [KaS,BKPS].

After this point, the analog of the rest of Section 8 follows similar steps. To obtain an analog of Lemma 9.2 we need to bound the mixing time of random walks on Cayley graphs generated by conjugacy classes (see [Ds]). This can be done in several ways, e.g., by combining general bounds in [AF,Lo] with the diameter bounds in [LaL]. The best general bounds of the order $O(n)$ were recently obtained in [LiS3].

Finally, we need to obtain an estimate on the probability that a random pair of elements of G commute. This follows from Frobenius' formula an upper bound on the number of conjugacy classes (see, e.g., [Sh1], or [Gl] when H is classical). Together this gives an analog of the results in Section 9 for simulation of Boolean operations on H_n. The remainder of the proof goes exactly as in Section 10. We omit the details. □

Acknowledgments

We thank Persi Diaconis, Walter Feit, Bill Kantor, Charles Leedham-Green, László Lovász, Alexander Lubotzky, Scott Murray, Alice Niemeyer, Aner Shalev, and Leonard Soicher for helpful conversations. Theorem 2.1 was initially announced in a joint paper [PaB] of Sergey Bratus and the second named author. Nathan Lulov pointed out reference [CaH]. The second named author was supported by NSF Postdoctoral Research Fellowship in Mathematical Sciences. Both authors acknowledge partial support by the NSF. The second named author also thanks the University of Chicago for the hospitality during his brief visit in March 1999.

References

[Al] M. Ajtai, Isomorphism and higher order equivalence, Ann. Math. Logic 16 (1979) 181–203.

[A2] M. Ajtai, $\sum_1^1$-formulae on finite structures, Ann. Pure Appl. Logic 24 (1983) 1–48.

[AKS] M. Ajtai, J. Komlós, E. Szemerédi, Sorting in $c \log n$ parallel steps, Combinatorica 3 (1983) 1–19.

[AF] D.J. Aldous, J. Fill, Time-reversible Markov chains and random walks on graphs, in preparation.

[AS] N. Alon, J.H. Spencer, The Probabilistic Method, Wiley, New York, 1992.

[As] M. Aschbacher, On the maximal subgroups of the finite classical groups, Invent. Math. 76 (1984) 469–514.

[Bbl] L. Babai, Local expansion of vertex-transitive graphs and random generation in finite groups, in: Proc. 23rd ACM STOC, 1991, pp. 164–174.

[Bb2] L. Babai, The probability of generating the symmetric group, J. Combin. Theory Ser. A 52 (1989) 148–153.

[Bb3] L. Babai, Randomization in group algorithms: conceptual questions, in: L. Finkelstein, W.M. Kantor (Eds.), Groups and Computation II, in: DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 28, Amer. Math. Soc., 1997, pp. 1–16.

[BKPS] L. Babai, W.M. Kantor, P.P. Pálfy, Á. Seress, Black-box recognition of finite simple groups of Lie type by statistics of element orders, in preparation.

[Bar] D.A. Barrington, Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1, J. Comput. System Sci. 38 (1989) 150–164.

[BoC] W. Bosma, J.J. Cannon, Handbook of Magma Functions, School of Mathematics and Statistics, University of Sydney, Sydney, 1997.

[CaH] A.R. Calderbank, P. Hanlon, A ratio of character values arising in the analysis of random shuffles, unpublished manuscript, circa 1985.

[CLMNO] F. Celler, C.R. Leedham-Green, S. Murray, A. Niemeyer, E.A. O'Brien, Generating random elements of a finite group, Comm. Algebra 23 (1995) 4931–4948.

[ChG] F.R.K. Chung, R.L. Graham, Random walks on generating sets for finite groups, Electron. J. Combin. 4/2 (1997), #R7.

[CCNPW] J.H. Conway, R.T. Curtis, S.P. Norton, R.A. Parker, R.A. Wilson, Atlas of Finite Simple Groups, Clarendon, Oxford, 1985.

[Ds] P. Diaconis, Group Representations in Probability and Statistics, in: Lecture Notes Monogr. Ser., vol. 11, IMS, Hayward, CA, 1988.

[DsG] P. Diaconis, R.L. Graham, The graph of generating sets of an abelian group, Colloq. Math. 80 (1999) 31–38.

[DsSl] P. Diaconis, L. Saloff-Coste, Walks on generating sets of groups, Probab. Theory Related Fields 105 (1996) 393–421.

[DsS2] P. Diaconis, L. Saloff-Coste, Walks on generating sets of abelian groups, Invent. Math. 134 (1998) 251–299.

[Dx] J.D. Dixon, The probability of generating the symmetric group, Math. Z. 110 (1969) 199–205.

[Fr] F.G. Frobenius, Über Gruppencharaktere, Sitzungsber. Berl. Akad. (1896) 985–1021.

[Gl] D. Gluck, Characters and random walks on finite classical groups, Adv. Math. 129 (1997) 46–72.

[Go] D. Gorenstein, Finite Simple Groups, Plenum, New York, 1982.

[Ha] P. Hall, The Eulerian functions of a group, Quart. J. Math. 7 (1936) 134–151.

[Ka] W.M. Kantor, Simple groups in computational group theory, in: Proc. Int. Congress of Math., in: Doc. Math., vol. II, Berlin, 1998, pp. 77–86, http://www.mathematik.uni-bielefeld.de/documenta/xvol-icm/ICM.html.

[KaL] W.M. Kantor, A. Lubotzky, The probability of generating a finite classical group, Geom. Dedicata 36 (1990) 67–87.

[KaS] W.M. Kantor, A. Seress, Black box classical groups, in: Mem. Amer. Math. Soc., 1999, in press.

[LaL] R. Lawther, M.W. Liebeck, On the diameter of a Cayley graph of a simple group of Lie type based on a conjugacy class, J. Combin. Theory Ser. A 83 (1998) 118–137.

[Le] C. Leedham-Green, personal communication, 1999.

[LiSl] M.W. Liebeck, A. Shalev, The probability of generating a finite simple group, Geom. Dedicata 56 (1995) 103–113.

[LiS2] M.W. Liebeck, A. Shalev, Simple groups, probabilistic methods, and a conjecture of Kantor and Lubotzky, J. Algebra 184 (1996) 31–57.

[LiS3] M.W. Liebeck, A. Shalev, Diameters of finite simple groups: sharp bounds and applications, 1999, preprint.

[Lo] L. Lovász, Random walks on graphs: a survey, in: Combinatorics: Paul Erdős is Eighty, in: Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 353–398 (distributed by the AMS).

[LoW] L. Lovász, P. Winkler, Mixing times, in: D. Aldous, J. Propp (Eds.), Microsurveys in Discrete Probability, in: DIMACS Ser. Discrete Math. Theoret. Comput. Sci., Amer. Math. Soc., Providence, RI, 1998.

[Pal] I. Pak, On the probability of generating a finite group, 1999, preprint.

[Pa2] I. Pak, What do we know about the product replacement algorithm? in: Groups and Computation III, de Gruyter, Berlin, 2000.

[Pa3] I. Pak, The product replacement algorithm is polynomial, in: Proc. 41st IEEE FOGS, Redondo Beach, CA, 2000.

[PaB] I. Pak, S. Bratus, On sampling generating sets of finite groups and product replacement algorithm (extended abstract), in: Proc. ISSAC '99, 1999, pp. 91–96.

[Ro] Y. Roichman, Upper bound on the characters of the symmetric group, Invent. Math. 125 (1996) 451–458.

[Sc] M. Schönert et al., GAP—Groups, Algorithms, and Programming, Lehrstuhl D für Mathematik, RWTH, Aachen, Germany, 1994.

[Sh1] A. Shalev, Probabilistic group theory, in: C.M. Campbell, E.F. Robertson, N. Ruskuc, G.C. Smith (Eds.), Groups St Andrews 1997 in Bath, II, in: London Math. Soc. Lecture Note Ser., vol. 260, Cambridge Univ. Press, 1999, pp. 648–678.

[Sh2] A. Shalev, Simple groups, permutation groups, and probability, in: Proc. Int. Congress of Math., Berlin, in: Doc. Math., vol. II, 1998, pp. 129–137, http://www.mathematik.uni-bielefeld.de/documenta/xvol-icm/ICM.html.

[Va] L.G. Valiant, Short monotone formulae for the majority function, J. Algorithms 5 (1984) 363–366.