# Simple Permutations Mix Well

Shlomo Hoory Avner Magen Steve Myers Charles Rackoff

Department of Computer Science*
University of Toronto

**Abstract**

We study the random composition of a small family of $O(n^3)$ simple permutations on $\{0,1\}^n$. Specifically we ask what is the number of compositions needed to achieve a permutation that is close to $k$-wise independent. We improve on a result of Gowers [7] and show that up to a polylogarithmic factor, $n^3k^3$ compositions of random permutations from this family suffice. Additionally, we introduce a new notion analogous to closeness to $k$-wise independence against adaptive adversaries and show the constructed permutation has the stronger property. This question is essentially about the rapid mixing of the random walk on a certain graph which we establish using a new approach to construct the so called canonical paths, which may be of independent interest. We also show that if we are willing to use a much larger family of simple permutations then we can guaranty closeness to $k$-wise independence with fewer compositions and fewer random bits.

## 1 Introduction

A question that occurs naturally in cryptography is how well the composition of permutations drawn from a simple distribution resembles a random permutation. Although this type of construction is a common source of security, the mathematical justification for it is troubling, and is one of the motivations of this work. This motivation is discussed in more detail in Section 6.

A source is *pseudo-random* if it is random in the computational sense, namely no computationally bounded machine can distinguish it from a truly random one. Another natural and well studied measure for randomness, although lacking an obvious linkage to computational considerations, is the notion of almost *$k$-wise independence*. In the context of permutations it means that the values of *any* $k$ distinct elements in the domain, is approximately uniformly distributed among the sets of $k$ distinct elements. We can now form the following question. Consider a small set of simple permutations, which we call *basic permutations*, on binary strings of length $n$, and compose random elements of this set $T$ times to get a permutation $f$. Is this permutation pseudo-random? How far is $f$ from a $k$-wise independent permutation? The second question is the focus of this paper; specifically we bound from above the number of times $T$ we need to compose the basic permutations in order to get a good enough approximation to a $k$-wise independent permutation.

In [7] T. Gowers studied this question. The basic permutations he considered were the ones that fix all but three coordinates of the $n$-bit strings, namely a set of size $O(n^3)$ (which is a tiny

* {shlomoh|avner|myers|rackoff}@cs.toronto.edu

fraction of the $2^n!$ possible permutations). Gowers managed to show that when composing [1] $\tilde{O}(n^3k(n^2+k)(n^3+k))$ basic permutations randomly, the result is a distribution over permutations that is close to $k$-wise independent, provided a certain divisibility condition regarding $n$ and $k$ applies. Here we show that by using this set of permutations (or in fact even a more restricted set) and without any additional conditions, that it is enough to compose the basic permutation $\tilde{O}(n^3k^3)$ times to get the above guarantee.

Our question is essentially one of the mixing rate of a random walk on the graph whose vertices are $k$-tuples of distinct $n$-bit strings, and whose edges are induced by the obvious operation of basic permutations on the vertices. The mixing rate of this graph is exactly that minimal number of composition $T$ we seek. Using the well known connections between the combinatorial/algebraic properties of graphs to the mixing rate of random walks on them, our goal is to show proper expansion/conductance/spectral gap for our graphs. In the course of showing that we improve Gowers' upper bound on the diameter of this graph from $O(kn^2)$ to $\tilde{O}(kn)$ which is tight. For estimating the conductance of the graph we present a new and general way to construct the so called *canonical paths* in a wide class of graphs (Cayley graphs or more generally Schreier graphs) and provide an "algorithmic" flavour to showing mixing. We believe that this technique (essentially Lemma 4) can be useful in showing rapid mixing for other Markov chains.

Another contribution of this work is the notion of *strong closeness to k-wise independence* which is a strengthening of the usual closeness to $k$-wise independence. Given a permutation $f$ drawn from a partricular distribution, how well can a computationally unbounded machine that is allowed to query $f$ $k$ times, distinguish it from a truly random permutation. We show an upper bound on the number of compositions needed to satisfy this stronger property.

To state our results we need to define our basic permutations. We look at permutations that change just one bit of their input, by XORing it with a function on few other bits. Formally, for $0 < w < n$ we define $\mathcal{F}_w$ to be the set of permutations $f_{i,J,h}$ where $i \in [n]$, $J = \{j_1, \ldots, j_w\}$ is a size $w$ index set disjoint from $i$, and $h$ is a boolean function on $\{0,1\}^w$. $f_{i,J,h}$ then maps $(x_1, \ldots, x_n) \in \Omega$ to $(x_1, \ldots, x_{i-1}, x_i \oplus h(x_{j_1}, \ldots, x_{j_w}), x_{i+1}, \ldots, x_n)$. Clearly $\mathcal{F}_2$ is a subset of Gowers' set of basic permutations. Also note that $|\mathcal{F}_w| = n \cdot \binom{n-1}{w} \cdot 2^{2^w}$. Our main results are.

**Theorem 1.** *Let $k = O(2^{n/4})$, and let $T$ be the minimal number of random compositions of independent and uniformly distributed permutations from $\mathcal{F}_2$ needed to generate a permutation which is strong $\epsilon$-close to $k$-wise independent. Then $T = \tilde{O}(n^2k^2 \cdot (\log(1/\epsilon) + nk))$.*

It is interesting to note that in some sense $\mathcal{F}_2$ or the set of basic permutations Gowers used are minimal in order for the graph decribed above is connected for $k \geq 4$ [2]. As Gowers notes, the connectedness for these sets is not immedaitely obvious; it does in fact follow from older results, eg [5]. If instead of striving to achieve the minimal set of basic permutations we want to use as little as possible true random bits in order to get $k$-wise independence, it is interesting to check other candidates to be our sets of basic-permutations. This number of random bits is simply the $\log_2$ of the number of basic permutations times the number of times we compose them. Therefore, Theorem 1 tells us $\tilde{O}(n^3k^3)$ random bits suffice to get the desired property. It follows from the next theorem that one can use as little as $\tilde{O}(n^2k^2)$ such bits, when insted of $\mathcal{F}_2$ we take $\mathcal{F}_{2\log k + \log n + \log\log n + 8}$.

[1] the tilde suppresses polylogarithmic factors in $n$ and $k$

[2] connectedness of the graph is clearly a necessary condition for generating almost $k$-wise independent distribution of permutations

**Theorem 2.** *Let $T$ be the minimal number of random compositions of independent and uniformly distributed permutations from $\mathcal{F}_w$ for $w \geq 2\log k + \log n + \log\log n + 8$, needed to generate a permutation which is $\epsilon$-close to $k$-wise independent. Then*

$$T = O(\log(1/\epsilon) \cdot n \cdot \log n \cdot (\log k + \log n)).$$

*If instead we consider strong $\epsilon$-closeness to $k$-wise independence, then*

$$T = O(\log(1/\epsilon) \cdot n^2 k \cdot \log n \cdot (\log k + \log n)).$$

## 2 Preliminaries

Let $f$ be a random permutation on some base set $X$. Denote by $X^{(k)}$ the set of all $k$-tuples of distinct elements from $X$. We say that $f$ is $\epsilon$-close to $k$-wise independent if for every $(x_1, \ldots, x_k) \in X^{(k)}$ the distribution of $(f(x_1), \ldots, f(x_k))$ is $\epsilon$-close to the uniform distribution on $X^{(k)}$. We measure the distance between two probability distributions $p, q$ by the total variation distance, defined by

$$d(p,q) = \frac{1}{2}||p-q||_1 = \frac{1}{2}\sum_{\omega}|p(\omega) - q(\omega)| = \max_A \sum_{\omega \in A} p(\omega) - q(\omega).$$

We sometimes abuse notation, and replace $p$ or $q$ by a random variable having this distribution.

Assume a group $H$ is acting on a set $X$ and let $S$ be a subset of $H$ closed under inversion. Then the *Schreier graph* $G = \mathrm{sc}(S, X)$ is defined by $V(G) = X$ and $E(G) = \{(x, xs) : x \in X,\ s \in S\}$. Also, for a sequence $\omega = (s_1, \ldots, s_l) \in S^l$ we denote $x\omega = xs_1 \cdots s_l$. We will sometimes refer by $x\omega$ also to the walk $x, xs_1, \ldots, xs_1 \cdots s_l$.

The *random walk* $(X_0, X_1, \ldots)$ assoicated with a $d$-regular graph $G$ is defined by the transition matrix $P_{vu} = \Pr(X_{i+1} = u | X_i = v)$ which is $1/d$ if $(v,u) \in E(G)$ and zero otherwise. The uniform distribution $\pi$ is stationary for this Markov process. If $G$ is connected and not bipartite, we know that given any initial distribution of $X_0$, the distribtion of $X_t$ tends to the uniform distribtion. We define the mixing time of $G$ as $\tau(\epsilon) = \max_{v \in V(G)} \min\{t : d(P^{(t)}(v, \cdot), \pi) < \epsilon\}$, where $P^{(t)}(v, .)$ is the probability distribution of $X_t$ given that $X_0 = v$.

It is not hard to prove (see for example Lemma 20 in [1]) that

$$\tau(2^{-l-1}) \leq l \cdot \tau(\frac{1}{4}). \tag{1}$$

## 3 Strong closeness to $k$-wise independence

Let $\mathcal{F}$ be a distribution of permutations $f : \Omega \to \Omega$. We can think of $k$-wise independence in the following terms: a (computationally unbounded) adversary chooses a tuple $\vec{x} \in \Omega^{(k)}$; it is then given either an element at random from $\Omega^{(k)}$ or the element $f(\vec{x})$ for a random $f \in \mathcal{F}$; and finally it is asked to distinguish the two distributions. To say that a distribution is $k$-wise independent (resp. $\epsilon$-close $k$-wise independent) is to say that the distinguishing probability is 0 (resp. less than $\epsilon$) for all adversaries. One can strengthen the notion of adversary to permit it to adaptively choose $\vec{x}$. Such an adversary is a tuple $A = (\alpha_1, ..., \alpha_k, \overline{A})$, where $\alpha_i : \Omega^{(i-1)} \to \Omega$ and $\overline{A} : \Omega^{(k)} \to \{0,1\}$.

The adversary iterates through $k$ steps, where in the $i$th step it requests $\alpha_i(r_1,..,r_{i-1})$ and gets response $r_i$. After the $k$th step it outputs $\overline{A}(r_1,...,r_k)$.

In the case of (strict) $k$-wise independence it can be shown that such a strengthening cannot help the adversary distinguish the distributions. This is not the case for $\epsilon$-close $k$-wise independence. Consider the uniform distribution over the set $\mathcal{F}$ of permutations $f : \Omega \to \Omega$ where $f = f^{-1}$, and the case $k = 2$. For every $(x_1, x_2) \in \Omega^{(2)}$, the process of choosing a random $f \in \mathcal{F}$ and outputting $(f(x_1), f(x_2))$ results in a distribution close to the uniform one over $\Omega^{(2)}$. In contrast, consider an adaptive adversary $A$ distinguishing the same distributions: let $A = (\alpha_0, \alpha_1, \overline{A})$, where $\alpha_0 = 0^n, \alpha_1(x) = x$ and $\overline{A}(r_1, r_2)$ outputs 1 if $r_2 = 0^n$ and outputs 0 otherwise. Note that if $(r_1, r_2) = (f(0^n), f(f(0^n)))$ then $\overline{A}$ outputs 1. Alternatively if $(r_1, r_2)$ is chosen uniformly from $\Omega^{(2)}$ then $\overline{A}$ outputs 0 w.h.p.. Therefore, $A$ effectively distinguishes between the two distributions.

This motivates the following definition: a distribution $\mathcal{F}$ is said to be *strongly* $\epsilon$-close $k$-wise independent if it is $\epsilon$-close to $k$-wise independent against *adaptive* adversaries. We will now show that any distribution of functions that is $\epsilon$-close to $k$-wise independent using a strong distance measure is also *strongly* $\epsilon$-close to $k$-wise independent. The distance notion we have in mind is *relative pointwise distance*. The *relative pointwise distance*, or $d_{rp}$, between probability distributions $p$ and $q$ over $\Omega$ is: $d_{rp}(p,q) = \max_{\omega\in\Omega} |p(\omega) - q(\omega)|/p(\omega)$.

Suppose that in an experiment an adversary, $A = (\alpha_1, ..., \alpha_k, \overline{A})$, has adaptively chosen $\vec{x}$ as its queries and received $\vec{r}$ as its replies. Note that $\vec{r}$ and $(\alpha_1, ..., \alpha_k)$ fix $\vec{x}$, and therefore there exists an associated function $\alpha : \Omega^{(k)} \to \Omega^{(k)}$ that maps replies $\vec{r}$ to the corresponding queries $\vec{x}$.

Let $p_{\vec{r}}$ and $q_{\vec{r}}$ denote the probabilities of $A = (\alpha_1, ..., \alpha_k, \overline{A})$ selecting $\vec{x}$ and receiving $\vec{r}$ in the respective experiments where $\vec{r}$ is chosen uniformly in $\Omega^{(k)}$ and where $\vec{r} = f(\vec{x})$ for $f$ chosen uniformly in $\mathcal{F}$. Let $I = \{\vec{r} | \overline{A}(\vec{r}) = 1\}$. Then:

$$\Pr_{\vec{r}\in\Omega^{(k)}}[\overline{A}(\vec{r}) = 1] - \Pr_{f\in\mathcal{F}}[\overline{A}(\vec{r}) = 1] = \sum_{\vec{r}\in I}(p_{\vec{r}} - q_{\vec{r}}) \leq$$

$$\frac{\sum_{\vec{r}\in I}(p_{\vec{r}} - q_{\vec{r}})}{\sum_{\vec{r}\in I} p_{\vec{r}}} \leq \max_{\vec{r}\in I} \frac{|p_{\vec{r}} - q_{\vec{r}}|}{p_{\vec{r}}} \leq \max_{\vec{r}} \frac{|p_{\vec{r}} - q_{\vec{r}}|}{p_{\vec{r}}} = v$$

Let $\vec{s}$ be a tuple attaining the last maximum, and let $\vec{y} = \alpha(\vec{s})$. Now let $q'$ be the distribution $f(\vec{y})$ where $f$ is chosen uniformly on $\mathcal{F}$. Observing that $q'_{\vec{y}} = q_{\vec{y}}$ we see that:

$$v = \frac{|p_{\vec{y}} - q_{\vec{y}}|}{p_{\vec{y}}} = \frac{|p_{\vec{y}} - q'_{\vec{y}}|}{p_{\vec{y}}} \leq d_{rp}(p, q').$$

Finally, we note that if a distribution is $\frac{\epsilon}{|\Omega^{(k)}|}$-close to $k-$wise independent then it is *strongly* $\epsilon$-close to $k$-wise independent.

# 4 Mixing with width two permutations

One method to prove that the random walk on $G$ mixes rapidly is to use the *Canonical Paths* method of Jerrum and Sinclair [9, 10, 8] to obtain a lower bound on its conductance

$$\Phi(G) = \min_{A\subseteq V(G),\, |A|\leq |V|/2} \frac{|E(A, \overline{A})|}{d \cdot |A|}, \tag{2}$$

where $\overline{A} = V(G) \setminus A$, and $E(A, \overline{A}) = \{(v,u) \in E(G) : v \in A \text{ and } u \notin A\}$. A fundamental result relating conductance and rate of mixing is the following. We say that a random walk is lazy if for some constant $\delta > 0$ we have $\Pr[X_{t+1} = v | X_t = v] \geq \delta$ for all $v \in V(G)$.

**Theorem 3.** *(Jerrum and Sinclair [9]) If the random walk on $G$ is lazy then $\tau(\epsilon) = O\left(\Phi^{-2} \cdot \log(|V(G)|/\epsilon)\right)$.*

One method to derive a lower bound on the conductance is the cannonical path technique of Jerrum and Sinclair [8]. This technique essentially states the following simple mincut $\geq$ maxflow fact. If one thinks of a $d$-regular graph as a network where edges have capacity $\Lambda$ and it is possible to transfer one unit of flow between every pair of vertices, then the conductance of the graph is at least $\frac{|V|}{2d\Lambda}$. That is, in order to bound the conductance one can show a valid flow that requires a small value of $\Lambda$ (this is sometimes referred to as the *load* of the flow).

Being a Schrier graph, our graph lends itself to a special type of flow that we now introduce. Let $G = sc(S, X)$ and consider a probability distribution $\mu$ over finite sequences of elements of $S$. For any $x \in X$, the distribution $\mu$ induces a distribution $\mu_x$ of the end points of paths starting at $x$, where the probability of the path $x\overline{s}$ is $\mu(\overline{s})$. Assume first that $\mu_x$ is the uniform distribution over $X$. Then for each $x, y \in X$ we can build the following flow. For any sequence $\overline{s}$ of elements from $S$, assign a flow $\mu(\overline{s})$ from $x$ to the path $x\overline{s}$, and to $y$ for $y\overline{s}$. We get a valid flow for $G$, where the load of the edge $e = (u, us)$ is $2 \cdot \sum_y \sum_x \eta_{x,u,s} = 2 \cdot |X| \cdot \sum_x \eta_{x,u,s}$, with $\eta_{x,u,s}$ being the expected number of occurences of $e$ in a random path $x\omega$ where $\omega$ has distribution $\mu$. The factor of 2 follows since the first and second halfs contribute the same load to $e$.

More generally, assume that that for all $x$ the distrubution $\mu_x$ is $\epsilon$-close to uniform in total variation distance . Then for any vertex $z$, we compare $\mu_y(z)$ and $\mu_x(z)$. We define the same flow from $x$ to $y$ as in the uniform case except that to get a valid flow we multiply the flow in the paths from $x$ to $z$ by $\min(1, \mu_y(z)/\mu_x(z))$, and the flow from $z$ to $y$ by $\min(1, \mu_x(z)/\mu_y(z))$. This will result in a flow of at least $1 - 2\epsilon$ from $x$ to $y$. By scaling back to 1, we get a valid flow, where the load of $e$ is bounded by $(1-2\epsilon)^{-1} \cdot 2 \cdot |X| \cdot \sum_x \eta_{x,u,s}$.

**Lemma 4.** *If $\mu, \mu_x, \Lambda$ are as above, and for every $x \in X$ the distribution $\mu_x$ is $\epsilon$-close to uniform, then $\Lambda \leq (1-2\epsilon)^{-1} \cdot |X| \cdot 2L$, where $L = \max_{s \in S} L(s)$ and $L(s)$ is the expected number of occurences of $s$ in a random sequence with distribution $\mu$.*

*Proof.* Since the load on the edge $e = (u, us)$ is bounded by $(1-2\epsilon)^{-1} \cdot 2 \cdot |X| \cdot \sum_x \eta_{x,u,s}$, it is sufficient to show that $\sum_x \eta_{x,u,s} \leq L$ for every $u, s$. Indeed, consider the process where we start from a randomly chosen $x \in X$ and follow a random sequence from $\mu$. Notice that $\frac{1}{|X|} \cdot \sum_x \eta_{x,u,s}$ is the expected number of times we hit $e$ in this process. Since the initial vertex is chosen according to the stationary distribution, the distribution of the vertex we traverse in the $l$'th move is always uniform. Hence $\sum_x \eta_{x,u,s} = |X| \cdot \frac{1}{|X|} \cdot L(s) \leq L$.

□

From Lemma 4 we get the desired lower bound on the conductance:

$$\Phi \geq \frac{|X|}{2d\Lambda} \geq \frac{1-2\epsilon}{4 \cdot |S| \cdot L}. \tag{3}$$

**Note 5.** *It is possible to improve (3) by a factor of two, if, rather than constructing a valid flow, we assign flow $\mu(\overline{s})$ to the path $x\overline{s}$ for all $x$ and $\overline{s}$. It is easy to see that for every vertex subset $Y \subset X$, the flow from $Y$ to its complement $\overline{Y}$ is at least $|Y| \cdot (|\overline{Y}|/|X| - \epsilon)$.*

Denote by $L(G,\epsilon)$ the minimal $L$ achievable by any distribution on sequences of elements from $S$ such that for every $x \in X$ the distribution of $x\omega$ is $\epsilon$-close to the uniform distiribution. Theorem 3 together with inequality 3 give

**Corollary 6.** $\tau(\epsilon) \leq O(|S|^2 \cdot L(G,1/4)^2 \cdot \log(|X|/\epsilon))$ *whenever the random walk is lazy.*

In order to prove that the composition of elements from $\mathcal{F}_2$ approaches $k$-wise independence quickly we construct the Schreier graph $G_{k,n} = \mathrm{sc}(\mathcal{F}_2, \Omega^{(k)})$, where $\Omega^{(k)}$ is the set of $k$-tuple with $k$ distinct elements from the base set $\Omega = \{0,1\}^n$. It is convenient to think of $\Omega^{(k)}$ as the set of $k$ by $n$ matrices with distinct rows. A basic permutation acts on $\Omega^{(k)}$ by acting on each of the rows.

Our goal now is to define a distribution over sequences of permutations from $\mathcal{F}_2$ with the following properties: (i) the application of a random sequence to any $x \in \Omega^{(k)}$ yields a matrix that is almost uniformly distibuted over $\Omega^{(k)}$ and (ii) the *load* (the expected number of occurences) is small for every $s \in \mathcal{F}_2$. More specifically, we want to show that

$$L(G_{k,n}, 1/4) = \tilde{O}(\frac{kn}{|\mathcal{F}_2|}) = \tilde{O}(\frac{k}{n^2}), \tag{4}$$

which by Corollary 6 proves Theorem 1.

For brevity, we denote $L(G_{k,n},\epsilon)$ by $L(k,n,\epsilon)$. Note that by (1) we have

$$L(k,n,\epsilon) \leq \lceil \log(1/\epsilon) \rceil \cdot L(k,n,1/4). \tag{5}$$

The rest of this section is devoted to proving (4). Here is an overview. A naive way to get a random sequence that will turn any matrix to random would be to go over all its entries one by one and to flip each entry independently with probability half. Such an approach ignores the fact that whenever we apply an element $s \in \mathcal{F}_2$ on the matrix we act simultaneously on all the rows, so independence is highly unlikely. But what if we apply what we call a *characteristic permutatuions*, which is a permutation that flips a bit exactly when a specified set of $a$ other bits have the values $\vec{\nu} = (\nu_1, \nu_2, \ldots, \nu_a)$? Intuitively most of the rows will not be affected by such a permutation. This leads to a way of approximating the naive scheme. Here is how. First notice that since characteristic permutatuions do not belong to $\mathcal{F}_2$ we need to compose elements of $\mathcal{F}_2$ in order to get them. To this end we adapt a theorem of Cleve [4] and show (Appendix A) that any such permutation is a composition of $O(a^2)$ elements from $\mathcal{F}_2$. We start our sequence by a relatively short sequence of elements from $\mathcal{F}_2$ achieving almost 2-wise independence. Therefore, taking a set of $a$ columns for sufficiently large $a$, we get that whp any string $\nu$ of length $a$ cannot occur in more than one row, and we get our required handle on the rows. This is done in Lemma 9. Unfortunatly the value of $a$ needed turns out to be big, making the length of the resulting sequences long. This issue is overcome in Lemmas 10 that bootstraps Lemma 9.

Next, with the benefit of foresight, we point out the following.

**Observation 7.** *In Lemmas 8, 9 and 10 we will present distributions $\mu$ on sequences of elements from $\mathcal{F}_2$ where certain $f \in \mathcal{F}_2$ seem to receive an undue load, as these permutations operate on specified indices (columns) of interest. This is easy to overcome when we simly imagine the lemmas applying over all possible permutations of the indices. Therefore, since there will always be three indicies of interest, we get that the load on any particular permutation in $\mathcal{F}_2$ is at most $O(\lambda/n^3)$ where $\lambda$ is the maximal length of the sequences of $\mu$.*

We turn to the lemmas establishing bounds on the needed load of the sequence distributions.

**Lemma 8.**

$$L(2, n, 1/4) = O(\log n/n^2).$$

*Proof.* Using Observation 7, it is enough to give a distribution over sequences $\omega$ of length $O(n \log n)$ of elements from $\mathcal{F}_2$ that take any initial $2 \times n$ matrix with two distinct rows to a matrix 1/4-close to a uniformly distributed matrix with two distinct rows. Lemma 11 shows that when we take sequences of random elements from $\mathcal{F}_2$ (specified by the random walk) we get the desired distribution.

□

We now get to two lemmas that embed "algorithms" in the construction of the stochastic sequences.

**Lemma 9.** *If* $k \leq 2^{(n-8)/4}$ *then*

$$L(k, n, 1/4) \leq L(2, n, 1/8k^2) + O(k^2 \cdot \log^2 k/n^2).$$

*Proof.* Let $a$ be the integer satisfying $8k^2 \leq 2^a < 16k^2$. We construct a random sequence $\omega$ by starting with $\omega_1$ which is an $L(2, n, 1/8k^2)$ sequence. Given any $k \times n$ matrix $x$ we know that the rows of $x\omega_1$ are $1/8k^2$-close to 2-wise independent. Let $X$ be the expected number of pairs of rows of $x\omega_1$ that coincide in their first $a$ coordinates. Then

$$E[X] \leq \binom{k}{2} \cdot (2^{-a} + \frac{1}{8k^2}) \leq \frac{k^2}{2} \cdot \frac{2}{8k^2} = \frac{1}{8}.$$

Therefore the probability that the first $a$ columns of $x\omega_1$ are do not have distinct rows is at most $\frac{1}{8}$. After $\omega_1$ we perform the following procedure $\omega_2$:

For $i = a+1, \ldots, n$
  For $\alpha \in \{0,1\}^a$
    with probability $\frac{1}{2}$ do $g_{i,\alpha}$,

where $g_{i,\alpha} : \{0,1\}^n \to \{0,1\}^n$ is the permutation that flips the $i$'th coordinate iff $(x_1, \ldots, x_a)$ is equal to $\alpha$. The permutation $g_{i,\alpha}$ is implemented as a concatenation of $O(a^2) = O(\log^2 k)$ basic permutations using lemma 12. If the first $a$ columns of $x\omega_1$ have distinct rows then the last $n - a$ columns of $x\omega_1\omega_2$ have a uniform distribution.

We end the sequence $\omega$ by performing $\omega_3$

For $i = 1, \ldots, a$
  For $\alpha \in \{0,1\}^a$
    with probability $\frac{1}{2}$ do $h_{i,\alpha}$,

where $h_{i,\alpha}$ flips the $i$'th coordinate iff the last $a$ coordinates are equal to $\alpha$. As before $h_{i,\alpha}$ is implemented as a concatenation of $O(\log^2 k)$ basic permutations. After applying $\omega_3$, the first $a$ columns have uniform distribution if all the rows of the last $a$ columns of $x\omega_1\omega_2$ are distinct. Given that the first condition holds, ie that all the rows of the first $a$ columns of $x\omega_1$ are distinct, the second condition fails with probability bounded by $\frac{k^2}{2} \cdot 2^{-a} \leq \frac{1}{16}$. Therefore, for $\omega = \omega_1\omega_2\omega_3$, we have that with probability at least $1 - \frac{1}{8} - \frac{1}{16}$ the distribution of $x\omega$ is uniform. Therefore the

distribution of $x\omega$ is $\frac{3}{16}$-close to uniform. [3] The only condition we have to check is that the first and last $a$ columns are disjoint, ie $2a \leq n$. This is guaranteed if $16k^2 \leq 2^{n/2}$.

The length of the sequence $\omega_2\omega_3$ is bounded by $O(k^2 n \log^2 k)$. By Observation 7 the load is $O(k^2 n \log^2 k/n^3)$.

□

**Lemma 10.** *If $k \leq 2^{(n-16)/4}$ then*

$$L(k, n, 1/4) \leq L(b, n, \epsilon) + O(\frac{k}{n^2} \cdot \log^2 k),$$

*where $b = 2 + \lceil \frac{1}{3} \log k \rceil$ and $\epsilon = \frac{1}{32} \cdot k^{-b-1}$.*

*Proof.* Let $a = 3 + \lceil \log k \rceil$. Since $4a \leq n$, we can partition the columns of the matrix to four sets $C_1, \ldots, C_4$ of size $a$ and the leftover $C$.

We start $\omega$ by $\omega_1$ which is an $L(b, n, \epsilon)$ sequence. For $p \in \{1, 2, 3, 4\}$, $i \notin C_p$ and $\alpha \in \{0,1\}^a$ let $g_{i,\alpha,p} : \{0,1\}^n \to \{0,1\}^n$ be the permutation that flips the $i$th bit of $x$ if $x_{|C_p} = \alpha$, where $x_{|C_p}$ denotes the restriction of $x$ to $C_p$. As before we implement $g_{i,\alpha,p}$ as the concatenation of $O(\log^2 k)$ basic permutations.

Let $\omega_2$ be the following randomized procedure.

For $i \in [n] \setminus (C_1 \cup C_2)$,
  For $\alpha \in \{0,1\}^a$ with probability $\frac{1}{2}$ do $g_{i,\alpha,C_1}$
  For $\beta \in \{0,1\}^a$ with probability $\frac{1}{2}$ do $g_{i,\beta,C_2}$.

We argue that for any $k \times n$ matrix $x$ the distribution of the columns $[n] \setminus (C_1 \cup C_2)$ of $x\omega_1\omega_2$ is uniform with high probability.

Given the matrix $x\omega_1$ we build a bipartite multi-graph $H$ over the sets $V_1$ and $V_2$ where $V_1 = V_2 = \{0,1\}^a$, and where $H$ has $k$ edges, one for each row of the matrix. The edge associated with a row of $x\omega_1$ is between $s_1 \in V_1$ and $s_2 \in V_2$ if its restriction to $C_p$ is $s_p$ for $p = 1, 2$. For perspetive we relate our schema here to the previous lemma. There, we essentially looked at a block of the size of $C_1 \cup C_2$ and went over all possible values to this number of bits, hence a range which is of size $k^2$ instead of $k$ here. In terms of $H$, the claim there was that whp it does not contain any multi edges and for that we needed the pairwise independence of the rows. Here we need a stronger property, namely that $H$ is cycle-free, and this will be possible to show using the stronger condition on $x\omega_1$, namely that it is an almost $b$ wise independent matrix.

We first argue if $H$ is cycle free then the distribution of the columns not in $C_1 \cup C_2$ of $x\omega_1\omega_2$ is uniform. Fix $i$ to be the column of interest. Let $r_{\alpha,i}$ and $s_{\beta,i}$ be the $2 \cdot 2^a$ random bits used to generate the part of $\omega_2$ that is responsible for column $i$. For any edge $e = (\alpha, \beta)$

$$(x\omega_1\omega_2)_{e,i} = (x\omega_1)_{e,i} \oplus r_{\alpha,i} \oplus s_{\beta,i}. \tag{6}$$

For a given $x\omega_1$ the probability of having a certain fixed column $v$ as the $i$'th column is therefore the number of solutions in the varaibles $r_{\alpha,i}, s_{\beta,i}$. This number is the same for every $v$ if the linear

[3] This argument actually proves that $x\omega$ is $\frac{3}{16}$-close to the uniform distribution on $\Omega^k$. However, the uniform distribution on $\Omega^k$ and $\Omega^{(k)}$ are $o(1)$-close.

system (6) is of full rank. It is easy to see that the matrix defining this system is exactly the incidence matrix of $H$. We now only need to use the well known fact that this matrix has a full rank iff $H$ does not contain a cycle.

We now turn to show that $H$ is cycle free whp. Recall that $H$ is a random bipartite graph with $k$ edges that is close to $b$-wise independent in the sense that any event in which at most $b$ edges are involved happens with almost the same probability it happens in a completely random graph with $k$ edges. Let $E_l$ be the expected number of $l$-cycles for $2 \leq l < b$ in the graph. We have $k \cdot (k-1) \cdots (k-l+1)$ ways to choose the $l$ edges of the cycle. The edges connect properly with probability at most $2^{-al} + \epsilon$. Thus

$$E_l \leq k^l \cdot (2^{-al} + \epsilon) \leq 8^{-l} + \frac{1}{32} \cdot k^{l-b-1}.$$

For cycles longer than $b$ we cannot use the $b$-wise independence in the same way. Instead we bound the probabiliy of having $b$ edges creating a path to get a bound on the expected number of all cycles of length $\geq b$ which is $k^b \cdot (2^{-a(b-1)} + \epsilon) \leq k \cdot 8^{-(b-1)} + \frac{1}{32} \leq \frac{3}{64}$. Therefore the total number of cycles is bounded by

$$\frac{3}{64} + \sum_{l=2}^{b-1} 8^{-l} + \frac{1}{32} \cdot k^{l-b-1} \leq \frac{1}{8},$$

for a sufficiently large $k$.

As in the proof of lemma 9, we continue with the sequence $\omega_3$, which uses the two column sets $C_3$ and $C_4$ to change the columns $C_1$ and $C_2$ to the uniform distribution. Assume that $H$ had no cycle and therefore that $\omega_2$ succeeded. Then the graph $H'$ formed by the $C_3$ and $C_4$ columns of $x\omega_1\omega_2$ has uniform distribution over all bipartite graphs with vertex sets of size $2^a$ and $k$ edges. Therefore the probability that $H'$ has a cycle is certainly smaller than $\frac{1}{8}$, and we get that with probability at least $\frac{3}{4}$ the matrix $x\omega_1\omega_2\omega_3$ is uniform. Therefore its distance from the uniform distribution is $\leq \frac{1}{4}$ (see footnote 3). Yet again, by Observation 7 we conclude the contribution of $\omega_1$, $\omega_2$ to $L$ is $O(k \log^2 k/n^2)$ and we are done. □

*Proof.* (of Theorem 1)

We combine lemmas 8, 9 and 10 with inequality (5) to get

$$\begin{aligned}
L(k,n,1/4) &\leq O(L(b,n,1/4)\log^2 k + \frac{k}{n^2}\log^2 k) \\
&\leq O(L(2,n,1/8b^2)\log^2 k + \frac{b^2}{n^2}\log^2 b \log^2 k + \frac{k}{n^2}\log^2 k) \\
&\leq O((\log n \log^2 k \log\log k + \log^4 k \log\log^2 k + k\log^2 k)/n^2) \\
&\leq O((\log n + k)\log^2 k/n^2).
\end{aligned}$$

By corollary 6, the mixing time of $G_{n,k}$ is bounded by

$$\tau(2^{-nk}) = O(n^6 \cdot ((\log n + k)\log^2 k/n^2)^2 \cdot nk) = O((\log n + k)^2 \cdot \log^4 k \cdot n^3 k) = \tilde{O}(n^3 k^3).$$

□

## 5 Mixing with logarithmic width

As before we construct the Schreier graph $G_{k,n} = \mathrm{sc}(\mathcal{F}_w, \Omega^{(k)})$.

**Lemma 11.** *For all $w \geq 1$ the mixing time of $G_{2,n}$ is $O(n \log n)$.*

*Proof.* Represent the state space of the walk by two vectors $(s, u)$, where $s$ is the first row of the matrix and $u$ is the mod 2 sum of the two rows. We describe the way we move in one step of the random walk in this new representation. We choose a coordinate $i$ at random, and then choose two bits *indpendently* $\alpha_s, \alpha_u$ with probability $1/2$ and $p_l = (1 - \prod_{j=1}^{w}(1 - \frac{l}{n-j}))/2$ respectively, where $l$ is the number of ones in $u$ not counting the $i$th bit. We then XOR to the $i$th bit of $s$ and $u$ the bits $\alpha_s$ and $\alpha_u$ respectively. To see that this is indeed the resulting walk we observe the fact that if $s$ and $t$ differ in one of the bits at which the random function $h$ look at, then the value of the $i$th coordinate of $u$ and of $t$ change independently with propability half. Otherwise they change simultaniously with probability $1/2$.

The $u$-component of this walk is a variant of the so called Aldous cube, and by the comment at the end of [3] it follows that this walk mixes in $O(n \log n)$ time. We are left to show that in this time the walk on both components mixes. The way to see it is to notice that in $O(n \log n)$ time the event $B$ where the indices $i$ assume all possible values in $1, 2, \ldots, n$ (coupon collector) occurs whp. Even conditioning on $B$ the walk on $u$ gets close to uniform after $O(n \log n)$ steps. We now observe that the walk on $s$ conditioned on any set of indices $i$ (satifying $B$) and on the bits $\alpha_u$, achieves the exact uniform distribution, since the bits $\alpha_s$ are independent of the bits $\alpha_u$ and indices $i$. □

*Proof.* (of Theorem 2)

To prove Theorem 2 it is enough to show that the mixing time of $G_{k,n}$ is $O(n \cdot \log n \cdot (\log k + \log n))$. Consider a length $T = T_1 + T_2$ random walk $\omega = \omega_1\omega_2$ on $G_{k,n}$, where $|\omega_1| = T_1$ and $|\omega_2| = T_2$. We will choose $T_1$, $T_2$ so that for any $x \in \Omega^{(k)}$, the matrix $x\omega_1$ will be close to 2-wise independent, and the distribution of the matrix $x\omega$ will be close to uniform on $\Omega^k$, and so close to $\Omega^{(k)}$. We let $T_1 = cn \log n \cdot (\log k + \log n)$ for some absolute constant $c$ and $T_2 = 2n \log n$ and show the claimed properties. By Lemma 11 we know that the mixing time of $G_{2,n}$ is $O(n \log n)$. Therefore by 1 we can choose $c$ so that after a length $T_1$ walk in $G_{2,n}$, we are $\delta$-close to the uniform distribution, for $1/100k^2T_2$.

Let $\omega_2 = g_1 g_2 \cdots g_{T_2}$, where $g_t = f_{i_t, J_t, h_t} \in \mathcal{F}_w$. Given any $x \in \Omega^{(k)}$, we know that the rows of the matrix $x\omega_1$ are $\delta$-close to 2-wise independent. We argue that the distribution of $x\omega_1\omega_2$ is close to uniform on $\Omega^k$. Again, by coupon collector argument we know that the event that the indices $i$ will not assume all possible values after $T_2$ steps is at most $\epsilon_1 = 1/100$. Instead of just proving that the distribution of $x\omega$ on $\Omega^k$ is close to uniform, we prove something stronger. During the walk $\omega_2$, at step $t$ we change column $i_t$. Let $C_t \in \{0,1\}^k$ be the new value of column $i_t$. We prove that the distribution of $C = (C_1, \ldots, C_{T_2})$ on $\{0,1\}^{kT_2}$ is close to uniform. We claim that conditioned on any specific values of $\overline{i} = i_1, \ldots, i_{T_2}$ and $\overline{J} = J_1, \ldots, J_{T_2}$ such that the index set $\{\overline{i}\}$ is $[n]$, the distribution of $C$ is $\epsilon_2$-close to uniform on $\{0,1\}^{kT_2}$, where $\epsilon_2 = 1/5$.

Once we prove this, we claim that $d(x\omega, U) \leq \epsilon_1 + \epsilon_2$, where $U$ is the uniform distribution over $\Omega^{(k)}$. First note that given $\overline{i}, \overline{J}$ satisfying $\{\overline{i}\} = [n]$ there are $n$ times $t_1, \ldots, t_n$ such that $t_l$ is the last occurrence of $l$ in the sequence $\overline{i}$. Then $x\omega = (C_{t_1}, \ldots, C_{t_n})$ which means that $x\omega$ is just a

marginal of $C$ and therefore is $\epsilon$-close to uniform on $\Omega^k$. Therefore, for every subset $A \subset \Omega^k$,

$$\begin{aligned}\Pr[x\omega \in A] - \frac{|A|}{2^{kn}} &= \sum_{\overline{i},\overline{J},\ \{\overline{i}\}\neq[n]} (\Pr[x\omega \in A|\overline{i},\overline{J}] - \frac{|A|}{2^{kn}}) \cdot \Pr[\overline{i},\overline{J}] + \\ &\quad \sum_{\overline{i},\overline{J},\ \{\overline{i}\}=[n]} (\Pr[x\omega \in A|\overline{i},\overline{J}] - \frac{|A|}{2^{kn}}) \cdot \Pr[\overline{i},\overline{J}] \\ &\leq \epsilon_1 + \epsilon_2 \leq \frac{1}{4}.\end{aligned}$$

We still have to prove that for any $\overline{i}, \overline{J}$ satisfying $\{\overline{i}\} = [n]$, the distribution of $C$ is $\epsilon_2$-close to $U'$ which is the uniform distribution on $\{0,1\}^{kT_2}$. For the following argument, fix the value of $\overline{i}, \overline{J}$. We would like to estimate $\Pr[C = y]$ for some $y = (y_1, \ldots, y_{T_2}) \in \{0,1\}^{kT_2}$, when the probability is taken on the possible choices of $\omega_1$ and of the random permutations $\overline{h} = h_1, \ldots, h_{T_2}$. Since $\overline{i}, \overline{J}$ are known, for any $t = 1, \ldots, T_2$ we can determine which of the $w$ indices in $J_t$ refer to columns in $x\omega_1$ and which refer to columns of $C$. Let $A$ be the event that for all times $t$, the matrix referred to by $J_t$ has distinct rows. Therefore, given $\omega_1 = \alpha$ and $C = y$ we can determine if $A$ happened. Let $S_y$ be the set of all $\alpha$ such that $A$ holds for $\alpha, y$. Then

$$\Pr[C = y|\omega_1 \in S_y] = \prod_{t=1}^{T_2} \Pr[C_t = y_t|\omega_1 \in S_y, C_{t'} = y_{t'} \text{ for all } t' < t] = 2^{-kT_2}.$$

We argue that for most values of $y$, the function $f(y) = \Pr[\omega_1 \notin S_y]$ is small. Consider $y \in \{0,1\}^{kT_2}$ picked uniformly at random, and assume that $w'$ of the $w$ indices in $J_t$, refer to columns in $x\omega_1$. Since $x\omega_1$ is $\delta$-close to 2-wise independent and since $y$ is uniformly distributed, the probability that any two rows of this matrix are identical is bounded by $(2^{-w'} + \delta) \cdot 2^{-(w-w')} \leq 2^{-w} + \delta$. Therefore, the expected value of $f(y)$ is bounded by $(2^{-w} + \delta) \cdot T_2 \cdot k^2/2 \leq 1/100$. Therefore the size of the set of bad $y$'s, $Y_B = \{y \in \{0,1\}^{kT2} : f(y) > 1/10\}$ is at most $2^{kT_2}/10$. If $y \notin Y_B$ then $\Pr[C = y] = \Pr[C = y|\omega_1 \in S_y] \cdot \Pr[\omega_1 \in S_y] \geq \frac{9}{10} \cdot 2^{-kT_2}$. From the last two it easily follows that $d(C, U') \leq \frac{1}{5} = \epsilon_2$. □

# 6 More on Motivation, Cryptography and Possible Extensions

A principle motivation for this work is the philosophy behind the construction of "permutation generators" such as DES and its successors. The goal is that the permutation generated from a random key should look like a randomly chosen permutation, when examined by a computationally limited adversary; this property is called "pseudo-randomness". The idea used by DES is to start with a very simple function generator $G$, and then compose functions independently and randomly chosen from $G$. (Actually, in order to keep the key short, the functions are not chosen independently, but we will ignore this for now.) Because the adversary is allowed much more time than was taken to compute the function, (almost) $k$-wise independence is neither necessary nor sufficient in order to achieve pseudo-randomness. Regardless, k-wise independence is a very natural measure of randomness, and one appealing question is what can (almost) k-wise independence tell us about pseudo-randomness.

Here is one possible conjecture. Let us assume that the generator $G$ we start with is such that each possible permutation is "simple", where "simple" might mean that each output bit depends on a

constant number of input bits. Say that $T$ compositions from $G$ suffice to achieve almost 4-wise independence. Then we conjecture that $T$ compositions suffice to achieve pseudo-randomness. Of course proving this would show P different from NP, so this is unlikely. The conjecture is, however, susceptible to disproof.

Why do we choose "4-wise" in the above conjecture? For one thing, it is not hard to find examples where 3-wise is not good enough. Also, there is a theorem – proven using the classification of finite simple groups – that any collection of permutations satisfying 4-transitivity will, when composed together, eventually yield at least the alternating group [2, 6].

## References


[1] D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs. http://stat-www.berkeley.edu/users/aldous/RWG/book.html.

[2] Peter J. Cameron. *Permutation groups*, volume 45 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999.

[3] F. R. K. Chung and R. L. Graham. Stratified random walks on the $n$-cube. *Random Structures Algorithms*, 11(3):199–222, 1997.

[4] R. Cleve. Complexity theoretic issues concerning block ciphers related to D.E.S. In A. Menezes and S. Vanstone, editors, *Advances in Cryptology - CRYPTO '90 Proceedings, Lecture Notes in Computer Science*, volume 537, pages 530–544. Springer-Verlag, 1990.

[5] Don Coppersmith and Edna Grossman. Generators for certain alternating groups with applications to cryptography. *SIAM J. Appl. Math.*, 29(4):624–627, 1975.

[6] John D. Dixon and Brian Mortimer. *Permutation groups*, volume 163 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

[7] W. T. Gowers. An almost $m$-wise independent random permutation of the cube. *Combin. Probab. Comput.*, 5(2):119–130, 1996.

[8] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.

[9] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989.

[10] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1(4):351–370, 1992.


## A An Adaptation of Cleve's Lemma

We showing that *characteristic permutatuions* are easy to construct. This follows as a special case of the following lemma, which is an adaptation of Theorem 5 in Cleve [4] to our setting.

**Lemma 12.** *Let $h$ be a boolean function on $s \leq n-3$ variables expressed as a circuit using fanin-2 xor/and gates at the nodes and variables, with zero/one constants at the leaves. Let $d$ be the depth of the circuit. Also, let $i$ be disjoint from $\{1, \ldots, s\}$. Then we can concatenate $O(4^d)$ basic $\mathcal{F}_2$ permutations to form the permutation $g_{i,h}$ mapping*

$$(x_1, \ldots, x_n) \to (x_1, \ldots, x_{i-1}, x_i \oplus h(x_1, \ldots, x_s), x_{i+1}, \ldots, x_n).$$

*Proof.* By induction on $d$. If $d = 0$ then $h$ is a variable or a constant and therefore $g_{i,h}$ is already a basic permutation. If $h = h_1 \oplus h_2$ then $g_{i,h} = g_{i,h_1} \circ g_{i,h_2}$. Otherwise $h = h_1 \wedge h_2$. It can be verified that

$$g_{i,h} = f_{i,j,k} \circ g_{j,h_1} \circ f_{i,j,k} \circ g_{k,h_2} \circ f_{i,j,k} \circ g_{j,h_1} \circ f_{i,j,k} \circ g_{k,h_2},$$

where $f_{i,j,k}$ is the basic permutation mapping $(x_1, \ldots, x_n)$ to $(x_1, \ldots, x_i \oplus x_j \wedge x_k, \ldots, x_n)$ and $j \neq k$ are two coordinates disjoint from $\{1, 2, \ldots, s\} \cup \{i\}$.

The length of the sequence $l(d)$ satisfies the recurrence $l(d) \leq 4 + 4 \cdot l(d-1)$ for $d > 0$ and $l(0) = 1$ and therefore $l(d) = O(4^d)$. □