

Simple Permutations Mix Even Better

Alex Brodsky
Department of Applied Computer Science
University of Winnipeg
abrodsky@acs.uwinnipeg.ca

Shlomo Hoory*
Haifa Research Laboratory
IBM
shlomoh@il.ibm.com

December 25, 2006

Abstract

We study the random composition of a small family of $O(n^3)$ simple permutations on $\{0, 1\}^n$. Specifically we ask how many randomly selected simple permutations need be composed to yield a permutation that is close to k -wise independent. We improve on the results of Gowers [13] and Hoory et al. [14] and show that it suffices to compose $\min(O(n^3 k^2), \tilde{O}(n^2 k^2))$ random permutations from this family for any $n \geq 3$ and $k \leq 2^n - 2$. The \tilde{O} notation suppresses a polylogarithmic factor in k and n .

Keywords: Mixing-time, k -wise independent permutations, cryptography, multicommodity flow, reversible computation.

A naturally occurring question in cryptography is how well the composition of simple permutations drawn from a simple distribution resembles a random permutation. Although such constructions are a common source of security for block ciphers like DES and AES, their mathematical justification (or lack thereof) is troubling.

This motivated the investigation of Hoory et al. [14] who considered the notion of almost *k -wise independence*. Namely, that the distribution obtained when applying a permutation from a given distribution to any k distinct elements is almost indistinguishable from the distribution obtained when applying a truly random permutation. Therefore, the question is how close is the composition of T random simple permutations to k -wise independent?

Another motivation is a fundamental open problem in the theory of expanding graphs, whose solution was recently announced by Kassabov [16]. Namely, the problem of constructing a constant degree expanding Cayley graph of the symmetric group. A possible relaxation of this problem is to ask whether one can find a small set of simple permutations such that its action on k points yields an expanding graph.

It turns out that these two problems reduce to bounding the mixing time and the spectral gap of the random walk on the *same* graph. This walk, P , is defined on the state space of k -tuples of distinct elements from the n -dimensional binary cube. In each step it randomly selects a simple permutation and applies it to each of the k elements at its current position. The mixing time, $\tau(\epsilon)$, is the number of steps needed to come ϵ -close to the uniform distribution (in total variation

*Research is supported in part by an NSERC grant and a PIMS postdoctoral fellowship.

distance), and the spectral gap, $\text{gap}(P)$, is the difference between the two largest eigenvalues of P 's transition matrix.

Following the construction of DES, and previous work by Gowers [13] and Hoory et al [14], we consider the class of *width 2 simple permutation*, denoted Σ . The action of such a permutation on an element of the n -dimensional binary cube is to XOR a single coordinate with a Boolean function of 2 other coordinates; there are $16n(n-1)(n-2)$ such permutations.

These problems were first considered by Gowers [13] who gave an $\tilde{O}(n^3k(n^2+k)(n^3+k))$ bound on the mixing time, by lower bounding the spectral gap $1/\text{gap}(P) = \tilde{O}(n^2(n^2+k)(n^3+k))$. Throughout the paper we use the \tilde{O} notation to suppress a polylogarithmic factor in n and k . Subsequently, Hoory et al. [14] improved the bound on the mixing time to $\tilde{O}(n^3k^3)$ by proving that $1/\text{gap}(P) = \tilde{O}(n^2k^2)$. Both results were achieved using the *canonical paths* technique, and neither result applies for $k > 2^{n/2}$. Using the comparison technique, in conjunction with the theory of reversible computation, we give better bounds for all values of k up to the largest conceivable value, $k = 2^n - 2$.

Theorem 1. $\tau(\epsilon) = \tilde{O}(n^2k^2 \cdot \log(1/\epsilon))$, as long as $k \leq 2^{n/50}$.

Theorem 2. $1/\text{gap}(P) = O(n^2k)$ for all $k \leq 2^n - 2$.

Using the well known connection between the mixing time and the spectral gap Theorem 2 implies:

Corollary 3. $\tau(\epsilon) = O(n^2k \cdot (nk + \log(1/\epsilon)))$ for all $k \leq 2^n - 2$.

The proofs of both Theorems are based on the comparison technique for Markov chains [8]. To prove Theorem 2 we compare the random walk P either to a Glauber dynamics Markov chain or to the random walk on the alternating group using 3-cycles. To prove Theorem 1 we observe that after a short preamble the random walk P is almost surely in a “generic” state. Consequently, it suffices to bound the mixing time of a Markov chain restricted to “generic” states. To this end we again employ the comparison technique, but with a better comparison constant. In all cases we construct the multicommodity flows required by the comparison technique using ideas from the theory of reversible computation.

It follows from [14, 18] that these results apply also in the more general setting of adaptive adversaries (see references for a definition).

Corollary 4. *Let T be the minimal number of random compositions of independent and uniformly distributed permutations from Σ needed to generate a permutation which is ϵ -close to k -wise independent against an adaptive adversary. Then $T = \tilde{O}(n^2k^2 \cdot \log(1/\epsilon))$ for $k \leq 2^{n/50}$, and $T = O(n^2k \cdot (nk + \log(1/\epsilon)))$ for $k \leq 2^n - 2$.*

After submission of this paper, in a private communication of the second author with Yuval Peres, we came with a promising direction for further improving the upper bound on the mixing time of P to $\tilde{O}(n^2k)$. The idea is to bound the log-Sobolev constant α of the Markov chain P , rather than its spectral gap. The easy fact that should enable one to bound $\alpha(P)$ is that the comparison constant of comparing P to \tilde{P} applies to log-Sobolev constants as well. Thus the only missing part is to determine the log-Sobolev constants of the two Markov chains used as \tilde{P} . One more benefit to gain from this approach is the elimination of the tedious calculations in Section 4. For more about the log-Sobolev constant of Markov chains and the comparison technique for log-Sobolev constants see [9] or [19].

1 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 ϵ -close to k -wise independent if for every $(x_1, \dots, x_k) \in X^{(k)}$ the distribution of $(f(x_1), \dots, f(x_k))$ is ϵ -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

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

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* $G = \text{sc}(S, X)$ is defined by $V(G) = X$ and $E(G) = \{(x, xs) : x \in X, s \in S\}$. For a sequence $\omega = (s_1, \dots, s_\ell) \in S^\ell$ we denote $x\omega = xs_1 \cdots s_\ell$, and we sometimes refer by $x\omega$ to the walk $x, xs_1, \dots, xs_1 \cdots s_\ell$.

The *random walk* X_0, X_1, \dots associated 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 π is stationary for this Markov chain. If G is connected and not bipartite, we know that given any initial distribution of X_0 , the distribution of X_t tends to the uniform distribution. The mixing time of G is $\tau(\epsilon) = \max_{v \in V(G)} \min\{t : d(P^{(t)}(v, \cdot), \pi) < \epsilon\}$, where $P^{(t)}(v, \cdot)$ is the probability distribution of X_t given that $X_0 = v$. It is not hard to prove (see [1, Lemma 20]) that

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

Let $1 = \beta_0 \geq \beta_1 \geq \dots \geq \beta_{|V(G)|}$ be the eigenvalues of the transition matrix P . We say that this random walk is *lazy* if for some constant $\delta > 0$ we have $P_{vv} \geq \delta$ for all $v \in V(G)$. We denote the spectral gap $1 - \beta_1$ of the Markov chain P by $\text{gap}(P)$.

Two fundamental results relating the spectral gap of a Markov chain to its mixing time are the following:

Theorem 5. ([11, Proposition 3]) *If the random walk on G is lazy then $\tau(\epsilon) = O(\log(|V(G)|/\epsilon) / \text{gap}(P))$.*

Theorem 6. ([21, Proposition 1.ii] or [1, Chapter 4]) *For any time reversible Markov chain P and $\epsilon > 0$, $\text{gap}(P) = \Omega(\log(1/2\epsilon) / \tau(\epsilon))$.*

2 Composing simple permutations

Another building block that we use are results on reversible computation that enables us to compose simple permutations to construct permutations that are easier to work with. A classical result of Coppersmith and Grossman [6] is that for $n > 3$ the set of width 2 simple permutations generates exactly the alternating group A_n . Thus, all compositions must be even permutations.

Formally, we define the set of width w simple permutations, Σ_w , as the set of permutations $f_{i,J,h}$ where $i \in [n]$, $J = \{j_1, \dots, j_w\}$ is a size w ordered subset of $[n] \setminus \{i\}$, and h is a Boolean function on $\{0, 1\}^w$. The permutation $f_{i,J,h}$ maps $(x_1, \dots, x_n) \in \{0, 1\}^n$ to $(x_1, \dots, x_{i-1}, x_i \oplus h(x_{j_1}, \dots, x_{j_w}), x_{i+1}, \dots, x_n)$. We are primarily interested in width 2 simple permutations, and denote $\Sigma = \Sigma_2$.

Theorem 7. (Barenco et al. [3]) *The permutation that flips the n -th bit of input x if and only if the first w bits of x are 1 can be implemented as a composition of $O(w)$ permutations from Σ , as long as $w \leq n - 2$.*

Theorem 8. (Brodsky [4]) *For any distinct $x, y, z \in \{0, 1\}^n$, one can compose $O(n)$ permutations from Σ to obtain the 3-cycle (xyz) .*

A length ℓ *implementation* of the permutation σ is a sequence of permutations $\sigma_1, \dots, \sigma_\ell$ from Σ whose composition is σ . Theorem 8 gives a length $O(n)$ implementation for 3-cycles. We would like to use this implementation to construct a multicommodity flow with low load on all edges. However, Theorem 8 does not guarantee this. We solve this problem by randomizing the implementation, enabling us to prove a stronger theorem.

A length ℓ *randomized implementation* of the permutation σ is a sequence of *random* permutations $\sigma_1, \dots, \sigma_\ell$ from Σ whose composition is σ . In Theorem 9 we give a randomized implementation for 3-cycles, such that applying any prefix $\sigma_1 \cdots \sigma_\ell$ of the randomized implementation of a uniformly random 3-cycle (xyz) to x yields a string that looks random. Namely, its *min-entropy* $H_\infty(\cdot)$ is high, which is the minimum amount of information revealed when exposing the value of a random variable X , that is $H_\infty(X) = \min_\chi(-\log_2(\Pr[X = \chi]))$.

Theorem 9. *Let $x, y, z \in \{0, 1\}^n$ be uniformly distributed and distinct. Then there is a length $L = O(n)$ randomized implementation $\sigma_1 \cdots \sigma_L$ of the 3-cycle (xyz) such that for all $\ell \in [L]$ the min-entropy of $(x\sigma_1 \cdots \sigma_{\ell-1}, \sigma_\ell)$ (which is a random variable on $\{0, 1\}^n \times \Sigma$) is at least $\log_2(2^n \cdot n^3) - O(1)$.*

Note, this implies that the min-entropy of the marginals is big, i.e., $H_\infty(x\sigma_1 \cdots \sigma_{\ell-1}) \geq n - O(1)$ and $H_\infty(\sigma_\ell) \geq \log_2(n^3) - O(1)$.

3 Proof of Theorem 2

In order to prove that the composition of random permutations from Σ approaches k -wise independence quickly we construct the Schreier graph $G_{k,n} = \text{sc}(\Sigma, X^{(k)})$, where $X^{(k)}$ is the set of k -tuple with k distinct elements from the base set $X = \{0, 1\}^n$. It is convenient to think of $X^{(k)}$ as the set of k by n binary matrices with distinct rows. A simple permutation acts on $X^{(k)}$ by acting on each of the rows. Then P is the transition matrix of the random walk on $G_{k,n}$. We prove that the random walk on this graph mixes rapidly.

To prove that $1/\text{gap}(P) = O(n^2k)$, we first observe that $\text{gap}(P)$ is monotone nonincreasing in k . This follows from the fact that the graph $G_{k+1,n}$ is a lift of $G_{k,n}$ and therefore inherits the spectrum of $G_{k,n}$. To see this, observe that any eigenfunction of $G_{k,n}$, can be lifted to an eigenfunction on $G_{k+1,n}$, where the value of the latter on some $k+1$ by n matrix is the value of the former on the matrix obtained by deleting the last row. The eigenvalues of these two eigenfunctions is the same. In light of this observation, it is sufficient to prove the following two lemmas:

Lemma 10. $1/\text{gap}(P) = O(n^2 \cdot 2^n)$ for $k = 2^n - 2$.

Lemma 11. $1/\text{gap}(P) = O(n^2k)$ for $k \leq 2^n/3$.

We obtain the lower bound on the spectral gap of P using the comparison technique [8]. This technique enables one to lower bound $\text{gap}(P)$ by $\text{gap}(\tilde{P})/A$, where \tilde{P} is some other Markov chain,

and A is the comparison constant. In our case, all chains are walks on regular graphs. An upper bound on A is obtained by constructing a multicommodity flow on the underlying graph of P . The flow flows a unit between all pairs of endpoints of edges of \tilde{P} such that the flow through each edge of P is small. To prove Lemmas 10 and 11, we compare P to two different Markov chains. We start with the first Lemma.

Proof. (of Lemma 10) For $k = 2^n - 2$, the state space of P comprises all even permutations of $\{0, 1\}^n$. Let \tilde{P} be a Markov chain on this state space, where in each step we pick three distinct elements of the cube $x, y, z \in \{0, 1\}^n$ and perform the permutation (xyz) . It follows from a result of Friedman [12], that $1/\text{gap}(\tilde{P}) = \Theta(2^n)$. Therefore, it is sufficient to prove that the comparison constant of P to \tilde{P} is $O(n^2)$.¹

To bound the comparison constant A , we need to construct a multicommodity flow f in $G_{k,n}$ that flows a unit between every two matrices M, M' such that $\tilde{P}(M, M') > 0$. Since the chains P and \tilde{P} correspond to random walks on regular graphs with degrees $d = \Theta(n^3)$ and $\tilde{d} = \Theta(2^{3n})$ respectively, the formula given in [8, Theorem 2.3] reduces to:

$$A = (d/\tilde{d}) \cdot \max_{(N, N') \in E(G_{k,n})} \left\{ \sum_{\gamma: (N, N') \in \gamma} f(\gamma) \cdot |\gamma| \right\}. \quad (2)$$

Let M, M' be two matrices such that $\tilde{P}(M, M') > 0$. Then M' can be obtained by applying some 3-cycle (xyz) to M . Recall that the randomized implementation given by Theorem 9 induces a probability distribution on the length L sequences of permutations from Σ whose composition is (xyz) . Such a distribution naturally translates to a distribution on length L paths from M to M' . We obtain a unit flow from M to M' by flowing through each such path γ an amount equal to its probability. We claim that the multicommodity flow obtained by repeating this process for all M, M' pairs satisfying $\tilde{P}(M, M') > 0$ yields a small comparison constant.

Since $|\gamma| \cdot (d/\tilde{d}) = \Theta(n \cdot |\Sigma|/2^{3n})$ for all paths γ with non-zero flow, the problem of bounding the sum in (2) reduces to bounding the total flow through a given edge $e \in E(G_{k,n})$. Let $\gamma = (M_0, \dots, M_L)$ be a path from M_0 to M_L , where M_L is obtained from M_0 by applying the 3-cycle (xyz) . Assume further that γ goes through the edge e at the ℓ -th step, and that x is the r -th row of M . For any of the $\Theta(2^{4n} \cdot n)$ possible assignments to x, y, z, ℓ, r , we can determine the distribution of the r -th row of the matrices M_0, \dots, M_L . In particular, the probability that $(M_{\ell-1}, M_L)$ is equal to e is bounded by the probability that they coincide in their r -th row. By Theorem 9, in average over all assignments to x, y, z, ℓ, r , this probability is $O(1/2^n |\Sigma|)$. Putting it all together yields that, up to a constant factor, the comparison constant A is bounded $(n \cdot |\Sigma|/2^{3n}) \cdot (2^{4n} \cdot n) \cdot (1/2^n |\Sigma|) = n^2$, as claimed. \square

Proof. (of Lemma 11) Let \tilde{P} be the a Markov chain on the same state space as P , which is the k by n binary matrices with distinct rows. If the current state of \tilde{P} is the matrix M , then the next state is determined by picking a row $r \in \{1, \dots, k\}$ and setting it to a random new value that is distinct from all other $k - 1$ rows. The process \tilde{P} is the Markov chain of coloring the clique on k vertices with 2^n colors defined in [15, section 4.1]. Proposition 4.5 therein bounds its mixing time by $\tilde{\tau}(\epsilon) = O(k \log(k/\epsilon))$ as long as $k \leq 2^n/3$. Setting $\epsilon = 1/4k$ in Theorem 6 implies

¹Alternately, one can define a transition of \tilde{P} as performing two random transpositions (not necessarily disjoint) and use a result of Diaconis and Shahshahani [10] that $1/\text{gap}(\tilde{P}) = \Theta(2^n)$.

that $\text{gap}(\tilde{P}) = \Omega(1/k)$. Therefore, as in the proof of Lemma 10, it is sufficient to prove that the comparison constant of P to \tilde{P} is $O(n^2)$.

Given matrices M, M' such that $\tilde{P}(M, M') > 0$, we know that M' is obtained from M by changing the value of some row r from x to y . To construct paths from M to M' , we note that M' can be obtained by applying the 3-cycle (xyz) to M for any $z \in \{0, 1\}^n$ that is distinct from all rows of M, M' . We choose z at random from the $2^n - (k+1)$ allowed values. As in the proof of Lemma 10, the randomized implementation of (xyz) , given by Theorem 9, defines a distribution on paths from M to M' and therefore a multicommodity flow. We turn to bound the comparison constant, given by (2).

As before, $|\gamma| \cdot (d/\tilde{d}) = \Theta(n \cdot |\Sigma|/k2^n)$ for all γ with non-zero flow, and it suffices to bound the flow through some edge $e \in E(G_{k,n})$. We enumerate over the choices of the position ℓ , row r and distinct x, y , which make a total of $\Theta(nk2^{2n})$ possible values. Again we apply Theorem 9 to argue that in average, the probability of agreement with e is bounded by $O(1/|\Sigma|2^n)$.² Therefore, up to a constant factor, $A = (n \cdot |\Sigma|/k2^n) \cdot (nk2^{2n}) \cdot (1/|\Sigma|2^n) = n^2$, as claimed. \square

4 Proof of Theorem 1

In light of inequality (1), it is sufficient to prove that $\tau(1/4) = \tilde{O}(n^2k^2)$. The outline of the proof is the following. We start by introducing the notion of a *generic* matrix, and as suggested by the name, most matrices are generic. The proof then proceeds by arguing that after a short random walk almost surely all matrices encountered are generic. Therefore, it is sufficient to bound the mixing time of a walk that is restricted to generic matrices. For such a walk, we can compare the chain to a chain defined *only* on generic matrices and achieve a much smaller comparison constant. This yields the desired bound, $\tilde{O}(n^2k^2)$.

Let $w = 10 \cdot (\log k + \log n)$. By assumption, we have $w \leq n/4$ for a sufficiently large n , and we set $p = \lceil n/2w \rceil$. Let C_1, \dots, C_p, C be a partition of $[n]$ such that $|C_i| = w$ for $i = 1, \dots, p$ and $|C| = n - pw$. Consequently, $n/4 \leq n/2 - w < |C| \leq n/2$.

We say that a k by n matrix is *generic*, if for all $j \in [p]$, its restriction to C_j has distinct rows. It is not difficult to check that a uniformly distributed matrix M is almost surely generic. In fact, it is sufficient that the rows of M are 2^{-w} -close to 2-wise independent, since then the probability that M is not generic is bounded by p times the probability that the restriction of M to C_j doesn't have distinct rows. This yields the bound $p \cdot \binom{k}{2} \cdot (2 \cdot 2^{-w}) = o(1/n^3k^3)$ and implies the following lemma:

Lemma 12. *If the rows of a random k by n matrix M are 2^{-w} -close to 2-wise independent, then M is generic with probability $1 - o(1/n^3k^3)$.*

It follows from a result of Chung and Graham about the mixing time of the ‘‘Aldous Cube’’ [5], that the number of steps needed to come close to 2-wise independence, which is the same as the mixing time of $G_{2,n}$, is $O(n \log n)$. This is stated in the following lemma (whose proof is deferred to Section 6).

Lemma 13. *For all $w \geq 1$ the ϵ mixing time of the Schreier graph $sc(\Sigma_w, X^{(2)})$ is $O(n \log n \log(1/\epsilon))$.*

²One should note that z is uniformly distributed only over $2^n - (k+1) > 2^{n-1}$ values. However, this is equivalent to conditioning a uniform z on an event with probability at least half and therefore (by Lemma 19) can only increase the probability of agreement with e by a factor of two.

Therefore, the matrix obtained after $T_1 = O(n \log n \cdot w) = O(n \log n \cdot (\log k + \log n))$ steps is 2^{-w} -close to 2-wise independent, and by Lemma 12 it is generic with probability $1 - o(1/n^3 k^3)$. This implies that if we proceed by $T_2 = O(n^3 k^3)$ steps, then all T_2 matrices encountered are generic with probability $1 - o(T_2/n^3 k^3) > 1 - \epsilon_1$, for any fixed $\epsilon_1 > 0$ and sufficiently large n .

We introduce a new Markov chain P' . The state space of P' consists of all generic k by n matrices. If the chain is currently at the matrix M , then the next state is determined as follows. We pick a uniformly distributed simple permutation $\sigma \in \Sigma$. If $M\sigma$ is generic, we move to $M\sigma$. Otherwise, we remain at M . Let $\tau'(\epsilon)$ denote the ϵ -mixing time of P' , and require that $T_2 \geq \tau'(\epsilon_2)$ for some fixed $\epsilon_2 > 0$.

We claim that as long as $2\epsilon_1 + \epsilon_2 < 1/4$ the mixing time of P can be bounded by $\tau(1/4) \leq T_1 + T_2$. To see this, let M be some k by n matrix with distinct rows, and consider the following two matrices. The first matrix M' obtained when starting at M and walking $T_1 + T_2$ steps using P . The second matrix M'' is defined as follows. Let \hat{M} be the matrix obtained when starting at M and performing T_1 steps of P . If \hat{M} is not generic, we set $M'' = \hat{M}$. Otherwise, M'' is the matrix reached by the length T_2 walk using P' that starts at \hat{M} . We claim that $d(M', M'') \leq \epsilon_1$ and that M'' is $(\epsilon_1 + \epsilon_2)$ -close to the uniform distribution over k by n matrices with distinct rows³. Proving those claims will imply that

$$\tau(1/4) \leq \tau'(\epsilon_2) + O(n \log n \cdot (\log k + \log n)), \quad (3)$$

as long as $\tau'(\epsilon_2) = O(n^3 k^3)$.

We start by checking that indeed $d(M', M'') \leq \epsilon_1$. It is convenient to think of the two length $T_1 + T_2$ walks from M to M' and M'' as defined over the same probability space $\Sigma^{T_1+T_2}$ which is the choice of a simple permutation in each of the $T_1 + T_2$ steps. Denote the P -walk by $(M_0 = M, M_1, \dots, M_{T_1+T_2} = M')$. Then, if all the matrices $M_{T_1}, \dots, M_{T_1+T_2}$ are generic, it coincides with the walk leading to M'' , and in particular we have $M' = M''$. By the previous arguments, this event happens at least with probability $1 - \epsilon_1$, implying that $d(M', M'') \leq \epsilon_1$.

The proof that M'' is $(\epsilon_1 + \epsilon_2)$ -close to uniform is more delicate. We know that the matrix \hat{M} is generic with probability at least $1 - \epsilon_1$. Also, since $T_2 \geq \tau'(\epsilon_2)$, we know that conditioned on \hat{M} being generic, M'' is ϵ_2 -close to the uniform distribution. Therefore M'' is $(\epsilon_1 + \epsilon_2)$ -close to the uniform distribution over matrices with distinct rows. This argument can be easily formalized using Lemma 18 of Section 6.

We are left with the proof of the following lemma.

Lemma 14. $\tau'(1/4) = \tilde{O}(n^2 k^2)$.

To bound the mixing time of the Markov chain P' , we apply the comparison technique [8]. We compare P' to the Markov chain \tilde{P} defined on the same state space, the k by n generic matrices. Given that \tilde{P} is at a matrix M , we determine the next state as follows. With probability half we pick a random column $c \in C$ and row $r \in [k]$ and flip the corresponding bit with probability half. Otherwise, we pick at random an index $i \in [p]$, a row $r \in [k]$ and a string $\alpha \in \{0, 1\}^w$ that is distinct from all other $k - 1$ rows in the restriction of M to the columns C_i . We set the bits at row r and columns C_i to α .

³Note that by our assumptions, the distance between the uniform distribution over matrices with distinct rows and generic matrices is $o(1)$

Consequently, the following two lemmas, imply Lemma 14. Note that we need not worry about the smallest eigenvalue of P' since a random permutation from Σ is the identity with probability $1/16$.

Lemma 15. $\text{gap}(\tilde{P}) = \Omega(1/nk)$.

Lemma 16. *The comparison constant A of \tilde{P} to P' satisfies $A = \tilde{O}(1)$.*

Proof. (of Lemma 15)

Consider two Markov chains \tilde{P}_1 and \tilde{P}_2 :

1. The state space of \tilde{P}_1 are the k by w binary matrices with distinct rows. At each step one chooses a random row and sets it to a random new value distinct from all other $k - 1$ rows. This chain is exactly the coloring chain of a clique on k vertices with 2^w colors of [15, Proposition 4.5], and as in the proof of Lemma 11, it satisfies $\text{gap}(\tilde{P}_1) = \Omega(1/k)$.
2. \tilde{P}_2 is the random walk on the $(n - wp) \cdot k$ dimensional binary cube, where in each step with probability half, one flips a random coordinate. Therefore, $\text{gap}(\tilde{P}_2) = \Omega(1/nk)$.

One can think of the chain \tilde{P} as the product of p copies of \tilde{P}_1 and one copy of \tilde{P}_2 . Indeed the state space of \tilde{P} is the direct product of the $p + 1$ state spaces. Moreover, a step of \tilde{P} performs a move of \tilde{P}_2 with probability $1/2$ and otherwise performs the move in a randomly selected copy of \tilde{P}_1 . It is straight forward to check that the spectral gap of \tilde{P} is $\min(\text{gap}(\tilde{P}_1)/p, \text{gap}(\tilde{P}_2))/2$, implying the desired bound. □

Proof. (of Lemma 16)

Let G' be the underlying graph of P' . The vertices of G' are the generic k by n matrices, and (N, N') is an edge of G' if $P'(N, N') > 0$. To bound the comparison constant A , we need to construct a multicommodity flow f in G' that flows a unit between every two matrices M, M' such that $\tilde{P}(M, M') > 0$. The chains P' and \tilde{P} correspond to random walks on regular graphs with degrees $d' = \Theta(n^3)$, $\tilde{d} = \Theta(kn2^w/w)$ respectively, and as before the comparison constant A is defined by (2).

To build a path γ from M to M' we need to distinguish two types of \tilde{P} transitions. Type (i) flips the bit at row r and column $c \in C$. Type (ii) changes the bits at row r and columns C_i from α to α' . We start by constructing the type (i) paths.

Let $j \in [p]$ be a random index, and let $\beta \in \{0, 1\}^w$ be the restriction of the r -th row of M to C_j . Also let S be a random sequence of $w - 1$ distinct elements from $C \setminus \{c\}$. The unit flow from M to M' is along paths $\gamma = \gamma_{M, M'}^{S, j}$. Each such path is defined by composing simple permutations from Σ to achieve the permutation that acts on $x \in \{0, 1\}^n$ by flipping coordinate c if the restriction of x to C_j is β . Clearly such a permutation maps M to M' . We follow the method of Barenco et al. [3] to build an AND gate with w inputs. This gate inverts its output bit (the coordinate c) if its w inputs (the coordinates C_j) have some fixed value β . The coordinates in the set S are used as “scratch”.

Let $C_j = \{j_1, \dots, j_w\}$, $S = \{s_1, \dots, s_{w-1}\}$ and $\beta = (b_1, \dots, b_w)$. Let σ_1 be the simple permutation that flips coordinate s_1 of $x \in \{0, 1\}^n$ if x_{j_1} is equal to b_1 , and let σ_ℓ for $2 \leq \ell \leq w - 1$ be the simple permutation that flips coordinate s_ℓ if $x_{s_{\ell-1}}$ is one and x_{j_ℓ} is equal to b_ℓ . Also, we denote

by τ_c the simple permutation that flips x_c if $x_{s_{w-1}}$ is one and x_{j_w} is equal to b_w . We claim that the following permutation flips coordinate c of $x \in \{0, 1\}^n$ if the restriction of x to C_j is equal to β :

$$\sigma = (\tau_c \sigma_{w-1} \cdots \sigma_2 \sigma_1 \sigma_2 \cdots \sigma_{w-1})^2$$

To see this, one checks by induction that $\sigma_\ell \cdots \sigma_1 \cdots \sigma_\ell$ flips coordinate s_ℓ if $x_{j_1}, \dots, x_{j_\ell}$ is equal to b_1, \dots, b_ℓ .

For the type (ii) paths, we need to change the bits at row r and columns C_i from α to α' . The problem is that if we change α to α' bit by bit, as suggested by the construction of type (i) paths, we might violate row distinctness. To solve this problem, we start our path by applying a length $L = O(w \log w \cdot (1 + 2 \log k))$ sequence ϕ of simple permutations with indices restricted to C_i . Let $\hat{M} = M\phi$ and $\hat{M}' = M'\phi$, and let C'_i and C''_i be the first and last $\lfloor (w-1)/2 \rfloor$ columns of C_i . We say that ϕ is valid if for both the restriction of \hat{M} to C''_i and for the restriction of \hat{M}' to C'_i , have distinct rows. By Lemma 13 we know for a random ϕ , both \hat{M} and \hat{M}' are $1/8k^2$ -close to 2-wise independence. Therefore, a random ϕ is not valid with probability bounded by $k^2 \cdot (2^{-w/2+1} + 1/8k^2) \leq 1/4$. If ϕ is valid we define a path $\gamma = \gamma_{M, M'}^{S, j, \phi}$ from M to M' , where $j \in [p] \setminus \{i\}$ and S is a length $w-1$ sequence of elements from C . The path is prefixed by ϕ to get from M to \hat{M} and is suffixed by ϕ^{-1} to get from \hat{M}' to M' . Let $\hat{\alpha}$ and $\hat{\alpha}'$ be the restriction of the r -th row of \hat{M} and \hat{M}' to C_i respectively, and let β be the restriction of the r -th row of M to C_j . Then the middle path connecting \hat{M} to \hat{M}' is defined as follows:

$$\sigma = \left[\left(\prod_{\{c \in C'_i : \hat{\alpha}_c \neq \hat{\alpha}'_c\}} \tau_c \right) \cdot \sigma_{w-1} \cdots \sigma_2 \sigma_1 \sigma_2 \cdots \sigma_{w-1} \right]^2 \cdot \left[\left(\prod_{\{c \in C_i \setminus C'_i : \hat{\alpha}_c \neq \hat{\alpha}'_c\}} \tau_c \right) \cdot \sigma_{w-1} \cdots \sigma_2 \sigma_1 \sigma_2 \cdots \sigma_{w-1} \right]^2,$$

where τ_c and σ_ℓ are as defined for the type (i) sequences. Therefore it is guaranteed that the matrices encountered along the first and second half of the sequence agree with \hat{M} on the columns C''_i and with \hat{M}' on the columns C'_i respectively. Since ϕ is valid, this implies that we never attempt to move to a non-generic matrix throughout the entire path. We define the unit flow from M to M' by splitting the flow uniformly between all valid paths γ designated by S, j, ϕ .

There are two points that need special attention in the constructed type (i) and type (ii) paths. The first point is that all indices of the simple permutations used in ϕ are in C_i . This is unacceptable for us, as it induces an undue load on a small subset of Σ . To solve this problem we replace each simple permutation used in ϕ by a constant length sequence that avoids that problem. For example, the permutation that flips coordinate i_1 if i_2 and i_3 are 1, denoted χ_{i_1, i_2, i_3} , is replaced by the sequence $(\chi_{s_2, s_1, i_3} \chi_{s_1, i_2}, \chi_{s_2, s_1, i_3}, \chi_{i_1, s_2})^2$ where permutation χ_{i_1, i_2} XORs coordinate i_1 with i_2 .

The second point is that some of the simple permutations used (σ_1 and some of the permutations in ϕ) do not use three indices. However, in the definition of Σ , we have three indices at our disposal even if we don't use all three. We use this to guarantee that all simple permutations used have one index in C_j and two from S or c for type (i) paths or C_i for type (ii) paths.

To complete the proof, we have to bound the comparison constant A given by (2). We have $d'/\tilde{d} = \theta(n^2 w/k2^w)$ and $|\gamma| = O(L)$. Also, $f(\gamma)$ is $\Theta(w/n(m)_{w-1})$ for type (i) paths and $\Theta(w/|\Sigma^{(w)}|^L n(m)_{w-1})$ for type (ii) paths, where we denote $m = |C|$, $(m)_q = m(m-1)(m-2) \cdots (m-q+1)$, and $\Sigma^{(w)}$ as the width 2 simple permutations restricted to the w -dimensional cube. Therefore, we only have to bound the maximal number of $\gamma_{M, M'}^{S, j}$ and $\gamma_{M, M'}^{S, j, \phi}$ paths through an edge (N, N') .

We start with type (i) paths. The first step is to extract as much information as possible about a path γ through (N, N') by considering the simple permutation s associated with (N, N') . Note

first that s determines j . Moreover, since only one of $\sigma_1, \dots, \sigma_{w-1}$ and τ_c can be equal to s , any path γ using s , must use it in one of $O(1)$ possible positions. Since a permutation σ_ℓ for $\ell \in [w-1]$ or τ_c determines two indices of S, c there are only $\Theta((m)_{w-2})$ choices for S, c that are consistent with s . The last thing still needed to reconstruct γ is the string $\beta \in \{0, 1\}^w$. Since the columns C_j are not modified throughout the entire sequence, β must be the restriction of some row of N to C_j , limiting β to one of k possible values. Therefore, the total number of type (i) paths through (N, N') is $O(k \cdot (m)_{w-2})$, and the contribution of the type (i) sequences to A is:

$$A_{(i)} = O\left(\overbrace{(n^2 w / k 2^w)}^{d'/\tilde{d}} \cdot \overbrace{(Lw / (m)_w)}^{f(\gamma) \cdot |\gamma|} \cdot \overbrace{(k \cdot (m)_{w-2})}^{\text{choices for } j, S, c, \beta \text{ and position}}\right) = O(Lw^2 / 2^w) = o(1).$$

For type (ii) paths we distinguish the cases where (N, N') is in the first middle or last sections of a path $\gamma_{M, M'}^{S, j, \phi}$. Consider the first section (and similarly the last). We enumerate over possible positions $\ell \in [L]$. Then we know two indices of the sequence S and one of the $3L$ indices in C_i that were used by ϕ . Therefore, we have $L \cdot (m)_{w-3} \cdot |\Sigma^{(w)}|^L / w$ possible values for S, i, ϕ and the position. This enables us to determine M and \tilde{M} . We still have to determine the row r , the two strings α, α' and the index j which have $O(kn2^w/w)$ possibilities. Therefore the contribution of the first and last sections of type (ii) paths is:

$$\begin{aligned} A_{(ii.\text{first, last})} &= O\left(\overbrace{(n^2 w / k 2^w)}^{d'/\tilde{d}} \cdot \overbrace{(Lw / |\Sigma^{(w)}|^L (m)_w)}^{f(\gamma) \cdot |\gamma|} \cdot \overbrace{(k 2^w L \cdot (m)_{w-2} \cdot |\Sigma^{(w)}|^L / w^2)}^{\text{choices for } j, S, i, \phi, \alpha, \alpha', \beta \text{ and position}}\right) \\ &= O(L^2) = O(w^2 \log^2 w \cdot (1 + \log k)^2). \end{aligned}$$

For the middle section of type (ii) paths, as for the type (i) argument, given (N, N') we first determine the position up to $O(1)$ possible choices. Then we determine the index i or j and two indices from S , then we have $O((m)_{w-2} \cdot |\Sigma^{(w)}|^L / w)$ possibilities for i, j, S, ϕ . Also we have $k2^w$ choices for the row and the strings β, α and α' . Therefore,

$$A_{(ii.\text{middle})} = O\left(\overbrace{(n^2 w / k 2^w)}^{d'/\tilde{d}} \cdot \overbrace{(Lw / |\Sigma^{(w)}|^L (m)_w)}^{f(\gamma) \cdot |\gamma|} \cdot \overbrace{(k 2^w \cdot (m)_{w-2} \cdot |\Sigma^{(w)}|^L / w)}^{\text{choices for } j, S, i, \phi, \alpha, \alpha', \beta \text{ and position}}\right) = O(Lw).$$

□

5 Proof of Theorem 9

First, we describe the randomized implementation of a 3-cycle (xyz) using the simple permutations in Σ . Second, we show that this randomized implementation satisfies the statement of the theorem. The randomness is introduced into the implementation of (xyz) by using a permutation $\phi \in S_n$ and two vectors v_4, v_5 .

Let ϕ be some permutation of the n coordinates. If $\omega = \sigma_1 \cdots \sigma_L$ implements $(x\phi, y\phi, z\phi)$, then ω^ϕ is an implementation of (xyz) , where $\omega^\phi = \phi\omega\phi^{-1}$ is the conjugation of ω with ϕ , i.e. the conjugation each of the permutations σ_i used in ω . Note that the set Σ of simple permutations is closed under conjugation by permutations from S_n , because this just relabels the indices.

For a vector $v \in \{0, 1\}^n$, we denote the first $n-2$ bits of v by $v' \in \{0, 1\}^{n-2}$ and the last two bits of v by $v'' \in \{0, 1\}^2$, i.e., $v = v'v''$. We call the last two bits the *control bits*. For convenience, the

notation $v'00$, $v'01$, $v'10$, and $v'11$ denotes bit vectors comprising the first $n - 2$ bits of v and the control bits 00, 01, 10, and 11, respectively. Let $(v)_j$ denote the j -th bit of a vector v . Finally, let $v_1 = x\phi$, $v_2 = y\phi$ and $v_3 = z\phi$.

If v'_1 is equal to v'_2 or to v'_3 then we say that ϕ is invalid. This can only occur if x , y , or z are less than Hamming distance 3 apart and ϕ maps all indices on which x and y (or z) differ to the control indices. For the rest of the description we assume that ϕ is valid. Let $v_4, v_5 \in \{0, 1\}^n$ be two additional vectors satisfying the validity requirement of being at least Hamming distance 3 from each other and from the former three vectors.

Observe that $(v_1, v_2, v_3) = \psi_1\psi_2$ where $\psi_1 = (v_1, v_2)(v_4, v_5)$ and $\psi_2 = (v_1, v_3)(v_4, v_5)$. Therefore it suffices to implement the two double transpositions ψ_1 and ψ_2 . These are implemented in an identical manner. Each implementation is divided into 15 blocks: a core block, which implements the permutation $\rho_{core} = (v'_500, v'_501)(v'_510, v'_511)$, and seven block pairs conjugating it.

The first four of these blocks, called π -blocks ensure that the control bits of each of the four vectors are distinct. Specifically, $v'_i v''_i$ is mapped to $v'_i c_i$, where $c_1 = 00$, $c_2 = c_3 = 01$, $c_4 = 10$ and $c_5 = 11$. If $v''_i = c_i$ then the corresponding block, labeled π_i performs a nop. Otherwise, block π_i performs the permutation $(v'_i v''_i, v'_i c_i)(v'_i a_i, v'_i b_i)$ where $\{a_i, b_i\} = \{0, 1\}^2 \setminus \{v''_i, c_i\}$.

The remaining three blocks, called τ -blocks, map v'_1, v'_2 (or v'_3), and v'_4 to v'_5 , using the control bits to distinguish between the four vectors. Block τ_i performs the permutation $\tau_i = \prod_{v' \in \{0, 1\}^{n-2}} (v' c_i, u' c_i)$, where $u' = v' \oplus v'_i \oplus v'_5$. Since it can easily be checked that $\tau_i = \tau_i^{-1}$, that $\pi_i = \pi_i^{-1}$, and that

$$\pi_1 \pi_2 \pi_4 \pi_5 \tau_1 \tau_2 \tau_4 \rho_{core} \tau_4 \tau_2 \tau_1 \pi_5 \pi_4 \pi_2 \pi_1 = \psi_1 \quad \text{and} \quad \pi_1 \pi_3 \pi_4 \pi_5 \tau_1 \tau_3 \tau_4 \rho_{core} \tau_4 \tau_3 \tau_1 \pi_5 \pi_4 \pi_3 \pi_1 = \psi_2,$$

we need only describe the implementation of each of these blocks.

Each of the blocks is implemented using $O(n)$ simple permutations. Each τ -block is implemented by concatenating $n - 2$ simple permutations, where for $j = 1 \cdots n - 2$, the j -th simple permutation is the identity if $(v'_i)_j = (v'_5)_j$, and otherwise flips the j -th bit of vector v if $v'' = c_i$.

The implementation of the ρ_{core} and π blocks is more involved. Permutation ρ_{core} flips bit $(v'')_2$ if and only if $v' = v'_5$. Barenco et al [3] showed how such permutations can be implemented using $O(n)$ simple permutations, comprising four sub-blocks: $\rho_{top}\rho_{bot}\rho_{top}\rho_{bot}$ where permutation ρ_{top} flips bit $(v'')_1$ if the first $\lceil (n - 2)/2 \rceil$ bits of v' match the first $\lceil (n - 2)/2 \rceil$ bits of v'_5 , and where permutation ρ_{bot} flips bit $(v'')_2$ if the latter $\lfloor (n - 2)/2 \rfloor$ bits of v' match the latter $\lfloor (n - 2)/2 \rfloor$ bits of v'_5 and $(v'')_1 = 1$. Each sub-block uses the remaining $\lceil (n - 2)/2 \rceil$ bits as “scratch”, returning them to their original state by the end of the sub-block. For details about the construction of the two sub-blocks see [3] or Lemma 16.

Each block π_i is implemented in a similar manner using two permutations that are nearly identical to the implementation of ρ_{core} . The first (second) permutation performs the identity if $(v'')_1 = (c_i)_1$ (respectively, $(v'')_2 = (c_i)_2$) and otherwise flips bit $(v'')_1$ (respectively, $(v'')_2$) if $v' = v'_i$.

The length of the implementations of ψ_1 and ψ_2 is $O(n)$, since each of the seven blocks can be implemented using $O(n)$ simple permutations from Σ . The randomized implementation of (xyz) is obtained by uniformly choosing at random a valid permutation ϕ and the two valid random vectors v_4, v_5 .

We now prove that this randomized implementation satisfies the statement of the theorem. Let $\Omega = \{x, y, z, v_4, v_5, \phi\}$ be the probability space obtained by uniformly choosing three distinct vectors x , y , and z , and then uniformly choosing a corresponding implementation, which is fixed by v_4 ,

v_5 , and ϕ . Each point $\omega = (x, y, z, v_4, v_5, \phi) \in \Omega$ corresponds to an implementation $\sigma_1 \cdots \sigma_L$ of the 3-cycle (xyz) . The size of Ω is $\Theta(2^{5n}n!)$, and although not uniform, the probability of each point in Ω is $O(1/2^{5n}n!)$. Thus, our problem of upper-bounding $\Pr[x\sigma_1\sigma_2 \cdots \sigma_{\ell-1} = \tilde{x}, \sigma_\ell = \tilde{\sigma}]$ reduces to a counting problem.

For all implementations $\omega \in \Omega$, the indices of the ℓ -th permutation σ_ℓ depend *only* on its position, ℓ , and ϕ . Moreover, as we change ϕ the indices of the ℓ -th permutation of the implementation of $(x, y, z, v_4, v_5, \phi)$ agree with the indices of some fixed permutation $\tilde{\sigma}$ only on a subset of S_n that is of size $O(n!/n^3)$ and depends only on ℓ and $\tilde{\sigma}$.

To establish the theorem we need to prove that for any given ϕ the number of choices of x, y, z, v_4 , and v_5 , such that $x\sigma_1\sigma_2 \cdots \sigma_{\ell-1} = \tilde{x}$, is $O(2^{4n})$, implying the number of points in Ω that agree with \tilde{x} and $\tilde{\sigma}$ is $O(2^{4n}n!/n^3) = O(|\Omega|/2^n n^3)$. This is accomplished by the following lemma:

Lemma 17. *Let $\phi \in S_n$ be fixed. Then the set of all x, y, z, v_4, v_5 such that implementation corresponding to $(x, y, z, v_4, v_5, \phi)$ satisfies the equality $x\sigma_1\sigma_2 \cdots \sigma_{\ell-1} = \tilde{x}$ is of size $O(2^{4n})$.*

Proof. (of Lemma 17)

Let $v_1 = x\phi$, $v_2 = y\phi$, $v_3 = z\phi$, and $\tilde{v} = \tilde{x}\phi$. Let $\Omega_{\tilde{v}, \ell}$ be the set of tuples (v_1, \dots, v_5) for which $x\sigma_1\sigma_2 \cdots \sigma_{\ell-1} = \tilde{x}$ is satisfied. Note that this set is independent of ϕ . Then the claim is that $|\Omega_{\tilde{v}, \ell}| = O(2^{4n})$.

The proof is via case analysis with respect to position ℓ . Without loss of generality we assume that the position is in the first half of the implementation, that which realizes permutation $(v_1, v_2)(v_4, v_5)$, otherwise, swapping v_2 and v_3 allows the same argument to be reused for the latter half of the implementation. Furthermore, due to symmetry, we assume that the position of ℓ is in or to the left of block ρ_{core} . There are four main cases: either ℓ is on a boundary between two blocks, ℓ is in block τ_i , ℓ is in the block ρ_{core} , or ℓ is in block π_i .

$$\underbrace{v_1 \xrightarrow{\pi_1} v'_1 c_1 \xrightarrow{\pi_2 \pi_4 \pi_5} v'_1 c_1}_{\tilde{v}' \text{ fixes } v'_1} \xrightarrow{\tau_1} \underbrace{v'_5 c_1 \xrightarrow{\tau_2 \tau_4} v'_5 c_1 \xrightarrow{\rho_{core}} v'_5 c_2 \cdots}_{\tilde{v}' \text{ fixes } v'_5}$$

Figure 1: The evolution of v_1 .

In the first case, the position, ℓ , is on a block boundary. Since each π -block only toggles bits $(v'')_1$ and $(v'')_2$, if position ℓ is adjacent to a π -block, then $\tilde{v}' = v'_1$. Thus, all but two bits of v_1 are fixed by \tilde{v} . If position, ℓ , is on a boundary but is not adjacent to a π -block, then it must occur after block τ_1 . Since block τ_1 maps $v'_1 00$ to $v'_5 00$, and none of the remaining blocks, τ_i or ρ_{core} , change the v' component to any other value, we have $\tilde{v}' = v'_5$. Thus, all but two bits of v_5 are fixed by \tilde{v} , implying that $|\Omega_{\tilde{v}, \ell}| = O(2^{4n})$.

In the second case, the position, ℓ , is inside block τ_i . If $i \neq 1$, then none of the simple permutations in block τ_i flips a bit. Therefore, the value of $\tilde{v}' = v'_5$; thus fixing all but two bits of v_5 , as before. If $i = 1$, then at position ℓ , we know exactly how many of the $n - 2$ simple permutations have already been performed. Let j be this number. Hence we know that $\tilde{v}' = (v'_5)_1, \dots, (v'_5)_j, (v'_1)_{j+1}, \dots, (v'_1)_{n-2}$. Therefore, j bits of v_5 and $n - 2 - j$ bits of v_1 are therefore fixed by \tilde{v} , implying that $|\Omega_{\tilde{v}, \ell}| = O(2^{4n})$ as well.

In the third case, the position, ℓ , is inside block ρ_{core} . In this case we must look at the sub-blocks of the block ρ_{core} . If the position occurs on a sub-block boundary, and since each of the sub-blocks

simply toggles the bits $(v'')_1$ and $(v'')_2$, the remaining bits of v'_5 are fixed by \tilde{v}' . If the position ℓ is inside a sub-block, then things are only slightly more complicated. Assume that position ℓ is in a ρ_{top} sub-block (similar arguments hold for ρ_{bot}). Then, ρ_{top} toggles bit $(v'')_1$ if the first half of v' matches the first half of v'_5 . The bits being matched are never modified and the other half of the bits of v' are used as “scratch”. We know that the first half of v' and v'_5 coincide throughout the block ρ_{top} , and therefore \tilde{v}' determines this half of v'_5 . The operations on the “scratch” half depends only on the fixed half and the position, and therefore can be reversed, reducing the problem to the position occurring at the beginning of ρ_{top} . Thus, \tilde{v} fixes all but two of the bits of v_5 .

In the last case, the position, ℓ , is inside a π -block. Block π_i comprises two blocks that are similar to ρ_{core} . Each of the two blocks is either the identity or toggles $(v'')_1$ or $(v'')_2$ if $v' = v'_i$. If $i = 1$ then the two blocks in Block π_i behave in the same manner as block ρ_{core} , except that \tilde{v} fixes all but two of the bits of v_1 rather than v_5 . If $i \neq 1$, then, for the most part, the argument remains the same. We need only consider what happens if the position, ℓ , is in one of the eight sub-blocks. As mentioned before, half of the bits of v' are not modified by the sub-block, while the other half are used as “scratch”. Again, without loss of generality, we assume that the sub-block does not modify the first half of v' . As before, \tilde{v}' fixes the first half of v'_1 . We enumerate on all choices for the first half of v'_i . This enables us to reverse the operations of the sub-block on the “scratch”, fixing the second half of v'_1 —as in the third case. This implies that $|\Omega_{\tilde{v}, \ell}| = O(2^{4n})$, and completes the proof. \square

6 Odds and Ends

Proof. (of Lemma 13)

We have to prove that for all $w \geq 1$ the mixing time of $G_{2,n}^{(w)} = \text{sc}(\Sigma_w, X^{(2)})$ is $O(n \log n)$.

Given a 2 by n matrix with rows s, t , we change basis to s, u with $u = s \oplus t$. Let $i \in [n]$ be a random coordinate, and consider the action of a width w permutations XORing the i -th bit with a random function h on w distinct coordinates from $[n] \setminus \{i\}$. We claim that its action on s, u is the same as XORing the i -th bit of s and u with two *independent* random bits α_s and α_u respectively. The bits α_s, α_u are one with probability $1/2$ and $p_\ell = 1 - \prod_{j=1}^w (1 - \frac{\ell}{n-j})$ respectively, where ℓ is the number of ones in u not counting the i -th bit. To see that this is indeed the resulting walk we observe the fact that if s and t differ on one of the input bits of the random function h , then the value of the i -th coordinate of s and of t change independently with probability half. Otherwise they change simultaneously with probability $1/2$.

The u -component of this walk is a variant of the Aldous cube, and by the comment at the end of [5] 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 A where the indices i assume all possible values in $1, 2, \dots, n$ (coupon collector) happens with high probability. Now since the bits α_s are independent of α_u , we get that even when we condition over the walk on the u component, the s component achieves uniform distribution conditioned on A , which ends the proof. \square

Lemma 18. *Let A be an event such that $\Pr[A] \geq 1 - \epsilon$, and let Z be a random variable over a domain Ω such that $d(Z|A, \text{uniform}) \leq \epsilon$. Then $d(Z, \text{uniform}) \leq 2\epsilon$.*

Proof.

$$d(Z, \text{uniform}) = \max_{S \subseteq \Omega} \Pr[Z \in S] - \frac{|S|}{|\Omega|} \leq \max_{S \subseteq \Omega} \Pr[Z \in S|A] + \Pr[\bar{A}] - \frac{|S|}{|\Omega|} \leq \epsilon + d(Z|A, \text{uniform}) \leq 2\epsilon.$$

□

Lemma 19. *Let X be a random variable and A an event. Then $\Pr[X|A] \leq \Pr[X]/\Pr[A]$. (Follows from the definition of conditional probability.)*

7 Some concluding remarks

Let us review what we currently know about the spectral gap of the Markov chain $P = P_{\Sigma}^{(k,n)}$. By Theorem 2, $\text{gap}(P) \leq \Omega(1/n^2k)$. On the other hand, $\text{gap}(P)$ is nonincreasing in k by the lifting argument from Section 3. Since for $k = 1$, P is the standard random walk on the cube, we have that $\text{gap}(P) \geq 1/n$.

In general, a generating set S for which the spectral gap is large becomes more difficult as k increases, until the largest conceivable k , which is $2^n - 2$. In this case, this is the random walk on the Cayley graph of the alternating group A_N for $N = 2^n$ with the generating set S . It is open whether one can find a constant size set for which A_N is an expander, [17, Problem 10.3.4].⁴ On the other hand, by Alon and Roichman [2], a random set of permutations of size $O(N \cdot \log N)$ will almost surely have a constant spectral gap. Although smaller expanding sets for A_N are not known to exist, the general belief is that such sets exist; Rozenman, Shalev, and Wigderson assume the existence of an $N^{1/30}$ expanding set for A_N , [20, section 1.4].

Our results suggest that width 2 permutations may be used to construct an $O(\log^3 N)$ expanding set for A_N . However, several obstacles stand in the way of achieving this goal. The first one is to prove that for width 2 permutations the spectral gap does not deteriorate with k , as we believe, and is $\Omega(1/n)$ for all k . The second problem is to achieve a constant gap. To this end, one has to overcome the inherent and obvious weakness of the width 2 simple permutations. Namely, that their action depends only on two coordinates and changes only one. This leads to poor expansion because there is only a small chance that the action will flip a specific bit or increase the distance between two similar vectors. One approach to avoiding this problem is to replace the standard set of generators of the cube e_1, \dots, e_n with some expanding set of size $O(n)$. Such an expanding set for the cube can readily be constructed from the generating matrix of a good code [7], and could then be used to define an $O(n^3)$ expanding set of permutations.

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] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994.

⁴The problem of finding a constant size expanding set for A_N or S_N is equivalent.

- [3] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter. Elementary gates for quantum computation. *Phys. Rev. A*, 52(5):3457–3467, 1995.
- [4] A. Brodsky. Reversible circuit realizations of boolean functions. In *proceedings of the 3rd IFIP International Conference on Theoretical Computer Science*, 2004.
- [5] F. R. K. Chung and R. L. Graham. Stratified random walks on the n -cube. *Random Structures Algorithms*, 11(3):199–222, 1997.
- [6] D. Coppersmith and E. Grossman. Generators for certain alternating groups with applications to cryptography. *SIAM J. Appl. Math.*, 29(4):624–627, 1975.
- [7] C. Delorme and P. Solé. Diameter, covering index, covering radius and eigenvalues. *European J. Combin.*, 12(2):95–108, 1991.
- [8] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.
- [9] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.
- [10] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.
- [11] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.
- [12] J. Friedman. On Cayley graphs on the symmetric group generated by transpositions. *Combinatorica*, 20(4):505–519, 2000.
- [13] W. T. Gowers. An almost m -wise independent random permutation of the cube. *Combin. Probab. Comput.*, 5(2):119–130, 1996.
- [14] S. Hoory, A. Magen, S. Myers, and C. Rackoff. Simple permutations mix well. In *31st International Colloquium on Automata, Languages and Programming (ICALP)*, 2004.
- [15] M. Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.
- [16] M. Kassabov. Symmetric groups and expander graphs. Arxiv math.GR/0505624.
- [17] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures*, volume 125 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 1994.
- [18] U. Maurer and K. Pietrzak. Composition of random systems: When two weak make one strong. In *The First Theory of Cryptography Conference*, 2004.
- [19] R. Montenegro and P. Tetali. *Mathematical Aspects of Mixing Times in Markov Chains*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2006.
- [20] E. Rozenman, A. Shalev, and A. Wigderson. A new family of Cayley expanders. *ACM Symposium on Theory of Computing (STOC)*, 2004.

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