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

László Babai*, Igor Pak†

Extended Abstract: July 12, 1999

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 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 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.*

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

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

*Department of Computer Science, University of Chicago, Chicago, IL 60637, E-mail: **laci@cs.chicago.edu**

†Department of Mathematics, Yale University, New Haven, CT 06520, E-mail: **paki@math.yale.edu**

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

Let U denote a uniform distribution over G. The *total variation distance* $\|\mathrm{Q}^k - \mathrm{U}\|_{tv}$ is defined as follows:

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

This quantity is between 0 and 1.

Theorem 1.1 *Let $G = A_n^m$, where $m = n!/8$. Then*

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

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 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 1.2 *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 $k = k(n) \geq 4$ 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 1.3 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).

2 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 ($(\log|G|)^5$), not suitable in practice. Therefore heuristic algorithms are used.

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

The product replacement algorithm works as follows [CLMNO]. Consider the Markov chain $\mathbf{M} = \{X_t\}$ on $\mathcal{N}_k(G)$ as follows. 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 ± 1 are determined by independent flips of a fair coin. The algorithm runs the Markov chain 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. It is known that g is distributed (nearly) uniformly if $k = \Omega(\log|G|)$ and T is large enough.

Let $\varkappa(G)$ and $\widetilde{\varkappa}(G)$ denote the smallest and the largest size, respectively, of a minimal generating set. It is known (see [CLMNO, DS2]) that when $k \geq \varkappa + \widetilde{\varkappa}$, the Markov chain $\mathbf{M}$ is reversible, aperiodic, irreducible, and has a uniform stationary distribution. Thus the chain is ergodic and can be used for approximate sampling from $\mathcal{N}_k(G)$. The empirical tests seem to indicate that the chain mixes rapidly (see [CLMNO, LG]) but no results are known 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.

While the former problem (a problem of mixing rate) has been studied by several authors (see [CG, DS1, DS2, Pa2, PB]), the present paper seems to be the first one to point out the second type of error.

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 $\mathrm{Q}_\circ^k$ 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 $T \to \infty$ and $\mathbf{M}$ is ergodic.

The following is an immediate consequence of ergodicty.

Proposition 2.1 *For any finite group G, $k \geq \varkappa(G) + \widetilde{\varkappa}(G)$, and $g \in G$ we have*

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

3 Direct product of groups

Evidently $\mathcal{N}_k(G) \subset (\mathcal{N}_k(A_n))^m$. The difference of these two sets is very small:

Lemma 3.1 *If $k \geq 4$ and $m \leq n!/8$ then*

$$|\mathcal{N}_k(G)| \;=\; |(\mathcal{N}_k(A_n))^m| \cdot (1 + O(1/n!)). \tag{1}$$

This result follows from classical work by P. Hall [Ha] and J. Dixon [Dx]. □

Corollary 3.2 *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 $(\mathcal{N}_k(A_n))^m$.) □

Remark 3.3 Let $\sigma^{(j)} = \left(\sigma_1^{(j)}, \ldots, \sigma_k^{(j)}\right)$ denote elements of $\mathcal{N}_k(A_n)$ $(1 \leq j \leq m)$. Corollary 3.2 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 $k \geq 4$ and $m \leq n!/8$, the error will be $O(1/n!)$.

4 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 4.1 *For $k \geq 2$ we have*

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

The idea of the proof is to show that the most frequent reason for $\sigma_1, \ldots, \sigma_k$ not to generate A_n is that all σ_i share a common fixed point. The probability of this is dominated by the term n^{-k+1}, the error being $O(n^{-2k+2})$. The proof uses inclusion-exclusion and the following estimate from [Dx] and [Bb2]: the probability that a pair of random permutations lies in a maximal subgroup 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. [Sh]). □

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

Proposition 4.2 *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

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

We conclude that

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

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

Proposition 4.3 *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. $\square$

5 Proof of Theorem 1.1

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

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

We claim that under the conditions of Theorem 1.1,

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

This immediately implies Theorem 1.1.

For the proof of the claim we note that the quantity on the right hand side 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, as stated by Alon and Spencer, [AS], Theorems A.11 and A.13 (pp. 237-238). For the result regarding Q^k we also need the strong approximate independence of the components, stated in Cor. 3.2 and Remark 3.3. □

6 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 4, 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 4.3 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

$$\mathbf{P}(Th_{s,t}(\mathcal{D}_1, \dots, \mathcal{D}_s)) < \exp(-n^k(1 - 1/4n^k)) < n^{-cn}, \tag{2}$$

while

$$\mathbf{P}(Th_{s,t}(\mathcal{D}_1, \dots, \mathcal{D}_s) \,|\, \mathcal{A}) > 1 - \exp(-n^k) + O(1/n!) > 1 - n^{-cn}. \tag{3}$$

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 6.1 *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 $\mathbf{P}(\mathbf{F} \mid \mathcal{A}) = O(n^{-cn})$, where $\mathbf{F} = \mathbf{F}_{n,k}(\mathcal{D}_1, \ldots, \mathcal{D}_s)$.* $\square$

7 Simulation of 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 Barrigton's simulation of Boolean operations by group operations [Br], 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 7.1 *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 $g, h, u_1, u_2, \ldots$ 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

$$z = \left(u_1^{-1}\, g\, u_1\right) \cdot \ldots \cdot \left(u_N^{-1}\, g\, u_N\right) \tag{4}$$

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 7.2 *Fix $g \in A_n$, $g \neq 1$. Let $N = \Omega(n^2 \log^2 n)$ and define z by equation (4). 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 $z \in A_n$. It follows then from a result of Roichman (see [Ro]) that

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

where $1 > c_1 > 0$, $N = c\, n \log n$, c, c_1 are universal constants[1]. Now use a standard bound which relates mixing in relative pointwise distance (or ℓ_∞ distance) with mixing in total variation distance (see e.g. [AF, LW].) 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 7.1. 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 7.1.

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

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

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,

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

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 $\eta(A_n) = O\left(e^{c\sqrt{n}}\right)$. We conclude that

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

8 Proof of Theorem 1.2

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 1.2 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 6.1. Substitute the expression x_i^n for the i-th Boolean input variable.

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

Substitute the words w_1, w_2 given in Section 7 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 1.2.

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 7.1 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 6.1 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 6.1 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 7.1.

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. □

Acknowledgements. We would like to thank Persi Diaconis, Walter Feit, Bill Kantor, Charles Leedham–Green, László Lovász, Alexander Lubotzky, Alice Niemeyer, Aner Shalev and Leonard Soicher for helpful conversations. Theorem 1.1 was initially announced in a joint paper [PB] of Sergey Bratus and the second named author. Nathan Lulov pointed out reference [CH].

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, vol. 52 (1989), 148–153

[Bb3] L. Babai, Randomization in group algorithms: Conceptual questions Groups and Computation II, DIMACS Series, vol. 28, AMS, Providence, 1997

[Br] D. Barrington, Bounded-Width Polynomial-Size Branching Programs Recognize Exactly Those Languages in NC^1, JCSS 38 (1989).

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

[Ca] J. Cannon, MAGMA, Sydney.

[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

[CG] F.R.K. Chung, R.L. Graham, Random walks on generating sets for finite groups, The Electronic J. of Comb., 4, no. 2 (1997) #R7

[Ds] P. Diaconis, *Group Representations in Probability and Statistics*, Lecture Notes Monograph Series, Vol 11, IMS, Hayward, California, 1988

[DS1] P. Diaconis, L. Saloff-Coste, Walks on generating sets of groups, Prob. Th. Rel. Fields, vol. 105 (1996), 393–421

[DS2] P. Diaconis, L. Saloff-Coste, Walks on generating sets of abelian groups, Invent. Math. vol. 134 (1998), 251–199

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

[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. Ak. 1896, 985–1021

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

[Ka] W. Kantor, Simple groups in computational group theory, Proc. Int. Congress of Math., Vol. II (Berlin, 1998)

[LL] 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, vol 83 (1998), 118–137

[LG] C. Leedham–Green, personal communication, (1999)

[Lo] L. Lovász, Random walks on graphs: a survey, in: *Combinatorics: Paul Erdős is Eighty*, Vol. 2, Bolyai Society Mathematical Studies 2, Budapest 1996, pp. 353–398.

[LW] L. Lovász, P. Winkler, *Mixing Times*, Microsurveys in Discrete Probability (ed. D. Aldous and J. Propp), DIMACS Series, AMS, 1998

[Pa1] I. Pak, On probability of generating a finite group, preprint (1999)

[Pa2] I. Pak, What do we know about product replacement algorithm?, in preparation (1999)

[PB] I. Pak, S. Bratus, On sampling generating sets of finite groups and product replacement algorithm. (Extended Abstract), to appear in *Proc. ISSAC Conf.*, 1999

[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.

[Sh] A. Shalev, Probabilistic group theory, St. Andrews Lectures, Bath, 1997