

# Strong bias of group generators: an obstacle to the “product replacement algorithm”

László Babai\*, Igor Pak†

## 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.

## 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 dis-

tribution 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 is wide open, we show that even the limiting distribution  $Q^k$  can be *very far from uniform*.

## 2 The main results

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

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

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

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

For a function  $T : G \rightarrow \mathbb{R}$  and a subset  $B \subseteq G$  we write  $T(B) = \sum_{g \in B} T(g)$ . The *total variation* of  $T$  is half of its  $\ell_1$ -norm:  $\|T\|_{tv} = \max_{B \subseteq G} |T(B)| = \frac{1}{2} \sum_{g \in G} |T(g)|$ . The *variation distance* of the probability distributions  $R$  and  $S$  over  $G$  is defined as  $\|R - S\|_{tv}$ . 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

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

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

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

$$\|Q^k - U\|_{tv} \rightarrow 1 \quad \text{as } n \rightarrow \infty$$

*assuming  $k \geq 4$  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.

\*Department of Computer Science, University of Chicago, Chicago, IL 60637, E-mail: [laci@cs.uchicago.edu](mailto:laci@cs.uchicago.edu)

†Department of Mathematics, Yale University, New Haven, CT 06520, E-mail: [paki@math.yale.edu](mailto:paki@math.yale.edu)

To prove the theorem we find an explicit set  $B$  such that  $Q^k(B) \rightarrow 0$  while  $U(B) \rightarrow 1$ . 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  $\{x_i^{\pm 1}, i = 1, 2, \dots\}$ . 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  $\omega(n) \rightarrow \infty$ ,  $\omega(n) = o(n)$ . Also, let  $k = k(n) \geq 4$  and  $k = o(n)$ . Set  $m = n^{k\omega(n)}$ . Let  $G = A_n^m$ . Then  $\mathbf{w}[Q^k] = 1$  is factorially likely (has probability  $1 - O(n^{-cn})$ ), while  $\mathbf{w}[U] = 1$  is factorially unlikely.*

**REMARK 2.1.** *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 [Bb1]. However, the

number of operations proven in [Bb1] 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<sup>1</sup>, 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  $X_t = (g_1, \dots, g_k) \in \mathcal{N}_k(G)$ . Define

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

where  $h_j = g_j g_i^{\pm 1}$  or  $h_j = g_i^{\pm 1} g_j$ , where the pair  $(i, j)$ ,  $1 \leq i, j \leq k$ ,  $i \neq j$  is chosen uniformly; the order of multiplication and the exponent  $\pm 1$  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  $k \geq \varkappa(G) + 1$ , the chain  $\mathbf{M}(G, k)$  is always connected. However, this problem remains wide open.

It was shown in [CLMNO] (cf. [DsS2]) that  $k \geq \varkappa + \tilde{\varkappa}$  suffices for connectedness 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  $k \geq 3$ , 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  $k \geq \varkappa(G) + 1$ , 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

<sup>1</sup>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].

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, DsS1, DsS2, Pa2, PaB]). 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 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  $(g_1, \dots, g_k) \in \mathcal{N}_k(G)$ . Let  $Q^k_\circ$  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  $T \rightarrow \infty$  if  $\mathbf{M}$  is ergodic.

The following is immediate and does not depend on the ergodicity of  $\mathbf{M}$ .

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

$$Q^k_\circ(g) = Q^k_*(g) = Q^k(g). \quad \square$$

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  $G = A_n^{n! / 8}$ ,  $n \geq 5$ . For  $k = o(n)$  we obtain

$$\|Q^k - U\|_{tv} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

It is known that  $G$  can be generated by two elements (see [Ha, KaL]). Thus taking  $k = \max\{10, \kappa(G)\}$  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  $\chi^2$  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  $Q^k(0) = 1/2 - \delta_k$  and  $Q^k(1) = 1/2 + \delta_k$  where  $\delta_k = 1/2(2^k - 1)$ , hence the bias in this case is  $\|Q^k - U\|_{tv} = \delta_k$ .

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  $|Q^k(H) - U(H)| = \delta_k + O(1/p^{k-1})$  and therefore the bias of the distribution  $Q^k$  is

$$\|Q^k - U\|_{tv} \geq \delta_k + O(1/p^{k-1})$$

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  $k \rightarrow \infty$  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.

#### 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

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

(see also [KaL]). The right hand side can be interpreted as the number of orbits of the diagonal action of  $\text{Aut}(H)$  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  $(g_1, \dots, g_k)$  the elements of  $\mathcal{N}_k(G)$ , and let  $g_i = (h_1^{(i)}, \dots, h_m^{(i)})$ , where  $h_j^{(i)} \in A_n$ ,  $1 \leq i \leq k$ ,  $1 \leq j \leq m$ . Observe that in order for the elements  $g_1, \dots, g_k$  to generate  $G$ , the elements  $h_j^{(1)}, \dots, h_j^{(k)}$  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  $\text{Aut}(H)$  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]) :

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

For simple groups  $H$  it is known that

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

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 [LiS1]. When  $H = A_n$  we conclude that  $\varphi_2 \geq 1/2$  when  $n$  is large enough. Therefore

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

where  $k \geq 2$ . In particular,  $d_2(A_n) \geq n!/8$ . Now let  $m \leq n!/8$ ,  $G = A_n^m$ . Then

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

We conclude:

LEMMA 5.1. *If  $k \geq 4$  and  $m \leq n!/8$  then*

$$(5.1) \quad \varphi_k(G) = (\varphi_k(A_n))^m \cdot (1 + O(1/n!)).$$

Equation (5.1) 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 5.1. *Let  $\mathcal{E} \subseteq (\mathcal{N}_k(A_n))^m$ . Then*

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

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

REMARK 5.1. *Let  $\sigma^{(j)} = (\sigma_1^{(j)}, \dots, \sigma_k^{(j)})$  denote elements of  $\mathcal{N}_k(A_n)$  ( $1 \leq j \leq m$ ). Corollary 5.1 means that for most calculations, we can treat the components  $\sigma^{(j)}$  of a uniform random element  $(\sigma^{(1)}, \dots, \sigma^{(m)}) \in \mathcal{N}_k(G)$  as independent; for  $k \geq 4$  and  $m \leq n!/8$ , 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  $n \rightarrow \infty$ .

Denote by  $x = (\sigma_1, \dots, \sigma_k)$  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  $k \geq 2$  we have*

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

*Proof.* As above, let  $x = (\sigma_1, \dots, \sigma_k) \in A_n^k$  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  $\sigma_i$  share a common fixed point.

The probability that each  $\sigma_i$  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  $k \geq 2$  we have

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

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

PROPOSITION 6.2. For  $k \geq 2$  we have

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

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

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

We estimate the conditional probability  $\mathbf{P}(\mathcal{A} | \mathcal{B})$  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

$$\begin{aligned} \mathbf{P}(\mathcal{A} | \mathcal{B}) &= 1 - (\mathbf{P}(\sigma(1) = 1))^{k-1} - (n-1) \cdot \\ &\quad \cdot (\mathbf{P}(\sigma(2) = 2))^{k-1} \cdot \mathbf{P}(\sigma(2) = 2 | \sigma(1) = 1) + \dots \\ &= 1 - \frac{1}{n^{k-1}} - (n-1) \cdot \frac{1}{(n-1)n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right) \\ &= 1 - \frac{2}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right). \end{aligned}$$

We conclude that

$$\begin{aligned} \mathbf{P}(\mathcal{B} | \mathcal{A}) &= \frac{1}{n} \cdot \frac{1 - \frac{2}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right)}{1 - \frac{1}{n^{k-1}} + O\left(\frac{1}{n^{2k-2}}\right)} \\ &= \frac{1}{n} - \frac{1}{n^k} + O\left(\frac{1}{n^{2k-1}}\right). \end{aligned}$$

This completes the proof.  $\square$

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

PROPOSITION 6.3. For  $k \geq 2$  we have

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

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

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

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  $\sigma_k$  is  $(S_2 \times S_{n-2}) \cap A_n$ . Therefore

$$\begin{aligned} \mathbf{P}(\bar{\mathcal{A}} | \mathcal{C}) &\leq \mathbf{P}(\bar{\mathcal{A}} \wedge \mathcal{C}) \\ &\leq \binom{n}{2} \cdot (\mathbf{P}(\sigma(1) = 2, \sigma(2) = 1))^k + \dots \\ &= O\left(\frac{1}{n^{2k-2}}\right). \quad \square \end{aligned}$$

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

PROPOSITION 6.4. For  $k \geq 2$  we have

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

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

## 7 Proof of Theorem 2.1

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

$$(7.2) \quad \mathbf{P}(\xi > pm + a) < \exp(-2a^2/pm + 4a^3/(pm)^2),$$

and

$$(7.3) \quad \mathbf{P}(\xi < pm - a) < \exp(-2a^2/pm).$$

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

$$(7.4) \quad \#\{j | \sigma_j(1) = 1, 1 \leq j \leq m\} > m \cdot \left(\frac{1}{n} - \frac{1}{2 \cdot n^k}\right).$$

LEMMA 7.1. Under the conditions of Theorem 2.1,

$$Q^k(B) \rightarrow 0, \text{ and } U(B) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

assuming  $k = o(n)$ .

*Proof.* For the proof we note that the quantity on the right hand side of inequality (7.4) is halfway between the expected value of the left hand side under uniform distribution ( $m/n$ ) and under  $Q^k$  ( $m(1/n - 1/n^k)$ ). Now both parts of the claim follow from Chernoff's bounds. Indeed, using the Chernoff bound (7.3) with  $p = 1/n$ ,  $a = m/(2n^{k-1})$ , we obtain

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

By definition  $m = n!/8$ . Thus when  $k = o(n)$  we have  $U(B) \rightarrow 1$  as  $n \rightarrow \infty$ . 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 5.1 and Remark 5.1 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 (7.2).  $\square$

Observe that Lemma 7.1 immediately implies Theorem 2.1. Indeed,

$$\|Q^k - U\|_{tv} \leq |Q^k(B) - U(B)| \rightarrow 1 \text{ as } n \rightarrow \infty. \quad \square$$

**REMARK 7.1.** *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  $z_1, \dots, z_s$  be Boolean variables. Let  $Th_{s,t}(z_1, \dots, z_s)$  denote the *threshold function* which takes value 1 if  $\sum_{i=1}^s z_i \geq t$  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  $x = (\sigma_1, \dots, \sigma_k) \in A_n^k$ . 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

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

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

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

Let us now take  $s$  independent samples  $x_1, \dots, x_s \in A_n^k$ , 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

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

We set  $s = 2n^{4k}$  and apply Chernoff's bounds with threshold  $t = 2n^{3k} + n^{2k}$ . This value of  $t$  is halfway

between the two expected values above for the given  $s$ . It follows that

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

while

$$\begin{aligned} \mathbf{P}(Th_{s,t}(\mathcal{D}_1, \dots, \mathcal{D}_s) | \mathcal{A}) &> 1 - \exp(-n^k) + O(1/n!) \\ &> 1 - n^{-cn}. \end{aligned}$$

Now we express the threshold function by a monotone Boolean circuit with suitable parameters. 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  $Th_{s,t}$ ; 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  $\leq n^{O(k)}$  and depth  $O(k \log n)$  with  $s = 2n^{4k}$  input variables such that, assuming  $k = k(n) = o(n)$ ,  $k \geq 4$ , we have  $\mathbf{P}(\mathbf{F}) = 1 - O(n^{-cn})$  and  $\mathbf{P}(\mathbf{F} | \mathcal{A}) = O(n^{-cn})$ , where  $\mathbf{F} = \mathbf{F}_{n,k}(\mathcal{D}_1, \dots, \mathcal{D}_s)$ .  $\square$*

## 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  $n \geq 5$ , 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

$$(9.5) \quad z = (u_1^{-1} g u_1) \cdot \dots \cdot (u_N^{-1} g u_N)$$

over a finite group  $H$ . For a fixed  $g \in A_n$  and randomly chosen  $u_i$  one can think of  $z$  as the  $N$ -th 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 equation (9.5). If the  $u_i$  are independent, uniformly distributed elements from  $A_n$  then

$$\left| \mathbf{P}(z = h) - \frac{1}{|A_n|} \right| < \frac{1}{2|A_n|}$$

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

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

where  $1 > c_1 > 0$ ,  $N = cn \log n$ ,  $c, c_1$  are universal constants<sup>2</sup>. Now use a standard bound which relates mixing in relative pointwise distance (or  $\ell_\infty$  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.  $\square$

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

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

$$\begin{aligned} \mathbf{P}(z(g) \cdot z(h) = 1) &= \sum_{f \in A_n} \mathbf{P}(z(g) = f) \cdot \mathbf{P}(z(h) = f^{-1}) \\ &\leq |A_n| \cdot \left( \frac{3/2}{|A_n|} \right)^2. \end{aligned}$$

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,

$$\begin{aligned} \mathbf{P}([z(g), z(h)] = 1) &= \sum_{v_1, v_2 \in A_n, [v_1, v_2] = 1} \mathbf{P}(z(g) = v_1) \cdot \\ &\quad \cdot \mathbf{P}(z(h) = v_2) \leq r(A_n) \cdot \left( \frac{3/2}{|A_n|} \right)^2 \end{aligned}$$

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

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

## 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  $n^{c_k}$  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  $i$ -th 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  $\mathbf{P}_U(\mathbf{w} = 1) = O(n^{-cn})$  (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  $\mathbf{w} = (\sigma_1, \dots, \sigma_m)$  is the identity (which is far more than what we need).

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

Under this assumption, it is factorially likely that

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

$\sigma_i = 1$  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.  $\square$

## 11 Direct products of simple groups of Lie type

It turns out that virtually all the results (including Theorems 2.1, 2.2) have an analog for products of large simple groups of Lie type. Here we will sketch the idea of the main steps of the proof, while omitting most details.

**THEOREM 11.1.** *Let  $H_n$  be a family of simple groups of Lie type,  $G_n = H_n^m$ , where  $m = \lfloor \sqrt{|H_n|} \rfloor$ . Let  $Q_n^k$  be the probability distribution of the first component of  $\mathcal{N}_k(G_n)$ . Then*

$$\|Q_n^k - U\|_{tv} \rightarrow 1 \quad \text{as } |H_n| \rightarrow \infty$$

given  $k = q^{o(rk(H_n))}$ .

It is known, on the other hand, that the groups  $G_n$  can be generated by only two generators when  $|H_n|$  is large enough (see [Ha, KaL, Pa1]). Therefore we again obtain a strong bias in the probability distribution of the output of the product replacement algorithm.

*Sketch of proof.* First, recall that a random pair of elements generates the simple group  $H$  of Lie type with probability  $\rightarrow 1$  as  $|H| \rightarrow \infty$  [KaL, LiS1] (see [Sh1]). Thus the components of elements in generating  $k$ -tuples are virtually independent. This is made explicit by the second author in [Pa1]). The analog of the result in [Bb2] used in Section 6 can be found in [KaL, LiS2]. Take as an analog of the stabizer subgroup of  $A_n$  a subgroup of  $H$  which fixes a subspace of codimension 1. Then much of Sections 6 and 7 goes through.  $\square$

Let us finish by elaborating on how one should be able to obtain an analog of Theorem 2.2 for classical groups and perhaps for all simple groups of Lie type.

To describe the analog of the event  $\mathcal{D}'$  in Section 8 we would need to work on a case-by-case basis, but the general idea is to consider elements of which the order characterizes them to belong to large conjugacy

classes. 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. This can be done in several ways. First, one can use bounds on characters (see estimates in [Gl] for classical groups). Also, one can use the known bound  $\Delta \leq 8(5k + 7)$  on diameter of the corresponding Cayley graph<sup>3</sup> in terms of the untwisted Lie rank  $k$  of  $H$  [LaL]. Then apply a standard  $O(\Delta^2 \log |H|)$  bound for the mixing time (see [AF, Lo]) and proceed as before. The analog of this approach for  $H = A_n$  would be to use the  $\Delta \leq n/2$  bound of Dvir (see [Dv]), that would give weaker bounds than those in Lemma 9.2. Finally, to obtain an estimate on the probability that a random pair of elements of  $G$  commute, we need an upper bound on the number of conjugacy classes (see e.g. [Gl] when  $H$  is classical).

**Acknowledgements.** We would like to 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 would also like to thank the University of Chicago for the hospitality during his brief visit in March 1999.

## References

- [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*, book, in preparation
- [AS] N. Alon, J.H. Spencer, *The Probabilistic Method*, Wiley 1992.
- [Bb1] L. Babai, Local expansion of vertex-transitive graphs and random generation in finite groups, *Proc. 23rd ACM STOC*, 1991, pp. 164–174.
- [Bb2] L. Babai, The probability of generating the symmetric group, *J. Comb. Th., Ser. A* **52** (1989), 148–153
- [Bb3] L. Babai, Randomization in group algorithms: conceptual questions. In: *Groups and Computation II*, (L.

<sup>3</sup>This diameter for conjugacy classes is also known as the covering number.



- Finkelstein and W. M. Kantor, eds.), DIMACS Ser. in Discr. Math. and Theor. Comp. Sci. Vol 28, A.M.S. 1997, pp. 1–16.
- [BKPS] L. Babai, W. M. Kantor, P. P. Pálffy, 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  $NC^1$ , *JCSS* **38** (1989), 150–164.
- [BoC] W. Bosma, J. J. Cannon, *Handbook of Magma functions*. School of Mathematics and Statistics, University of Sydney.
- [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 and E. A. O’Brien, Generating random elements of a finite group, *Comm. Alg.* **23** (1995), 4931–4948.
- [ChG] F. R. K. Chung, R. L. Graham, Random walks on generating sets for finite groups, *The Electronic J. of Comb.* **4/2** (1997) #R7.
- [Ds] P. Diaconis, *Group Representations in Probability and Statistics*, Lecture Notes Monograph Series, Vol 11, IMS, Hayward, California, 1988.
- [DsG] P. Diaconis, R. L. Graham, The graph of generating sets of an abelian group, *Colloq. Math.* **80** (1999), 31–38.
- [DsS1] P. Diaconis, L. Saloff-Coste, Walks on generating sets of groups, *Prob. Th. Rel. Fields* **105** (1996), 393–421.
- [DsS2] P. Diaconis, L. Saloff-Coste, Walks on generating sets of abelian groups, *Invent. Math.* **134** (1998), 251–199.
- [Dx] J. D. Dixon, The probability of generating the symmetric group, *Math. Z.* **110** (1969), 199–205.
- [DxM] J. D. Dixon, B. Mortimer, *Permutation Groups*, Springer, 1996.
- [Dv] Y. Dvir, Covering properties of permutation groups. In: *Products of Conjugacy Classes in Groups* (Z. Arad and M. Herzog, eds.), Springer LNM vol. 1112 (1985), 197–221.
- [Fr] F. G. Frobenius, Über Gruppencharaktere, *Sitzungsber. der Berl. Akad.* 1896, 985–1021.
- [Gl] D. Gluck, Characters and random walks on finite classical groups, *Adv. Math.* **129** (1997), 46–72.
- [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, *Proc. Int. Congress of Math.*, Documenta Math., Berlin, 1998, Vol. II, pp. 77–86, [www.mathematik.uni-bielefeld.de/documenta/xvol-icm/ICM.html](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, *Memoirs of the AMS*, 1999, to appear.
- [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. Comb. Theory, Ser A* **83** (1998), 118–137.
- [Le] C. Leedham-Green, personal communication, 1999.
- [LiS1] 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.
- [Lo] L. Lovász, Random walks on graphs: a survey, In: *Combinatorics: Paul Erdős is Eighty*, Vol. 2, Bolyai Society Mathematical Studies 2, Budapest 1996 (distributed by the AMS), pp. 353–398.
- [LoW] L. Lovász, P. Winkler, *Mixing Times*, Microsurveys in Discrete Probability (D. Aldous and J. Propp, eds.), DIMACS Series, AMS, 1998.
- [Pa1] I. Pak, On the probability of generating a finite group, preprint (1999).
- [Pa2] I. Pak, What do we know about the product replacement algorithm?, in preparation (1999).
- [PaB] I. Pak, S. Bratus, On sampling generating sets of finite groups and product replacement algorithm (Extended Abstract), in *Proc. ISSAC’99*.
- [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: *Groups St Andrews 1997 in Bath, II* (C. M. Campbell, E. F. Robertson, N. Ruskuc, G. C. Smith, eds.), London Math. Soc. Lect. Notes 260, Cambr. U. Press, 1999, pp. 648–678.
- [Sh2] A. Shalev, Simple groups, permutation groups, and probability, *Proc. Int. Congress of Math.*, Documenta Math., Berlin, 1998, Vol. II, pp. 129–137, [www.mathematik.uni-bielefeld.de/documenta/xvol-icm/ICM.html](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.