When Simple Permutations Mix Poorly: Limited Independence Does Not Imply Pseudorandomness

Jesko Dujmovic* Angelos Pelecanos† Stefano Tessaro‡

February 24, 2026

Abstract

Over the past two decades, several works have used (almost) k-wise independence as a proxy for pseudorandomness in block ciphers, since it guarantees resistance against broad classes of statistical attacks. For example, even the case $k = 2$ already implies security against differential and linear cryptanalysis.

Hoory, Magen, Myers, and Rackoff (ICALP '04; TCS '05) formulated an appealing conjecture: if the sequential composition of T independent *local* randomized permutations is (close to) four-wise independent, then it should also be a pseudorandom permutation. Here, "local" means that each output bit depends on only a constant number of input bits. This conjecture offers a potential strong justification for analyses of block ciphers that establish (almost) k-wise independence of this type of constructions.

In this work, we disprove the conjecture in full generality by presenting an explicit local randomized permutation whose sequential composition is four-wise independent, but not a pseudorandom permutation. Our counterexample in fact extends to k-wise independence for any constant k.

*Northeastern University. Email: mail@ind-jesko.net. This paper describes work while at CISPA Helmholtz Center for Information Security and Saarland University.

†UC Berkeley. Email: apelecan@berkeley.edu

‡University of Washington. Email: tessaro@cs.washington.edu

Contents

1 Introduction **3**

2 Technical overview **5**
2.1 Round-dependent permutations 5
2.2 Round-independent permutations 7

3 Preliminaries **10**
3.1 Linear algebra and probability 10
3.2 Markov chains 11
3.3 Comparison using multicommodity flows 12

4 Block cipher construction **14**
4.1 Mixing step 15
4.2 Conditional brickwork step 16
4.3 Increase counter step 17
4.4 Randomize special strings step 17
4.5 Remarks 18

5 The distinguishing attack **19**

6 Reaching k-wise independence **21**
6.1 The spectral gap of the conditional brickwork step 22
6.2 Proof of k-wise independence 27

7 Non-special strings reach high min-entropy quickly **30**
7.1 Distance growth after each mixing step 31
7.2 Min-entropy growth after mixing 35

8 Discussion **38**

A Code **43**

1 Introduction

One of the most fundamental questions in cryptography is how to design secure block ciphers. An inherent tension stems from the fact that theoretical cryptography, starting with the seminal work of Goldreich, Goldwasser and Micali [GGM84], has succeeded in constructing pseudorandom functions (and permutations [LR86]) for which security reduces to the hardness of well-studied computational problems (see [NR97, NRR02, BPR12, BP14] for example) and even, generically, to the existence of one-way functions [HILL99]. These give theoretically sound candidates for block ciphers, but the resulting constructions are usually not efficient enough to be used in practice. Instead, practical designs, such as AES [AES01], are designed without resorting to the framework of provable security. Rather, the goal here is resisting a broad class of cryptanalytic attacks developed over the years, which includes linear [MY92] and differential [BS91] cryptanalysis, algebraic attacks [CP02], and many others (e.g., [Knu94, Knu98, JK97, KW02, BKR11], only to mention a few). It is surprising that, to date, many designs have evaded attacks—for example, the best known attack [BKR11] against 128-bit AES only achieves a factor four speedup over brute-force key search.

The lack of provable security for a cipher like AES is widely believed to be somewhat inherent. Indeed, these ciphers adopt a round-based structure, with each round meant to add additional security, but otherwise, on its own, each round is weak. For instance, the AES round adds first a key-dependent linear shift to the input, then applies to each byte, in parallel, a public permutation, the so-called "S-Box", and then mixes the output with a publicly known linear layer. The only possible proofs are those that idealize a block cipher component. Idealizing a full round (such as in [BKL$^+$12, Ste12, LS14, CS14, HT16]) gives weak guarantees of resistance against very limited generic attacks, as actual rounds are not ideal. Idealizing small components such as S-Boxes [DSSL16, DKS$^+$17, CDK$^+$18, CL18] is only meaningful if this component has sufficiently large input, and the extent to which such analyses apply to actual block ciphers can at best be conjectured [DKW22].

Limited independence. Therefore, the next best alternative approach is to prove security against restricted classes of attacks. To this end, one observation is that many (but not all) of the considered attacks (such as linear and differential cryptanalysis) ultimately aim to find statistical deviations in a small number of block cipher outputs (under a single random key). Several works [Gow96, HMMR05, BH08] focused on the notion of *k-wise independence* as a potential target for a partial security proof that validates the security of a block cipher against large classes of statistical attacks. Here, one demands that the outputs on any k distinct inputs, under the same secret key, are sufficiently close to random. One can, for example, easily show that k-wise independence for $k = 2$ (i.e., pairwise independence) is already sufficient to ensure security against linear and differential cryptanalysis (cf. e.g. [LTV21]). For small k's, and assuming that rounds behave independently (which is a common heuristic in the cryptanalytic literature), one can hope to prove that a concrete block cipher is indeed k-wise independent after a sufficiently large number of rounds, or more precisely, to prove it negligibly far from k-wise independent (by which we mean that the statistical distance of its outputs is negligibly far from the uniform distribution over k-tuples of distinct values). Such a proof would generally be statistical, without resorting to computational assumptions. This has led to a number of security proofs for k-wise independence, both for ciphers that implement random local circuits [Gow96, HMMR05, BH08, CHH$^+$24, GHP25, GHKO25] as well as for substitution-permutation networks (SPNs) like AES [LTV21, LPTV23, Din25, BLS25]. Vaudenay [Vau03] also introduced "decorrelation theory" which, in a more modern light, can be seen as proving forms of k-wise independence.

Does limited independence imply pseudorandomness? As much as ensuring resistance to differential and linear cryptanalysis is not sufficient to ensure a block cipher is a pseudorandom permutation, it is equally unreasonable to expect that a k-wise independent block cipher, for some sufficiently small k, also happens to be a pseudorandom permutation. For example, if we ignore the permutation constraint, for simplicity, it is well known that a random degree $k-1$ polynomial is k-wise independent, but seeing $k+1$ input-output pairs allows us to *efficiently* distinguish it from a random function.

However, it appears more reasonable to conjecture that k-wise independence (for a sufficiently large, but

not too large k), in conjunction with some extra properties (typically related to the resistance of algebraic attacks), is a better heuristic to validate the pseudorandomness of a block cipher. For example, the simple argument we gave above with a random polynomial leverages our ability to efficiently interpolate, but such an efficient attack may not exist if we can enforce the right type of structure. This viewpoint is somewhat implicit in works proving k-wise independence of block ciphers.

Hoory, Magen, Myers, and Rackoff (HMMR) [HMMR04] went one step further and made an explicit conjecture.[1] They consider the composition of *simple permutations*, i.e., round-based block ciphers obtained by composing a sufficiently simple (i.e., local) round transformation, where each round uses independent randomness. Concretely, let $\mathcal{P}$ be a randomized permutation over n-bit strings which we assume to be *local*, i.e., each output bit only depends on a small (constant) number of input bits. We consider then the T-fold sequential composition of $\mathcal{P}$, i.e., the randomized permutation $\mathcal{P}^T$ over n-bit permutations which implements

$$\mathcal{P}^T : x \mapsto (\mathcal{P}_T \circ \cdots \circ \mathcal{P}_1)(x) \ ,$$

where each $\mathcal{P}_i$ is an independent copy of $\mathcal{P}$. Then, HMMR [HMMR04] made the following explicit conjecture:

> **Conjecture 1** (HMMR's Conjecture)**.** *If $\mathcal{P}^T$ is negligibly far from 4-wise independent, then it is also a pseudorandom permutation.*

This conjecture may appear surprisingly strong, but it is yet to be refuted, and it appears challenging to do so by using existing cryptanalytic techniques. This is because four-wise independence already suffices to exclude statistical attacks such as differential cryptanalysis. Furthermore, the locality constraint on $\mathcal{P}$ forces us to look at variants of random local circuits (as in [Gow96, HMMR05, BH08, CHH$^+$24, GHP25, GHKO25]) and the typical choice of T to achieve four-wise independence is enough, in all known constructions, to defeat obvious algebraic attacks. The motivation behind four (as opposed to other constants) is rather arbitrary, but HMMR point out that counterexamples exist if we were to use 3-wise independence, so in some sense four is the smallest choice that is meaningful. The fact that a statement as simple as Conjecture 1 cannot be refuted is a clear indication of our limited understanding of such analyses.

We note that if one does not restrict their attention to local round transformations, then even Feistel constructions instantiated with perfect k-wise independent round functions (i.e., random degree $k-1$ polynomials) contradict the conjecture. In particular, it is known (e.g., Theorem 1 of [Mau92]) that three such Feistel rounds suffice to reach negligibly far from k-wise independence, for any constant k. However, the resulting permutation is distinguishable from random using a simple attack with $k+1$ inputs. For any fixed right-half value r, the left output of a three-round Feistel on inputs (ℓ, r) is an explicit degree $k-1$ polynomial in ℓ. An adversary issuing $k+1$ chosen plaintexts (ℓ_i, r) can then test whether the corresponding left halves lie on a polynomial of degree at most $k-1$, which they always do under the Feistel construction, but with only negligible probability for a uniformly random permutation. This gives an efficient $(k+1)$-query distinguisher.

Our results. Our main result is a refutation of Conjecture 1. We give an explicit construction of a local randomized permutation $\mathcal{P}$ and a value T, such that (1) its T-fold sequential composition $\mathcal{P}^T$ is negligibly far from 4-wise independent, but (2) it is not a pseudorandom permutation, and in fact, not even for adversaries evaluating $\mathcal{P}^T$ on five inputs. (We stress that we are not happy here with just showing that $\mathcal{P}^T$ is not five-wise independent, as we want to give an *efficient* distinguisher.)

We give more details in the overview below. Here, we simply specify that our permutations will be expressed as mapping $\Sigma^n \to \Sigma^n$, for a constant-size alphabet Σ, and that $T = \Theta(n)$ for us. Any ℓ-local permutation over Σ-wires can be translated into a $\ell \log |\Sigma|$-local permutation over bits, which preserves constant locality for a constant-sized alphabet Σ.

[1] This conjecture is often attributed originally to Rackoff, but we dispense with this naming, as there is no published reference giving this concrete attribution.

2 Technical overview

In this section, we lay out the ideas behind the design and analysis of the permutation that refutes Conjecture 1. The core idea is to use families of permutations with constant round locality for which we have proofs that they are 4-wise independent and plant trapdoors that allow an attacker to efficiently break pseudorandomness with just 5 inputs. For simplicity, we focus on 4-wise independence during the technical overview, and we generalize our construction to k-wise independence for any constant k in the main part of this paper.

2.1 Round-dependent permutations

First, we elaborate on the core principles of our counterexample with a simplified conjecture. In particular, this warm-up conjecture allows each permutation $\mathcal{P}_i$ to depend on the round index i.

Conjecture 2 (Round-dependent permutations)**.** *Let $(\mathcal{P}_i)_{i\in[T]}$ be randomized permutations on $\{0,1\}^n$ such that each output bit only depends on a constant number of input bits. If $\mathcal{P}_T \circ \cdots \circ \mathcal{P}_1$ is negligibly far from 4-wise independent, then $\mathcal{P}_T \circ \cdots \circ \mathcal{P}_1$ is also a pseudorandom permutation.*

For this construction, it is again useful to think of wires that carry values from a constant-sized alphabet Σ. The exact members of the alphabet are not too important, except for the following six special characters: the numbers 1 to 5 and the $\heartsuit$ symbol. Moreover, we will define addition over this alphabet by interpreting the symbols as elements of $\mathbb{Z}_{|\Sigma|}$, and addition over strings in Σ^n by interpreting them as vectors in $\mathbb{Z}^n_{|\Sigma|}$.

Our construction is split into two phases. Throughout the first phase, we preserve the property that the strings of the form $\heartsuit^{n-1}x$ remain invariant for all $x \in [5]$. We call these strings *special*. In this phase, we also make sure that the permutation is 4-wise independent on all inputs but $\{\heartsuit^{n-1}x | x \in [5]\}$. Then, in the second phase, all that is left to do is to map the special strings to five distinct strings $w^1, \ldots, w^5$ such that they are 4-wise independent but all five of them are efficiently distinguishable from random.

As a result, phases one and two combined result in a 4-wise independent distribution, but an attacker can query the permutation at the five special strings $\{\heartsuit^{n-1}x | x \in [5]\}$ and receive $w^1, \ldots, w^5$, which are efficiently distinguishable from random. This concludes that the resulting randomized permutation is not pseudorandom.

In the following, we discuss how to implement these high-level ideas in permutations with constant locality.

Randomizing non-special strings. Our starting point is the permutation families of random reversible brickwork circuits as analyzed in [CHH+24, GHKO25]. In these permutations, we have n wires arranged on a one-dimensional line to which we apply layers of 3-local random permutations (gates) that touch neighboring wires. The layers of local gates together form a brickwork-like pattern, hence the name (see Figure 1).

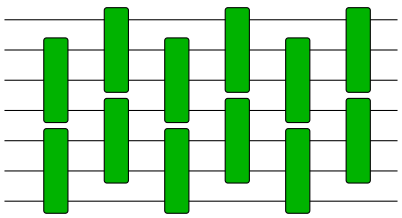

Figure 1: 7 Layers of Brickwork Permutation on 7 wires

We would like to apply brickwork permutations to inputs that are not special strings. While an input v being special is a global property, we can check it locally by picking a random wire $i^* \in [n]$ and then only applying a brickwork layer of gates to all other wires if v_{i^*} is inconsistent with a special string. Specifically, v_{i^*} is inconsistent with a special string if (1) $i^* \in [n-1]$ and $v_{i^*} \neq \heartsuit$, or (2) $i^* = n$ and $v_{i^*} \notin [5]$. We use the letter $\mathcal{B}$ to refer to this "conditional brickwork" permutation.

Our goal now is to show that applying this conditional brickwork permutation will randomize any inputs that are not special. The standard way to achieve this is by showing that the Markov chain defined by this random process has a large enough spectral gap. Prior work studying similar brickwork-related permutations [BHH16, CHH$^+$24, GHKO25] involves lengthy arguments and non-trivial results from expander graphs or quantum many-body physics.

We would like instead to use the spectral gap of the brickwork architecture as a black-box, and use that to infer that the conditional brickwork randomizes the non-special strings. A first observation is that whenever we have a string that is inconsistent with a special string in many places, it is very likely that an almost full layer of brickwork gates is applied to it. On the other hand, a string that only differs from a special string in one position will likely remain unchanged after $o(n)$ applications of the conditional brickwork process.

One way to achieve our goal is to consider the first few rounds as "warm-up", which will take every input to a string that is far from a special string. Then, the conditional brickwork circuit acts in a very similar way to the original brickwork architecture.

Formally, we define the distance $D(v)$ of a string v from a special string to be the number of indices i^* at which v is inconsistent with a special string. Then, at every round during the first phase, a layer of random brickwork gates gets applied to $n-1$ of the wires with probability $D(v)/n$. Thus, after $\mathrm{poly}(n)$ many such rounds, every non-special input will have a brickwork layer applied to it except with negligible probability.

Once all inputs have gone through at least one brickwork layer with high probability, the "warm-up" rounds are done. Indeed, the marginal min-entropy of each input is very high (since at least $n-1$ of the wires have been completely randomized). Thus, a Chernoff bound argument implies that the expected number of $\heartsuit$ symbols in each input will be small, especially when the alphabet size $|\Sigma|$ is a large enough constant. We will call the strings that have a small number of $\heartsuit$ symbols as "high-distance" strings. Then we can conclude the 4-wise independence of the non-special strings by establishing a lower bound on the spectral gap of the Markov chain that is defined by this conditional brickwork step. It turns out that when we restrict the state space of this Markov chain to high-distance strings, its spectral gap can be shown to be at least a constant, by using a "comparison by multicommodity flows" argument with the original brickwork Markov chain of [GHKO25]. Since an overwhelming fraction of the state space consists of inputs that are high-distance, and the spectral gap of the conditional brickwork in this space is a constant, we can show that an additional $\mathrm{poly}(n)$ rounds suffice for the distribution of the outputs on the non-special strings to become almost 4-wise independent.

Randomizing special strings. The second phase only consists of a single round, which we call $\mathcal{W}$, in which we map the five special strings $\{\heartsuit^{n-1}x | x \in [5]\}$ to five distinct strings $w^1, \ldots, w^5$ such that the distribution of $(w^1, \ldots, w^5)$ is close to 4-wise independent, but efficiently distinguishable from 5-wise independent.

In more detail, we use the randomness of this round to sample 4 distinct strings

$$w^1, w^2, w^3, w^4 \sim \Sigma^n$$

uniformly at random, under the condition that there exists an index $i^* \in [n-1]$ for which the symbols $w^1_{i^*}, w^2_{i^*}, w^3_{i^*}, w^4_{i^*}$ and $(w^1_{i^*} + w^2_{i^*} + w^3_{i^*} + w^4_{i^*})$ are distinct. Note that this condition is met with overwhelming probability over a uniformly random choice of w^1, w^2, w^3, w^4. We define $w^5 = w^1 + w^2 + w^3 + w^4$ and observe that the distribution of w^1, w^2, w^3, w^4, w^5 is almost 4-wise independent, but not 5-wise independent. This is because the 5 strings satisfy a linear relation that only a negligible fraction of 5-tuples satisfy.

Now that we have these five strings, we present the mechanism with which we map $\heartsuit^{n-1}x$ to w^x for all $x \in [5]$. For every wire $i \notin \{i^*, n\}$ we sample five permutations $\pi_{i,x}$ that take wire i, control on the value x of wire n for each $x \in [5]$, and output

$$\pi_{i,x}(\heartsuit) = w^x_i.$$

Further, we sample a uniformly random permutation $\pi_{i^*} : \Sigma^2 \to \Sigma^2$ such that $\pi_{i^*}(\heartsuit, x) = w^x_{i^*} w^x_n$. Because we have that the symbols $w^1_{i^*}, w^2_{i^*}, w^3_{i^*}, w^4_{i^*}$ and $w^1_{i^*} + w^2_{i^*} + w^3_{i^*} + w^4_{i^*}$ are distinct, π_{i^*} is, indeed, a permutation.

Finally we apply π_{i,v_n} to v_i for all $i \notin \{i^*, n\}$ and then we apply π_{i^*} to $v_{i^*} v_n$. Overall, this has the effect that each special input $\heartsuit^{n-1}x$ is mapped to w^x. The output wire $i \notin \{i^*, n\}$ depends on input wires i and

n and the output wires i^* and n depend on the input wires i^* and n together. Therefore, phase two has constant locality.

2.2 Round-independent permutations

In this section, we present how to extend the previously elucidated ideas to a full refutation of Conjecture 1. The only difference from Conjecture 2 to Conjecture 1 is that the permutation has to be the same, independent of the round. In Section 2.1, we crucially relied on the knowledge of the round to make sure that we apply the second phase, which attends to the special strings, exactly in the final round.

Counter. A natural solution to this problem is to keep track of which round you are in by reserving one wire for a counter. Then, in each round of the permutation, one would increase this counter wire by 1.

This solution, however, runs into the simple barrier that to guarantee 4-wise independence, or even pairwise independence, the round function needs to be iterated $\Omega(\log(n))$ times. That is because each round only has constant locality, and to guarantee pairwise independence, every input wire needs to influence every output wire.

If we were to use a single wire as a binary counter, then the wire would have to have an alphabet of super-constant size $\Omega(\log\log(n))$. Therefore, we implement a unary counter. Specifically, we introduce another significant character ♣ to our alphabet and change our set of special strings to

$$\{\heartsuit^i\clubsuit^{n-1-i}x \mid i \in [n-2], x \in [5]\}.$$

The idea is to increase the number of ♡'s in every round by one, and when $\heartsuit^{n-2}\clubsuit x$ is reached, apply the second phase as discussed in Section 2.1. This design has the advantage over a binary counter that checking if a special string v has the form $\heartsuit^{n-2}\clubsuit x$ for some $x \in [5]$ can be done locally, simply by verifying that $v_{n-2} = \heartsuit$. No other *special* string has this property. Therefore, if we apply phase two every time $v_{n-2} = \heartsuit$, we might erroneously apply the second phase to non-special strings; crucially, we apply it to special strings if and only if $v = \heartsuit^{n-2}\clubsuit x$ for some $x \in [5]$.

Implementing counters, locally. Next, we discuss how to implement such a counter with local permutations.

Assume that we know the remainder of the round in which we are, modulo 3. This holds without loss of generality because we can always create three rounds and combine them into one. If each round had locality at most ℓ, then the merged rounds have locality at most ℓ^3.

We start with the case that we are in a round r such that $r \equiv 1 \pmod 3$ and want to move the counter from $\heartsuit\clubsuit^{n-2}1$ to $\heartsuit^2\clubsuit^{n-3}1$. We choose the permutation $\pi : \Sigma^3 \to \Sigma^3$ that satisfies:

- π maps $\heartsuit\clubsuit\clubsuit$ to $\heartsuit\heartsuit\clubsuit$,
- π maps $\heartsuit\heartsuit\clubsuit$ to $\heartsuit\clubsuit\clubsuit$, and
- π maps every other string $x_1x_2x_3$ to itself $x_1x_2x_3$.

Then we apply π to the input string v and output:[2]

$$\pi(v_1, v_2, v_3), \pi(v_4, v_5, v_6), \ldots, \pi(v_{n-3}, v_{n-2}, v_{n-1}), v_n.$$

This permutation, indeed, maps $\heartsuit\clubsuit^{n-2}1$ to $\heartsuit^2\clubsuit^{n-3}1$; it also maps $\heartsuit^4\clubsuit^{n-5}1$ to $\heartsuit^5\clubsuit^{n-6}1$, etc. Importantly, the way π is structured, there is no way for it to turn a special string into a non-special string or vice versa. Similarly, if $r \equiv 2 \pmod 3$, then we map v to

$$v_1, \pi(v_2, v_3, v_4), \ldots, \pi(v_{n-5}, v_{n-4}, v_{n-3}), v_{n-2}, v_{n-1}, v_n$$

and to

$$v_1, v_2, \pi(v_3, v_4, v_5), \ldots, \pi(v_{n-4}, v_{n-3}, v_{n-2}), v_{n-1}, v_n$$

if $r \equiv 0 \pmod 3$. We combine these three local permutations into one permutation $\mathcal{C}$ by composition. It holds that $\mathcal{C}$ forms a cycle over the $n-2$ special strings for each $x \in [5]$, only one of which has $v_{n-2} = \heartsuit$.

[2] We assume $n \equiv 1 \pmod 3$.

Detecting non-special strings. With our new definition of special strings, we also need to adjust the conditional brickwork step that locally detects and randomizes them using $\mathcal{B}$. In Section 2.1, the special strings were $\{\heartsuit^{n-1}x \mid x \in [5]\}$, and it was enough to look at one wire at a time and check if it looks like a special string, locally. That is because every string v that satisfies $v_i = \heartsuit$ for all $i \in [n-1]$ and $v_n \in [5]$ must be a special string.

Any condition that checks a single wire at a time cannot work, since the string $v = \heartsuit\clubsuit\heartsuit^{n-3}1$ is not special even though $v_i \in \{\heartsuit, \clubsuit\}$ for all $i \in [n-1]$, and $v_n \in [5]$. Fortunately, for our new set of special strings $\{\heartsuit^i\clubsuit^{n-1-i}x \mid i \in [n-2], x \in [5]\}$, a 2-local check suffices to detect a special string. In particular, it holds that a string v is special if and only if

- $v_1v_2 \in \{\heartsuit\heartsuit, \heartsuit\clubsuit\}$,
- $v_iv_{i+1} \in \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$ for all $i \in \{2, \ldots, n-3\}$,
- $v_{n-2}v_{n-1} \in \{\heartsuit\clubsuit, \clubsuit\clubsuit\}$, and
- $v_n \in [5]$.

All of these conditions are locally checkable. As a result, at each step, we will pick a random pair of consecutive wires and check if they satisfy any of the conditions above. If the two wires do not satisfy the condition above, we will apply a brickwork layer of gates to all other wires. This will be our new conditional brickwork layer.

Rate of convergence. Trying to follow the proof from Section 2.1, we can define the distance $D(v)$ to be the number of conditions that are violated. Then, every time we try to apply a conditional brickwork layer, the conditions will fail and apply a brickwork layer of gates with probability $D(v)/n$. This reveals a new issue: our unary counter can only count to $n-2$, but to make sure that we apply the first layer of brickwork gates with overwhelming probability may require $\omega(n)$ rounds, e.g., if we start from a non-special string with $D(v) = 1$.

Therefore, we introduce a permutation $\mathcal{M}$ that ensures the growth of D. We pick $\Theta(n)$ disjoint pairs of conditions c and wires i. On an input v, for every condition-wire pair (c, i), we will apply a random shift to v_i if v does not meet condition c. These modifications do not cause any changes to special strings because they meet all conditions.

Intuitively, sampling $\Theta(n)$ conditions will give us a constant probability of hitting a violated condition, even for strings v with a small distance $D(v)$ from a special string. Each such sampled violated string will randomize some random character of the string, which is very likely to increase $D(v)$ for strings with a small distance. In particular, we show that applying $\mathcal{M}$ to a string v with small distance $D(v) \in [1, n/d]$ (where d is a sufficiently large constant) will increase $D(v)$ exponentially fast with high probability. Once $D(v) = \Theta(n)$, the conditional brickwork layer can detect the difference and apply random gates to it with constant probability. Therefore, $\Theta(n)$ additional rounds of brickwork suffice to randomize the non-special strings.

Putting everything together. We describe how the different components fit together to construct the randomized permutation

$$\mathcal{P} = \mathcal{W} \circ \mathcal{C} \circ \mathcal{B}^m \circ \mathcal{M}.$$

Here m is some constant that we will specify later, and each permutation is chosen independently according to the distribution described above. As each permutation has constant locality and we are composing a constant number of permutations, their composition $\mathcal{P}$ has constant locality, too.

The main theorem is that the $(n-2)$-wise composition of the above permutation $\mathcal{P}$, for some constant m, is negligibly far from a 4-wise independent permutation, but it is very far from a 5-wise independent permutation, since the following relation

$$\begin{aligned}&\mathcal{P}^{n-2}(\heartsuit^{n-2}\clubsuit 5) = \\ &\mathcal{P}^{n-2}(\heartsuit^{n-2}\clubsuit 1) + \mathcal{P}^{n-2}(\heartsuit^{n-2}\clubsuit 2) + \mathcal{P}^{n-2}(\heartsuit^{n-2}\clubsuit 3) + \mathcal{P}^{n-2}(\heartsuit^{n-2}\clubsuit 4)\end{aligned}$$

always holds, and can be distinguished from random efficiently.

Our proof for the 4-wise independence of $\mathcal{P}^{n-2}$ follows closely the ideas developed above. We prove the 4-wise independence of the distribution of

$$(y^1, y^2, y^3, y^4) = (\mathcal{P}^{n-2}(x^1), \mathcal{P}^{n-2}(x^2), \mathcal{P}^{n-2}(x^3), \mathcal{P}^{n-2}(x^4))$$

by keeping track of the 2-norm of the distribution of the intermediate y values after each round r of the block cipher for $r \in [n-2]$. A standard application of the Cauchy-Schwarz inequality will imply a statistical distance bound from the uniform distribution. On this end, we show that after $O(\log^3 n)$ rounds, the distribution of y is concentrated on tuples of strings that have either high distance $D(\cdot)$ from a special string, or are special themselves. Since the set of high-distance strings is an overwhelming fraction of the space of 4-tuples of strings, it suffices to restrict our attention to these tuples and show that after $n-2$ rounds the tuple y is negligibly far from a uniformly random tuple of high-distance strings.

To bound the 2-norm of the intermediate distribution, we use spectral gap techniques. As mentioned, we relate the spectral gap of the conditioned brickwork circuit when restricted to high-distance strings to the spectral gap of the random brickwork circuit of [GHKO25] using the comparison method. A spectral gap implies a reduction of the 2-norm of the distribution for each application of the $\mathcal{B}$ layer. We choose the constant m to make the spectral gap of the $\mathcal{B}^m$ step low enough such that $n-2$ applications of the spectral gap can decrease the 2-norm to a sufficient amount, which is only possible since the spectral gap of $\mathcal{B}$ was constant in the first place. It is also worth noting that since all components of $\mathcal{P}$ are permutations, the 2-norm of the distribution cannot increase, and thus $\mathcal{M}, \mathcal{C}$, and $\mathcal{W}$ do not hurt our progress in the 2-norm.

We deal with the case when some of the inputs are special in a different way. In particular, by the construction of our block cipher, all special strings are mapped in a deterministic way until they get mapped to the w^x strings that are chosen to be almost 4-wise independent. This allows us to directly argue that their output is almost independent from the other outputs, and thus 4-wise independence is still achieved.

3 Preliminaries

We use $[n] = \{1, \ldots, n\}$. We are working over strings with n characters (or symbols) drawn from an alphabet Σ of constant size. The exact alphabet does not matter, but we will often refer to the following $k+3$ special symbols:

- Start and end symbols $\heartsuit, \clubsuit$.
- Special values $1, 2, \ldots, k+1$.

Given a string $x \in \Sigma^n$, we use x_i to refer to its i-th symbol, for $i \in [n]$. When referring to multiple inputs, we will often use a superscript to distinguish between them, e.g., $x^1, \ldots, x^k$.

Definition 3.1 (Randomized Permutation)**.** *Let $\mathcal{P}$ be an algorithm that takes a string $v \in \Sigma^n$ and randomness r. We call $\mathcal{P}$ a* randomized permutation *if $\mathcal{P}(\,\cdot\,;\mathsf{r})$ is a permutation of Σ^n for all r. Alternatively, this can be seen as a distribution over the permutations $\mathcal{P}(\,\cdot\,;\mathsf{r})$ where r is sampled uniformly at random. We will often omit r when it is chosen uniformly at random and write $\mathcal{P}(v)$ for $\mathcal{P}(v;\mathsf{r})$.*

We use the notation $\mathcal{P}^t(c)$ for t-wise composition of $\mathcal{P}$ with uniform and independent randomness. More precisely,

$$\mathcal{P}^t = \mathcal{P}(\,\cdot\,;\mathsf{r}_t) \circ \cdots \circ \mathcal{P}(\,\cdot\,;\mathsf{r}_1),$$

where $\mathsf{r}_1, \ldots, \mathsf{r}_t$ are chosen independently and uniformly at random.

Definition 3.2 (Distinct ordered k-tuples)**.** *Given a set A, we will define $A^{(k)}$ to be the set of* distinct ordered k-tuples *with elements in A. That is,*

$$A^{(k)} = \{(x^1, \ldots, x^k) \in A^k \mid x^i \neq x^j,\ \forall i, j \in [k], i \neq j\}.$$

3.1 Linear algebra and probability

Definition 3.3 (Stochastic Dominance)**.** *Let X, Y be random variables if $\Pr[X \geq t] \geq \Pr[Y \geq t]$ for all t then we say X has stochastic dominance over Y or $X \geq_{\mathsf{st}} Y$.*

Theorem 3.4 (Markov's inequality)**.** *Let X be a non-negative random variable and $a > 0$ then*

$$\Pr(X \geq a) \leq \mathbf{E}(X)/a.$$

Theorem 3.5 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $a \leq X_i \leq b$ for all $i \in [n]$ and $t > 0$ then*

$$\Pr(X_1 + \cdots + X_n \leq \mathbf{E}(X_1 + \cdots + X_n) - t) \leq \exp\left(-\frac{t^2}{n(b-a)^2}\right).$$

Theorem 3.6 (Multiplicative Chernoff bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables taking the values $\{0, 1\}$, $\delta > 0$, and $\mu = \mathbf{E}(X_1 + \cdots + X_n)$ then*

$$\Pr(X_1 + \cdots + X_n \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}.$$

Definition 3.7 (Vector p-norm)**.** *Given a distribution x over a space Ω, its* p-norm *is defined as*

$$\|x\|_p = \left(\sum_{i \in \Omega} x_i^p\right)^{1/p}.$$

In particular, the 1-norm of a vector is given by

$$\|x\|_1 = \sum_{i \in \Omega} |x_i|,$$

which is related to the total variation distance between two distributions that we see below.

Definition 3.8 (Total variation distance)**.** *The* total variation distance *between two distributions p, q over a set Ω is defined as*

$$d_{\mathrm{TV}}(p, q) = \frac{1}{2} \cdot \|p - q\|_1 = \frac{1}{2} \sum_{i \in \Omega} |p_i - q_i|.$$

We will also refer to it as the statistical distance*.*

Definition 3.9 (Negligibly far from k-wise independence)**.** *Given a set A, we say that a distribution x over $A^{(k)}$ is* negligibly far *(with respect to n)* from k-wise independence *if its total variation distance from the uniform distribution μ over $A^{(k)}$ is negligible in n. That is,*

$$d_{\mathrm{TV}}(x, \mu) \leq \mathrm{negl}(n).$$

We will also need the Cauchy-Schwarz inequality to translate between the 2-norm to the 1-norm.

Theorem 3.10 (Cauchy-Schwarz inequality)**.** *Let x be a distribution over a space Ω. Then*

$$\|x\|_1 \leq \sqrt{|\Omega|} \cdot \|x\|_2.$$

Notation 1. *Given a real symmetric matrix A, we will use $\lambda_i(A)$ to refer to its i-th largest eigenvalue.*

Definition 3.11 (Operator norm)**.** *For a symmetric matrix A of shape $m \times m$, its* operator norm *is equal to its maximum absolute eigenvalue*

$$\|A\|_\infty = \max\{|\lambda_1(A)|, \ldots, |\lambda_m(A)|\}.$$

Remark 3.12. *The operator norm of a symmetric matrix A with positive entries is upper bounded by the maximum row or column sum.*

We will also need the following standard results on matrices and their eigenvalues.

Theorem 3.13 (Weyl's inequality)**.** *Let A, B be real symmetric matrices of shape $m \times m$. Then for all $i \in [m]$,*

$$|\lambda_i(A + B) - \lambda_i(A)| \leq \|B\|_\infty.$$

Theorem 3.14 (Cauchy's interlacing theorem)**.** *Let A be a real symmetric matrix of shape $m \times m$ and projection matrix Π. Then*

$$\lambda_i(A) \geq \lambda_i(\Pi A \Pi).$$

3.2 Markov chains

Let $\mathcal{K}$ be the transition matrix of a Markov chain over a finite state space Ω, and let π denote its stationary distribution. We overload $\mathcal{K}$ to refer to the transition matrix, the edge set, but also the Markov chain itself.

Definition 3.15 (Reversible Markov chain)**.** *A Markov chain $\mathcal{K}$ on state space Ω is* reversible *if it satisfies*

$$\pi(x)\mathcal{K}(x, y) = \pi(y)\mathcal{K}(y, x), \quad \forall x, y \in \Omega.$$

Not all of the chains we consider in this paper are reversible, and we will mention their reversibility explicitly.

On the other hand, since we are dealing with permutations, it will hold that the Markov chains we consider will have a doubly stochastic transition matrix, and thus $\mathcal{K}^T\mathcal{K}$ is a symmetric matrix with real and non-negative eigenvalues. Moreover, the largest eigenvalue of $\mathcal{K}^T\mathcal{K}$ is equal to 1 and corresponds to the all-ones vector on the state space Ω.

Definition 3.16 (Spectral gap)**.** *Given a Markov chain $\mathcal{K}$ with a doubly stochastic transition matrix, we define its* spectral gap $\lambda(\mathcal{K})$ *as follows:*

$$\lambda(\mathcal{K}) = 1 - \lambda_2(\mathcal{K}^T\mathcal{K}).$$

Since the spectral gap is related to the second eigenvalue of a matrix, it can be used to bound the 2-norm of a distribution after one step of $\mathcal{K}$ (see, for example, [Din25, Proposition 2]).

Theorem 3.17 (Spectral gap implies 2-norm contraction)**.** *Let $\mathcal{K}$ be a Markov chain over a state space Ω with a doubly stochastic transition matrix. Moreover, let p be a probability distribution over Ω. Then*

$$\|\mathcal{K}p\|_2^2 \leq (1 - \lambda(\mathcal{K})) \cdot \|p\|_2^2 + \lambda(\mathcal{K}) \cdot \frac{1}{|\Omega|}.$$

Random reversible brickwork circuits. One of the Markov chains that is central in our analysis is the one that corresponds to a layer of a random reversible brickwork circuit. This has been recently studied by [GHKO25], in which the authors proved a constant spectral gap for this chain.

Definition 3.18 (Random reversible brickwork circuits)**.** *Let n be an integer such that $n \equiv 1 \pmod 3$, and let $n' = (n-1)/3$. The chain $\mathcal{R}^k_{\Sigma,n}$ on the state space $(\Sigma^n)^{(k)}$ maps the k-tuple of strings $(x^1, \ldots, x^k)$ to $y = (y^1, \ldots, y^k)$, by drawing a value $i \in \{1, 2\}$ and n' random permutations $\pi_0, \ldots, \pi_{n'-1}$. Each string y^r is defined as follows:*

- *If $i = 1$, then $y^r_n = x^r_n$.*
- *Else if $i = 2$, then $y^r_1 = x^r_1$.*
- *For all $c \in \{0, \ldots, n' - 1\}$, set*

$$(y^r_{3c+i}, y^r_{3c+i+1}, y^r_{3c+i+2}) = \pi_c(x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}).$$

We will often drop the dependence on the alphabet Σ for simplicity.

The chain we define above is not quite the same as the one considered in [GHKO25]. In particular, the authors consider a chain over an alphabet of size 2. It turns out that their spectral gap bound translates directly to our larger alphabet setting, albeit with a different constant. This can be seen by viewing each symbol in Σ as a bit string of length $\log|\Sigma|$, and each gate now touches $3\log|\Sigma|$ consecutive wires at once.

Theorem 3.19 (Corollary of [GHKO25])**.** *The spectral gap of $\mathcal{R}^k_n$ is a constant:*

$$\lambda(\mathcal{R}^k_n) = \Theta(1).$$

3.3 Comparison using multicommodity flows

We briefly describe the "comparison using multicommodity flows" method of Sinclair, Diaconis, and Saloff-Coste [Sin92, DSC93]. Let $\mathcal{K}^{\mathrm{ref}}$ and $\mathcal{K}^{\mathrm{tar}}$ be two reversible Markov chains on the same ground set Ω, with stationary distributions $\pi^{\mathrm{ref}}, \pi^{\mathrm{tar}}$. Our goal is to relate the spectral gap of $\mathcal{K}^{\mathrm{tar}}$ to the spectral gap of $\mathcal{K}^{\mathrm{ref}}$.

Definition 3.20 (Path)**.** *For an edge $(x, y) \in \mathcal{K}^{\mathrm{ref}}$, we define the* path γ_{xy} *to be the sequence*

$$(x = a_0, a_1, \ldots, a_t = y),$$

where each (a_i, a_{i+1}) is an edge in $\mathcal{K}^{\mathrm{tar}}$. We say that such a path has length $|\gamma_{xy}| = t$.

Let $\mathcal{P}_{xy}$ be the set of all simple paths connecting x to y, and $\mathcal{P} = \cup_{(x,y)\in\mathcal{K}^{\mathrm{ref}}}\mathcal{P}_{xy}$ be the set of all simple paths between all pairs of vertices that are connected by an edge in $\mathcal{K}^{\mathrm{ref}}$. Our goal is to construct a flow over the paths that connect all pairs of vertices connected by an edge in $\mathcal{K}^{\mathrm{ref}}$.

Definition 3.21 (Flow)**.** *A function $F : \mathcal{P} \to \mathbb{R}_{\geq 0}$ is called a* $(\mathcal{K}^{\mathrm{tar}}, \mathcal{K}^{\mathrm{ref}})$-flow *if for all edges $(x, y) \in \mathcal{K}^{\mathrm{ref}}$,*

$$\sum_{\gamma\in\mathcal{P}_{xy}} F(\gamma) = \mathcal{K}^{\mathrm{ref}}(x, y)\pi^{\mathrm{ref}}(x).$$

Given such a flow, we can obtain a relationship between the spectral gaps of $\mathcal{K}^{\mathrm{ref}}$ and $\mathcal{K}^{\mathrm{tar}}$. In particular, if the flow has low congestion, this intuitively means that one can "simulate" the edges of $\mathcal{K}^{\mathrm{ref}}$ using short paths in $\mathcal{K}^{\mathrm{tar}}$, and thus $\lambda(\mathcal{K}^{\mathrm{tar}})$ cannot be much smaller than $\lambda(\mathcal{K}^{\mathrm{ref}})$.

Formally, for each edge $(a, b) \in \mathcal{K}^{\mathrm{tar}}$, let $\mathcal{P}(a, b) = \{\gamma \in \mathcal{P} \mid (a, b) \in \gamma\}$ be the set of paths that use the edge (a, b). Then the quantity $A(F)$ in the following theorem captures the congestion of a flow.

Theorem 3.22 ([DSC93, Theorem 2.3])**.** *Let $\mathcal{K}^{\mathrm{tar}}$, $\mathcal{K}^{\mathrm{ref}}$ be Markov chains on a finite set Ω reversible with respect to $\pi^{\mathrm{tar}}, \pi^{\mathrm{ref}}$ respectively. For any $(\mathcal{K}^{\mathrm{tar}}, \mathcal{K}^{\mathrm{ref}})$-flow F, the spectral gaps satisfy*

$$\lambda(\mathcal{K}^{\mathrm{tar}}) \geq \frac{1}{A(F)} \cdot \lambda(\mathcal{K}^{\mathrm{ref}})$$

with

$$A(F) = \max_{(a,b)\in\mathcal{K}^{\mathrm{tar}}} \left\{ \frac{1}{\pi^{\mathrm{tar}}(a)\mathcal{K}^{\mathrm{tar}}(a, b)} \sum_{\gamma\in\mathcal{P}(a,b)} |\gamma| \cdot F(\gamma) \right\}.$$

4 Block cipher construction

In this section, we outline the construction of the block cipher that proves our main result in Theorem 4.1 and refutes Conjecture 1.

Theorem 4.1. *Let k be a positive integer. There exists a block cipher $\mathcal{P}$ that acts on length-n strings from a constant-sized alphabet Σ such that:*

(i) Each round of $\mathcal{P}$ has constant locality. That is, each output bit depends on a constant number of input bits from the round before.

(ii) Composing $(n-2)$ independent copies of $\mathcal{P}$ constructs a randomized permutation that is negligibly far (statistically) from k-wise independence.

(iii) There exists an efficient distinguishing attack against $\mathcal{P}^{n-2}$ that uses $k+1$ inputs and can distinguish their outputs from a uniformly random tuple of $k+1$ distinct strings with constant probability.

Proof. The construction of our block cipher is given below. We observe that in each step, every output bit only depends on a constant number of input bits, and thus each round of $\mathcal{P}$ has constant locality.

The k-wise independence of the block cipher is proved in Theorem 6.12.An efficient attack against the block cipher that uses $k+1$ inputs is highlighted in Theorem 5.1. □

$$\to \boxed{\mathcal{M}} \to \underbrace{\boxed{\mathcal{B}} \to \cdots \to \boxed{\mathcal{B}}}_{m \text{ times}} \to \boxed{\mathcal{C}} \to \boxed{\mathcal{W}} \to$$

Figure 2: Schematic representation of one round of the $\mathcal{P}$ permutation.

The construction consists of four parts with differing roles in the proof. For all these permutations, we assume for simplicity that $n \equiv 1 \pmod 3$. The breakdown of one round of the block cipher into four parts is as follows:

Algorithm 1 $\mathcal{P}$ on input x

1: Let $x^{(a)} \leftarrow \mathcal{M}(v)$
2: Let $x^{(b)} \leftarrow \mathcal{B}^m(v)$
3: Let $x^{(c)} \leftarrow \mathcal{C}(v)$
4: Let $x^{(d)} \leftarrow \mathcal{W}(v)$
5: **return** $x^{(d)}$

In addition to the four parts, our construction is parametrized by a constant m that will be chosen later. This parameter is needed to make sure that $\mathcal{P}$ reaches negligibly far from k-wise independence in less than $(n-2)$ rounds, without any hidden constant factors.

Remark 4.2 (Conditional permutations)**.** *Similar to some block cipher constructions (e.g. the Feistel cipher [LR88]), our design includes steps that apply a conditional permutation, that is, a permutation whose action depends on the value of some wire. Such a permutation is well-defined, as long as it does not modify the value that it conditions on. In particular, we will often apply permutation π_T on a subset of wires $S \subseteq [n]$, where π_T depends on a disjoint subset of wires $T \subseteq [n]$.*

Before we go into the construction, we introduce several pieces of notation.

Definition 4.3 (Counter and value of input)**.** *Given an input $x \in \Sigma^n$, we will refer to its first $n-1$ characters as its* counter*, and to the last character as its* value*.*

Definition 4.4 (Valid counter)**.** *A* valid counter *has the form* $\heartsuit^i\clubsuit^{n-1-i}$ *for any* $i \in [n-2]$*. We extend this to say that a string with a valid counter is a* valid *string. Otherwise, a string and its counter are called* invalid*.*

We emphasize that a valid counter must start with $\heartsuit$ and end with $\clubsuit$.

Definition 4.5 (Detectable index)**.** *For a counter* $c \in \Sigma^{n-1}$*, we call an index* i detectable *in* c *if*

- $i \in [2, n-3]$ *and* $c_i c_{i+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$*,*
- $i = 1$ *and* $c_1 c_2 \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit\}$*, or*
- $i = n-2$ *and* $c_{n-2} c_{n-1} \notin \{\heartsuit\clubsuit, \clubsuit\clubsuit\}$*.*

This captures whether an adjacent pair of symbols does not look like a valid counter.

Definition 4.6 (Distance of counter from valid)**.** *For a counter* $c \in \Sigma^{n-1}$*, we define its* distance $D(c)$ from valid *as the number of detectable indices. Specifically,*

$$D(c) = \Big|\big\{i \mid i \text{ is detectable in } c\big\}\Big|.$$

Similarly, whenever we refer to the distance of a string, we mean the distance of its counter from valid.

Definition 4.7 (Special input)**.** *An input* $x \in \Sigma^n$ *is called* special *if it has a valid counter, and its value is in the special set* $[k+1]$*. We denote the set of special strings using* $\mathsf{Special}$*.*

We proceed with a detailed description of each step of the block cipher.

4.1 Mixing step

The first step of a round of the block cipher is the mixing step $\mathcal{M}$. This randomized permutation tries to increase the distance $D(\cdot)$ from valid of the non-special strings, while leaving the special strings untouched. The random coins that $\mathcal{M}$ takes are of the form:

- Letters $a_1, \ldots, a_{n/5}$ drawn independently and uniformly at random from Σ.
- Indices $y_1, \ldots, y_{n/5}$ drawn uniformly without replacement from $[n-2]$.
- Indices $z_1, \ldots, z_{n/5}$ drawn uniformly without replacement from $[n-1] \setminus (\cup_{i \in [n/5]} \{y_i, y_i + 1\})$.

Intuitively, each application of the $\mathcal{M}$ permutation will check $n/5$ indices y_i and check if they are detectable. If y_i is detectable, then it applies a random linear shift by a_i to the z_i index. The choice of having $\Theta(n)$ indices is to make sure that even non-special strings with a small distance can increase their distance in a constant number of rounds.

Algorithm 2 $\mathcal{M}$ on input x with randomness $\{a_i\}, \{y_i\}, \{z_i\}$

```
for i ∈ [n/5] do
    if y_i is detectable in x then
        x_{z_i} ← x_{z_i} + a_i
    end if
end for
return x
```

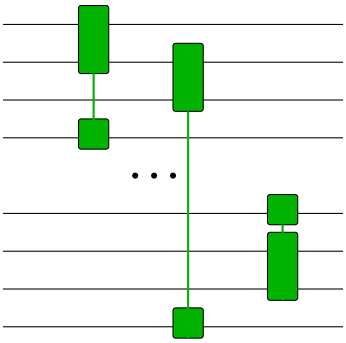

Figure 3: An example of how the mixing step can be implemented. Each gate corresponds to some i in Step 1 of Algorithm 2 and touches a consecutive pair of wires that corresponds to y_i and $y_i + 1$, and a single wire that corresponds to z_i. The gate only modifies wire z_i, conditioned on the value of the other two wires. The gates may overlap on the y_i wires, and since these wires are not modified, all $n/5$ of them can be implemented in one layer. They are illustrated above in different layers for clarity.

4.2 Conditional brickwork step

The second step of a round of the block cipher is the conditional brickwork step $\mathcal{B}$. This randomized permutation applies a layer of random brickwork gates to inputs with a high $D(\cdot)$ value with good probability. This step is repeated m times in each round, where m is a constant that we will choose later.

The random coins that $\mathcal{B}$ takes are of the form:

- Index i^* drawn uniformly from $[n-1]$. If $i^* \equiv 1, 2 \pmod 3$, then set $i = 1$, else set $i = 2$.
- Uniformly random permutations $\pi_0, \ldots, \pi_{n'-1}$ on Σ^3, where $n' = (n-1)/3$.

Algorithm 3 $\mathcal{B}$ on input x with randomness $i^*, i, \{\pi_c\}$

```
if i* is detectable in x then
    for c ∈ {0, ..., n' − 1} do
        if {i*, i* + 1} ⊈ {3c + i, 3c + i + 1, 3c + i + 2} then
            Apply π_c to wires (3c + i, 3c + i + 1, 3c + i + 2) of x.
        end if
    end for
end if
return x
```

On a high level, this step chooses a random pair of consecutive wires $i^*, i^* + 1$ and if it detects that they are different than a valid counter, it applies a brickwork layer to all wires except the triple that includes i^*, i^*+1. In particular, there exists a unique index c, which satisfies $\{i^*, i^*+1\} \subseteq \{3c+i, 3c+i+1, 3c+i+2\}$, such that π_c is never applied, because it would act on wires $i^*, i^* + 1$.

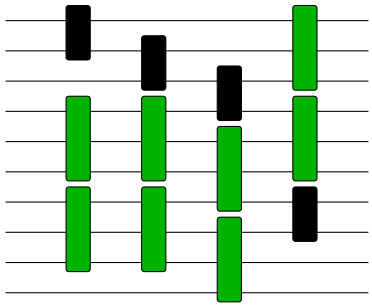

Figure 4: An example of how the conditional brickwork step can be implemented. In the above we show 4 layers of random brickwork, e.g. $\mathcal{B}^4$. Each layer contains a dark square that corresponds to the pair i^*, i^*+1 that is used to detect if i^* is detectable. The remaining wires then have gates according to the brickwork layout. Depending on the remainder of i^* modulo 3, the gates have an offset of 0 or 1, as shown in the figure.

4.3 Increase counter step

The third step of the block cipher round is the "increase counter" step $\mathcal{C}$, which increments the counter part of the special strings. This step is deterministic. In particular, we define the permutation π_{count} on Σ^3 that advances the counter. That is, π_{count} implements the following mapping:

$$\heartsuit\clubsuit\clubsuit \mapsto \heartsuit\heartsuit\clubsuit, \qquad \heartsuit\heartsuit\clubsuit \mapsto \heartsuit\clubsuit\clubsuit, \qquad a_1a_2a_3 \mapsto a_1a_2a_3,$$

for all other $a_1a_2a_3$.

Algorithm 4 $\mathcal{C}$ on input x

1: Let $x \leftarrow (\pi_{\text{count}}(x_1, x_2, x_3), \pi_{\text{count}}(x_4, x_5, x_6), \ldots, \pi_{\text{count}}(x_{n-3}, x_{n-2}, x_{n-1}), x_n)$
2: Let $x \leftarrow (x_1, \pi_{\text{count}}(x_2, x_3, x_4), \ldots, \pi_{\text{count}}(x_{n-5}, x_{n-4}, x_{n-3}), x_{n-2}, x_{n-1}, x_n)$
3: Let $x \leftarrow (x_1, x_2, \pi_{\text{count}}(x_3, x_4, x_5), \ldots, \pi_{\text{count}}(x_{n-4}, x_{n-3}, x_{n-2}), x_{n-1}, x_n)$
4: **return** x

Remark 4.8. *Applying the increase counter step $\mathcal{C}$ to an input x does not change its distance from a valid counter. That is,*

$$D(\mathcal{C}(x)) = D(x).$$

4.4 Randomize special strings step

The fourth and final part of a round of $\mathcal{P}$ permutation we define is the "randomizing special strings" $\mathcal{W}$, which maps the $k+1$ special strings to the random strings $w^1, \ldots, w^{k+1}$ that are sampled from an approximately k-wise independent distribution, but they are efficiently distinguishable from $(k+1)$-wise independent. The random coins that $\mathcal{W}$ takes are of the form:

- Strings $w^1, \ldots, w^k, w^{k+1} \in \Sigma^n$. Here, the first k string are sampled uniformly at random from Σ^n such that there exists an index $i^* \in [n-3]$ for which $w^1_{i^*}, \ldots, w^k_{i^*}, w^{k+1}_{i^*}, \heartsuit, \clubsuit$ are all distinct.
- Permutations $\pi_{i,j} : \Sigma \to \Sigma$ uniformly at random for all $i \in [n-1]$ and $j \in \Sigma$. If $j \in [k+1]$, the sampled permutations $\pi_{i,j}$ must map $\heartsuit$ symbols to the respective symbol of w^j. Formally:

  $$\pi_{i,j}(\heartsuit) = w^j_i \quad \text{for } i < n-1 \qquad \text{and} \qquad \pi_{n-1,j}(\clubsuit) = w^j_{n-1}.$$

- Permutation $\tau : \Sigma^3 \to \Sigma^3$ uniformly at random such that for all $x \in [k+1]$, the permutation maps:

  $$\heartsuit\heartsuit x \mapsto w^x_{i^*} w^x_{n-2} w^x_n, \qquad w^x_{i^*} w^x_{n-2} w^x_n \mapsto \heartsuit\heartsuit x, \qquad a_1a_2a_3 \mapsto a_1a_2a_3,$$

  for all other $a_1a_2a_3$.

Algorithm 5 $\mathcal{W}$ on input x with randomness $\{w^i\}$, $\{\pi_{i,j}\}$, τ

1: **if** $x_{i^*} = \heartsuit$, $x_{n-2} = \heartsuit$, and $x_n \in [k+1]$ **then**
2: Let s be such that

$$s \leftarrow \pi_{1,x_n}(x_1), \ldots, \pi_{i^*-1,x_n}(x_{i^*-1}), x_{i^*}, \\ \pi_{i^*+1,x_n}(x_{i^*+1}), \ldots, \pi_{n-3,x_n}(x_{n-3}), x_{n-2}, \pi_{n-1,x_n}(x_{n-1}), x_n.$$

3: **else**
4: $s \leftarrow x$
5: **end if**
6: Let $o_{i^*}, o_{n-2}, o_n \leftarrow \tau(s_{i^*}, s_{n-2}, s_n)$
7: **return** $s_1, \ldots, s_{i^*-1}, o_{i^*}, s_{i^*+1}, \ldots, s_{n-3}, o_{n-2}, s_{n-1}, o_n$.

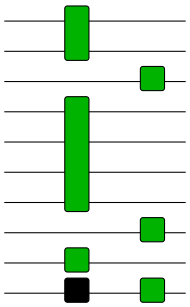

Figure 5: A possible implementation of $\mathcal{W}$ with $i^* = 3$. The first layer will condition on the last wire and modify all wires except $\{i^*, n-2, n\}$ using the π permutations. The second layer will then apply the τ permutation to the remaining wires.

In particular, Step 1 checks if the string looks like a special string whose counter has reached the last value (i.e. $\heartsuit^{n-2}\clubsuit$). In that case, it applies a random permutation that maps a finished counter to one of the random strings w^j. The index of the random string j is chosen by having π depend on the last symbol of the string. To make sure this conditional permutation is well-defined, from Remark 4.2, π will not be acting on the wires $n-2$ and n.

Then, the permutation τ is chosen such that it maps wires $n-2$ and n to the correct values, and keeps them unchanged if they don't correspond to a finished counter. Of course, for such a τ to be valid, it must be the case that $w^x_{n-2}w^x_n$ is different for all $x \in [k+1]$. This will make the distribution of the w^j's far from k-wise independent, since we expect a collision between any two fixed indices with constant probability. This is why we introduce the additional wire i^* in τ. By making τ act on the wires $i^*, n-2, n$, we deduce that the w^j's must have at least one index in which they all differ, which happens with probability $1 - \mathrm{negl}(n)$.

4.5 Remarks

Lemma 4.9. *The distribution of the strings $w^1, \ldots, w^k$ sampled in $\mathcal{W}$ is* $\mathrm{negl}(n)$*-far from the uniform distribution on $(\Sigma^n)^{(k)}$ when k is constant, and the alphabet size satisfies $|\Sigma| \geq (k+3)^2$.*

Proof. We will show that if we choose the k strings $w^1, \ldots, w^k$ at random with replacement, then there always exists some index $i^* \in [n-3]$ for which $w^1_{i^*}, \ldots, w^k_{i^*}, w^{k+1}_{i^*}, \heartsuit, \clubsuit$ are all distinct, except with negligible probability. This implies the statement of the lemma, since sampling with replacement is negligibly far from sampling without replacement whenever k is constant.

For uniformly random $w^1, \ldots, w^k$, the probability that there exists a collision between the characters $w^1_i, \ldots, w^k_i, w^{k+1}_i, \heartsuit, \clubsuit$ is at most $\binom{k+3}{2}/|\Sigma| \leq \frac{1}{2}$. Thus, the probability that all indices $i \in [n-3]$ have at least one collision is $(1/2)^{n-3} = \mathrm{negl}(n)$. □

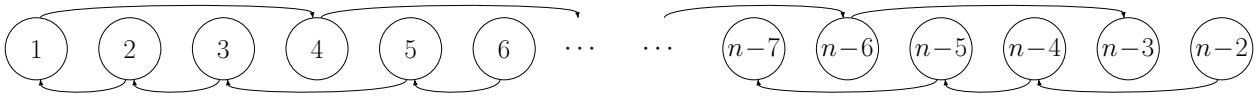

Figure 6: A schematic representation of the deterministic mapping between the special strings. A node with label i corresponds to the string $\heartsuit^i\clubsuit^{n-1-i}x$. Starting from $n-2$ will allow us to reach node $n-3$ in $n-2$ rounds. From node $n-3$, we enter the second phase and get mapped to w^x.

Remark 4.10. *If x is a special string, then x remains unchanged after the mixing $\mathcal{M}$ and conditional brickwork $\mathcal{B}$ steps.*

Since special strings do not change after mixing and conditional brickwork steps, we direct our attention to the last two steps $\mathcal{C}, \mathcal{W}$. We observe that any special string whose counter is not equal to $\heartsuit^{n-2}\clubsuit$ will not go into Step 2 of $\mathcal{W}$, and the permutation τ in Step 6 will not modify it either. Similarly, $\mathcal{C}$ will only modify its counter in a deterministic way.

On the other hand, a special string with the counter $\heartsuit^{n-2}\clubsuit$ will go into Step 2 of $\mathcal{W}$ in the next counter step and get mapped to some w^j, which is chosen (almost) uniformly at random. This is the first, and sometimes the only step of the block cipher, in which a special string undergoes a non-deterministic mapping. For this reason, we give this event the name "second phase".

Definition 4.11 (Second phase)**.** *An input x goes into* second phase *when it goes through Step 2 of $\mathcal{W}$.*

We show in Section 5 below that every special string will pass to phase two in at most $n-2$ rounds.

5 The distinguishing attack

In this section, we show that on specific inputs, the $(n-2)$-wise composition of $\mathcal{P}$ is not a pseudorandom permutation.

Theorem 5.1. *There exists an efficient distinguishing attack against $\mathcal{P}^{n-2}$ that uses $k+1$ inputs and can distinguish their outputs from a uniformly random tuple of $k+1$ distinct strings.*

Proof. The efficient attacker will use the inputs $\heartsuit^{n-2}\clubsuit x$ for all $x \in [k+1]$. Let the outputs of these inputs be $z^1, \ldots, z^{k+1}$. Then the distinguisher will output "Pseudorandom" if z^{k+1} is equal to the sum of z^1 up to z^k, and "Random" otherwise. It is clear that this distinguisher is efficient. We proceed by analyzing its distinguishing advantage.

Throughout this proof, we will be considering only special strings (Definition 4.7). Therefore, they have a valid counter. As per Remark 4.10, the steps $\mathcal{M}$ and $\mathcal{B}$ do not change any input that has a valid counter. Consequently, it is enough to consider the changes that $\mathcal{C}$ and $\mathcal{W}$ apply to these strings. To understand the following it is useful to remember that $n \equiv 1 \pmod 3$. We identify the following cases:

- Let $i \in [5, n-2]$, $i \equiv 2 \pmod 3$. Then for all $x \in [k+1]$ we have
$$\mathcal{P}(\heartsuit^i\clubsuit^{n-i-1}x) = \heartsuit^{i-2}\clubsuit^{n-i+1}x.$$
- Let $i \in [6, n-4]$, $i \equiv 0 \pmod 3$. Then for all $x \in [k+1]$ we have
$$\mathcal{P}(\heartsuit^i\clubsuit^{n-i-1}x) = \heartsuit^{i-1}\clubsuit^{n-i}x.$$
- It holds that for all $x \in [k+1]$:
$$\mathcal{P}(\heartsuit\heartsuit\heartsuit\clubsuit^{n-4}x) = \heartsuit\heartsuit\clubsuit^{n-3}x,$$
and
$$\mathcal{P}(\heartsuit\heartsuit\clubsuit^{n-3}x) = \heartsuit\clubsuit^{n-2}x.$$

- Let $i \in [n-6]$, $i \equiv 1 \pmod 3$. Then for all $x \in [k+1]$ we have

$$\mathcal{P}(\heartsuit^i \clubsuit^{n-i-1} x) = \heartsuit^{i+3} \clubsuit^{n-i-4} x.$$

- It holds that

$$\mathcal{P}(\heartsuit^{n-3} \clubsuit\clubsuit x) = \mathcal{W}(\heartsuit^{n-2} \clubsuit x) = w^x,$$

  for all $x \in [k+1]$, where $w^1 + \cdots + w^k = w^{k+1}$.

We conclude that for all $x \in [k+1]$, there exist strings $w^1, \ldots, w^{k+1}$ such that $w^1 + \cdots + w^k = w^{k+1}$, and $\mathcal{P}^{n-2}(\heartsuit^{n-2} \clubsuit x) = w^x$, or equivalently that

$$\mathcal{P}^{n-2}(\heartsuit^{n-2} \clubsuit 1) + \cdots + \mathcal{P}(\heartsuit^{n-2} \clubsuit k) = \mathcal{P}(\heartsuit^{n-2} \clubsuit (k+1)).$$

For a random permutation $\mathcal{P}_{\text{rand}}$ the probability that

$$\mathcal{P}_{\text{rand}}(\heartsuit^{n-2} \clubsuit 1) + \cdots + \mathcal{P}_{\text{rand}}(\heartsuit^{n-2} \clubsuit k) = \mathcal{P}_{\text{rand}}(\heartsuit^{n-2} \clubsuit (k+1))$$

is exponentially small in n. Thus, the claim follows. □

6 Reaching k-wise independence

In this subsection, we show that the block cipher $\mathcal{P}$ becomes almost k-wise independent after $n-2$ rounds for some constant value of m. The proof will follow by dividing the $n-2$ rounds into two parts: the first $O(\log^3 n)$ rounds will be referred to as the "warm-up" and the final $n - O(\log^3 n)$ rounds will be the "core" rounds.

The reason for this division is Lemma 6.4, in which we show that after the warm-up rounds, the probability mass is concentrated on a specific subset of inputs. In particular, any input x^i that is special initially will be either special or have a high distance from a valid counter after $\Omega(\log^3 n)$ rounds with high probability. On the other hand, any non-special input x^i will have a high distance from a valid counter with high probability.

Definition 6.1 (High-distance k-tuple)**.** *We call a k-tuple of distinct strings $(x^1, \ldots, x^k)$ to be a* high-distance k-tuple *if $D(x^i) \geq n(1 - 3/|\Sigma|)$ for all $i \in \{1, \ldots, k\}$. We denote the set of high-distance k-tuples as $\mathsf{HighDist}_n^k$.*

It is not hard to show that an overwhelming fraction of strings are high-distance strings.

Lemma 6.2. *Let $z = (z^1, \ldots, z^k) \in (\Sigma^n)^{(k)}$ be sampled uniformly at random from distinct k-tuples for constant k. Then*

$$\Pr[z \notin \mathsf{HighDist}_n^k] \leq \mathrm{negl}(n).$$

Proof. It suffices to show that a single z^i is not high-distance with high probability, and the lemma statement follows by applying a union bound over all k inputs. It holds that the marginal distribution of each z^i is uniform over Σ^n.

The first observation is that for a z^i to be low-distance, it must have at least $3n/|\Sigma| + O(1)$ symbols that are equal to $\heartsuit$ or $\clubsuit$. Sampling z^i uniformly at random means that each symbol is in $\{\heartsuit, \clubsuit\}$ with probability $2/|\Sigma|$. Thus, the number of symbols in z^i that are in the set $\{\heartsuit, \clubsuit\}$ is a sum of independent indicator random variables with mean $2/|\Sigma|$. A standard application of the Chernoff bound implies that the probability that this random variable is at least $3n/|\Sigma| + O(1)$ is exponentially small in $\Theta(n)$. □

The same proof can be used to extend this result to the case when we are sampling z from a distribution with sufficiently high min-entropy.

Corollary 6.3. *Let $z = (z^1, \ldots, z^k) \in (\Sigma^n)^{(k)}$ be sampled from a distribution such that each z^i has min-entropy at least $n \log |\Sigma| - c$. Then for a constant k,*

$$\Pr[z \notin \mathsf{HighDist}_n^k] \leq 2^c \cdot \mathrm{negl}(n).$$

We are now ready to formally prove that after the first $O(\log^3 n)$ rounds, the probability distribution of the intermediate strings of the block cipher is concentrated on special and high-distance strings.

Lemma 6.4. *Let $x = (x^1, \ldots, x^k)$ be an input k-tuple to $\mathcal{P}$, and $S \subseteq [k]$ be the subset of inputs that are special. Then, for $y = \mathcal{P}^t(x)$ for any $t = \Omega(\log^3 n)$, it holds that*

$$\Pr\left[\bigwedge_{i \in S}\left(y^i \in \mathsf{Special} \cup \mathsf{HighDist}_n^k\right) \wedge \bigwedge_{i \notin S}\left(y^i \in \mathsf{HighDist}_n^k\right)\right] \geq 1 - \mathsf{negl}(n).$$

Moreover, if $y^i \in \mathsf{Special}$, then y^i is a deterministic function of x^i with high probability.

Proof. We will prove the statement above for each pair (x^i, y^i) separately. The final statement will follow from a union bound over all $k = O(1)$ input-output pairs.

First, if x^i is a special string, it evolves in a deterministic way until it passes into the second phase. If it goes into the second phase, its distribution becomes negligibly far from a uniformly random string from Lemma 4.9.[3] Concretely, it will have min-entropy at least $n \log |\Sigma| - O(1)$. Since the rounds of our

[3]Since only a negligible fraction of strings cannot be sampled as w^x during the second phase of the $\mathcal{W}$ step.

block cipher are permutations, the marginal distribution of y^i remains almost uniformly random for the subsequent rounds, and thus Corollary 6.3 implies that y^i is a high-distance string with high probability. If a special string does not go into the second phase, it remains a special string with certainty, and it is a deterministic function of the original input x^i.

Second, if x^i is not a special string, we show in Lemma 7.13 of Section 7 below that after $O(\log^3 n)$ steps y^i will have min-entropy at least $n \log |\Sigma| - O(1)$. This again implies from Corollary 6.3 that y^i is a high-distance string with high probability. □

In light of this theorem, it suffices to restrict our attention to the case when our input $x = (x^1, \ldots, x^k)$ to the block cipher $\mathcal{P}^{n-O(\log^3 n)}$ consists only of high-distance and special strings. Moreover, it suffices to show that the outputs of $\mathcal{P}^{n-O(\log^3 n)}$ are negligibly far from the uniform distribution over $\mathsf{HighDist}_n^k$. Indeed, since this space captures all but a negligible fraction of $(\Sigma^n)^{(k)}$ from Lemma 6.2, proving approximate k-wise independence in this space suffices.

6.1 The spectral gap of the conditional brickwork step

The way we prove approximate k-wise independence in this space is by keeping track of the 2-norm of the current distribution of our input. Since we started with special and high-distance strings, we will show that after each conditional brickwork step, the distribution over high-distance strings will get more and more uniform. On the other hand, the special strings behave deterministically until they enter phase two, when they become almost uniformly random themselves. To keep track of the 2-norm of the high-distance strings, we will establish a spectral gap bound for our conditional brickwork layer when restricted to high-distance strings.

We start by defining the restricted versions of our Markov chains for high-distance k-tuples.

Definition 6.5 (High-distance random reversible brickwork circuits)**.** *The chain $\widetilde{\mathcal{R}}_n^k$ on the state space of high-distance k-tuples has a transition matrix, such that for any $x, y \in \mathsf{HighDist}_n^k$:*

$$\widetilde{\mathcal{R}}_n^k(x,y) = \begin{cases} \mathcal{R}_n^k(x,y) & x \neq y, \\ \mathcal{R}_n^k(x,y) + \sum_{z \notin \mathsf{HighDist}_n^k} \mathcal{R}_n^k(x,z) & x = y. \end{cases}$$

In other words, the high-distance version of the chain performs the same transition as $\mathcal{R}_n^k$, but if it transitions to a low-distance state, then $\widetilde{\mathcal{R}}_n^k$ remains in the same state.

We observe that the probability of transitioning from a high-distance to a low-distance tuple is small when k is small, and thus the transition probabilities of the high-distance chain almost agree with the original chain.

Lemma 6.6. *For any $x \in \mathsf{HighDist}_n^k$, it holds that $\sum_{z \notin \mathsf{HighDist}_n^k} \mathcal{R}_n^k(x,z) = \mathrm{negl}(n)$.*

Proof. We will show this by restricting our attention to each input x^i separately and applying a union bound over all $k = O(1)$ inputs.

Observe that each input x^i that goes through a brickwork layer is mapped to a string z^i whose marginal distribution has high min-entropy. In particular, since a brickwork layer randomizes $n - O(1)$ of the symbols, the min-entropy of z^i is at least $(n - O(1)) \log |\Sigma|$. Then Corollary 6.3 shows that z^i is high-distance with high probability. By the union bound over all k inputs, we conclude that $z = (z^1, \ldots, z^k)$ consists only of high-distance strings with high probability. □

Lemma 6.7. *It holds that $\lambda(\widetilde{\mathcal{R}}_n^k) \geq \lambda(\mathcal{R}_n^k) - \mathrm{negl}(n)$.*

Proof. For simplicity, we will drop the n, k in the superscript and subscript throughout this proof. Moreover, let us define $\psi_x = \sum_{z \notin \mathsf{HighDist}_n^k} \mathcal{R}(x,z)$, $\delta_{xy} = 1$ if and only if $x = y$, and 0 otherwise, and we use $\widehat{\mathcal{R}}$ to denote the restriction of $\mathcal{R}$ to the high-distance k-tuples (meaning that $\widehat{\mathcal{R}}$ is not necessarily a doubly stochastic matrix).

We first bound the difference between the second eigenvalues of $\mathcal{R}^T\mathcal{R}$ and $\widehat{\mathcal{R}}^T\widehat{\mathcal{R}}$ using the Weyl inequality. Let us first view $\widehat{\mathcal{R}}$ as a matrix on the entirety of $(\Sigma^n)^{(k)}$ by padding it with zeros. Then

$$|\lambda_2(\mathcal{R}^T\mathcal{R}) - \lambda_2(\widehat{\mathcal{R}}^T\widehat{\mathcal{R}})| \leq \|\mathcal{R}^T\mathcal{R} - \widehat{\mathcal{R}}^T\widehat{\mathcal{R}}\|_\infty.$$

By writing $\mathcal{R} = \widehat{\mathcal{R}} + E$, we observe that

$$\begin{aligned}\|\mathcal{R}^T\mathcal{R} - \widehat{\mathcal{R}}^T\widehat{\mathcal{R}}\|_\infty &= \|(\widehat{\mathcal{R}} + E)^T(\widehat{\mathcal{R}} + E) - \widehat{\mathcal{R}}^T\widehat{\mathcal{R}}\|_\infty \\ &= \|\widehat{\mathcal{R}}^T E + E^T\widehat{\mathcal{R}} + E^T E\|_\infty \\ &\leq 2\|\widehat{\mathcal{R}}\|_\infty \cdot \|E\|_\infty + \|E\|_\infty^2,\end{aligned}$$

where we used the fact that $\mathcal{R}, \widehat{\mathcal{R}}$, and thus E, are symmetric matrices and the operator norm is sub-additive and sub-multiplicative. Recall Remark 3.12, which states that the operator norm of a symmetric matrix with positive entries is bounded above by the maximum row and column sum. From Lemma 6.6, this means that $\|E\|_\infty = \text{negl}(n)$, and $\|\widehat{\mathcal{R}}\|_\infty \leq 1$, since it is a restriction of a doubly stochastic matrix. We conclude that the second eigenvalues of these matrices are negligibly far:

$$|\lambda_2(\mathcal{R}^T\mathcal{R}) - \lambda_2(\widehat{\mathcal{R}}^T\widehat{\mathcal{R}})| \leq \text{negl}(n).$$

The definition of $\widetilde{\mathcal{R}}$ can be written compactly as

$$\widetilde{\mathcal{R}}(x, y) = \widehat{\mathcal{R}}(x, y) + \delta_{xy}\psi_x.$$

Hence

$$\begin{aligned}(\widetilde{\mathcal{R}}^T\widetilde{\mathcal{R}})(x, y) &= \sum_{z \in \mathsf{HighDist}_n^k} \widetilde{\mathcal{R}}(z, x)\widetilde{\mathcal{R}}(z, y) \\ &= \sum_{z \in \mathsf{HighDist}_n^k} (\widehat{\mathcal{R}}(z, x) + \delta_{xz}\psi_z)(\widehat{\mathcal{R}}(z, y) + \delta_{yz}\psi_z) \\ &= \sum_{z \in \mathsf{HighDist}_n^k} \widehat{\mathcal{R}}(z, x)\widehat{\mathcal{R}}(z, y) + \psi_x \cdot \widehat{\mathcal{R}}(x, y) + \psi_y \cdot \widehat{\mathcal{R}}(y, x) + \delta_{xy}\psi_x\psi_y \\ &= (\widehat{\mathcal{R}}^T\widehat{\mathcal{R}})(x, y) + \psi_x \cdot \widehat{\mathcal{R}}(x, y) + \psi_y \cdot \widehat{\mathcal{R}}(y, x) + \delta_{xy}\psi_x\psi_y. \qquad (1)\end{aligned}$$

where Ψ is the diagonal matrix on $\mathsf{HighDist}_n^k$ such that $\Psi(x, x) = \psi_x$. Using the fact that $\widehat{\mathcal{R}}$ and Ψ are symmetric matrices, Equation (1) can be written as

$$\widetilde{\mathcal{R}}^T\widetilde{\mathcal{R}} = \widehat{\mathcal{R}}^T\widehat{\mathcal{R}} + \widehat{\mathcal{R}} \cdot \Psi + \Psi \cdot \widehat{\mathcal{R}} + \Psi^2.$$

We are now ready to bound the second eigenvalue of $\widetilde{\mathcal{R}}^T\widetilde{\mathcal{R}}$. From the Weyl inequality in Theorem 3.13:

$$\begin{aligned}|\lambda_2(\widetilde{\mathcal{R}}^T\widetilde{\mathcal{R}}) - \lambda_2(\widehat{\mathcal{R}}^T\widehat{\mathcal{R}})| &\leq \|\widehat{\mathcal{R}} \cdot \Psi + \Psi \cdot \widehat{\mathcal{R}} + \Psi^2\|_\infty \\ &\leq \|\widehat{\mathcal{R}} \cdot \Psi\|_\infty + \|\Psi \cdot \widehat{\mathcal{R}}\|_\infty + \|\Psi^2\|_\infty \\ &\leq 2\|\widehat{\mathcal{R}}\|_\infty \cdot \|\Psi\|_\infty + \|\Psi\|_\infty^2.\end{aligned}$$

Since Ψ is a diagonal matrix, $\|\Psi\|_\infty$ is equal to its maximum absolute eigenvalue, which is $\max_x \psi_x = \text{negl}(n)$ from Lemma 6.6. We conclude that the second eigenvalues of the two matrices are negligibly far, and thus the spectral gap of the high-distance brickwork circuit is at most negligibly smaller than the spectral gap of the original brickwork circuit. □

The conditional brickwork step $\mathcal{B}$ of our block cipher naturally gives rise to the following Markov chain.

Definition 6.8 (Conditional brickwork circuit)**.** *Let n be an integer such that $n \equiv 1 \pmod 3$, and let $n' = (n-1)/3$. The chain $\mathcal{B}_n^k$ on the state space $(\Sigma^n)^{(k)}$ has a transition matrix, such that for any $x = (x^1, \ldots, x^k) \in (\Sigma^n)^{(k)}$:*

- *Choose a random index i^* from $[n-1]$ and let $i^* \equiv j \pmod 3$. Set $i = 1$ if $j = 1, 2$ and $i = 2$ otherwise.*
- *Sample random permutations $\pi_0, \ldots, \pi_{n'-1}$.*
- *For every string x^r, if i^* is* detectable*, that is $x^r_{i^*} x^r_{i^*+1}$ is different from a valid counter, then:*
  - *If $i = 1$, then $y^r_n = x^r_n$.*
  - *Else if $i = 2$, then $y^r_1 = x^r_1$.*
  - *For all $c \in \{0, \ldots, n'-1\}$*
    - *If $i^*, i^*+1 \notin \{3c+i, 3c+i+1, 3c+i+2\}$, set*
    $$(y^r_{3c+i}, y^r_{3c+i+1}, y^r_{3c+i+2}) = \pi_c(x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}).$$
    - *Else set*
    $$(y^r_{3c+i}, y^r_{3c+i+1}, y^r_{3c+i+2}) = (x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}).$$
- *The new state is $y = (y^1, \ldots, y^k)$.*

We will refer to the index c for which $\{i^*, i^*+1\} \subseteq \{3c+i, 3c+i+1, 3c+i+2\}$ as the skipped index*. Each edge is determined by the indices i^*, i, c and the permutations π. We will thus refer to it using the tuple $(i^*, i, c, \pi_0, \ldots, \pi_{n'-1})$.*

Intuitively, this chain chooses a random pair of consecutive wires i^*, i^*+1 and if it detects that they are different than a valid counter, it applies a brickwork layer to all wires except the triple that includes i^*, i^*+1. For our purposes, we consider the high-distance restriction of this chain.

Definition 6.9 (High-distance conditional brickwork circuit)**.** *The chain $\widetilde{\mathcal{B}}_n^k$ on the state space of high-distance k-tuples has a transition matrix, such that for any $x, y \in \mathsf{HighDist}_n^k$:*

$$\widetilde{\mathcal{B}}_n^k(x,y) = \begin{cases} \mathcal{B}_n^k(x,y) & x \neq y, \\ \mathcal{B}_n^k(x,y) + \sum_{z \notin \mathsf{HighDist}_n^k} \mathcal{B}_n^k(x,z) & x = y. \end{cases}$$

Using a comparison argument, we can show that the spectral gap $\widetilde{\mathcal{B}}_n^k$ is only a constant factor away from the spectral gap of $\widetilde{\mathcal{R}}_n^k$.

Lemma 6.10. *The spectral gap of $\widetilde{\mathcal{B}}_n^k$ and $\widetilde{\mathcal{R}}_n^k$ satisfy:*

$$\lambda(\widetilde{\mathcal{B}}_n^k) \geq \Theta(1) \cdot \lambda(\widetilde{\mathcal{R}}_n^k).$$

Proof. We prove this using a comparison by multicommodity flows argument. For every edge (x, y) in $\widetilde{\mathcal{R}}_n^k$, we will construct a set of paths $\mathcal{P}_{xy}$ that consist of two edges in $\widetilde{\mathcal{B}}_n^k$. First, from the transitions of $\widetilde{\mathcal{R}}_n^k$, each edge (x, y) is determined by an offset value $i \in \{1, 2\}$ and n' random permutations $\pi_0, \ldots, \pi_{n'-1}$.

The paths that we construct will consist of the pair of edges (x, z) and (z, y) from $\widetilde{\mathcal{B}}_n^k$, for some intermediate node z. In particular, the first edge is determined by

$$(i_0^*, i_0, c_0, \sigma_0, \ldots, \sigma_{n'-1}),$$

where $i_0 = i$ and $x^r_{i_0^*} x^r_{i_0^*+1}$ is different from a valid counter for all $r \in [k]$. Let this edge take the node $x = (x^1, \ldots, x^k)$ to the node $z = (z^1, \ldots, z^k)$. The second edge is determined by

$$(i_1^*, i_1, c_1, \tau_0, \ldots, \tau_{n'-1}),$$

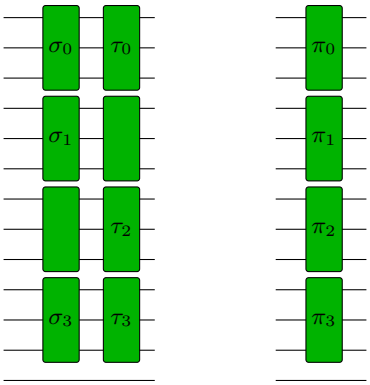

Figure 7: An example of the length-2 paths that we construct. On the right-hand side, we have an edge in $\widetilde{\mathcal{R}}$ that is defined by a layer of brickwork gates. On the left-hand side, we use two edges of the conditional brickwork $\widetilde{\mathcal{B}}$. In particular, each one of these layers must have one gate that is trivial and corresponds to the gate that touches the wires that we are conditioning on. Empty rectangles represent these. Then we can see that as long as $\tau_0 \circ \sigma_0 = \pi_0$, $\sigma_1 = \pi_1$, and so on, the two figures compute the same permutation.

where $i_1 = i$, the skipped indices c_0, c_1 are distinct, i_0^*, i_1^* have the same remainder modulo 3, and $z^r_{i_1^*} z^r_{i_1^*+1}$ is different from a valid counter for all $r \in [k]$. Moreover, the permutations sampled satisfy:

$$\sigma_{c_1} = \pi_{c_1}, \qquad \tau_{c_0} = \pi_{c_0}, \qquad \tau_c \circ \sigma_c = \pi_c, \ \forall c \notin \{c_0, c_1\}.$$

We first show that these edges define a valid path from $x = (x^1, \ldots, x^k)$ to $y = (y^1, \ldots, y^k)$. From the description of the first edge, we can write z^r as:

- If $i = 1$, then $z^r_n = x^r_n$. Else if $i = 2$, then $z^r_1 = x^r_1$.
- For all $c \in \{0, \ldots, n' - 1\}$:
  - If $c \neq c_0$, set:
$$(z^r_{3c+i}, z^r_{3c+i+1}, z^r_{3c+i+2}) = \sigma_c(x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}).$$
  - Else if $c = c_0$, set:
$$(z^r_{3c+i}, z^r_{3c+i+1}, z^r_{3c+i+2}) = (x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}).$$

Let now the second edge take $z = (z^1, \ldots, z^k)$ to a node $w = (w^1, \ldots, w^k)$. We can write w^r as:

- If $i = 1$, then $w^r_n = z^r_n = x^r_n = y^r_n$. Else if $i = 2$, then $w^r_1 = z^r_1 = x^r_1 = y^r_1$.
- For all $c \in \{0, \ldots, n' - 1\}$:
  - If $c \neq c_1$, set:
$$\begin{aligned}(w^r_{3c+i}, w^r_{3c+i+1}, w^r_{3c+i+2}) &= \tau_c(z^r_{3c+i}, z^r_{3c+i+1}, z^r_{3c+i+2}) \\ &= \pi_c(x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}) \\ &= (y^r_{3c+i}, y^r_{3c+i+1}, y^r_{3c+i+2})\end{aligned}$$
  - Else if $c = c_1$, set:
$$\begin{aligned}(w^r_{3c+i}, w^r_{3c+i+1}, w^r_{3c+i+2}) &= (z^r_{3c+i}, z^r_{3c+i+1}, z^r_{3c+i+2}) \\ &= \sigma_c(x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}) \\ &= \pi_c(x^r_{3c+i}, x^r_{3c+i+1}, x^r_{3c+i+2}) \\ &= (y^r_{3c+i}, y^r_{3c+i+1}, y^r_{3c+i+2}).\end{aligned}$$

Hence, these edges form a path from x to y.

Let us now show that this flow has a small comparison constant. In particular, from Definition 3.21, the total flow that has to go through $\mathcal{P}_{xy}$ is equal to

$$\widetilde{\mathcal{R}}_n^k(x,y) \cdot \frac{1}{|\mathsf{HighDist}_n^k|} = \frac{\Theta(1)}{((|\Sigma|^3)!)^{n/3}} \cdot \frac{1}{|\mathsf{HighDist}_n^k|},$$

where the factor $((|\Sigma|^3)!)^{n/3}$ is from our choice of $\pi_0, \dots, \pi_{n'}$, the fact that we have a constant-size alphabet, and that $n' = \Theta(n/3)$. We will divide the total amount of flow uniformly over all the paths in $\mathcal{P}_{xy}$. The total number of such paths is

$$\Theta(n^2) \cdot ((|\Sigma|^3)!)^{n/3},$$

where the second factor appears because we can choose all but one of the σ permutations uniformly at random. On the other hand, the τ permutations are fully determined by the π and σ permutations. The n^2 factor appears because we are dealing with high-distance strings, and thus for each node $x = (x^1, \dots, x^k) \in \mathsf{HighDist}_n^k$, there are at least $n(1 - 3k/|\Sigma|) = \Theta(n)$ indices i^* for which $x^r_{i^*} x^r_{i^*+1}$ does not look like a valid counter for all $r \in [k]$. Hence, the number of i^*'s that give possible edges for our paths is $\Theta(n)$ for all starting vertices. Thus, we obtain a factor of $\Theta(n)$ from both the starting vertex x and the intermediate vertex z. We conclude that the total amount of flow that passes through each path γ is equal to

$$F(\gamma) = \frac{1}{\Theta(n^2) \cdot ((|\Sigma|^3)!)^{2n/3} \cdot |\mathsf{HighDist}_n^k|}.$$

We now bound the number of paths γ that pass through an edge (a,b) in $\widetilde{\mathcal{B}}_n^k$. Consider a path $\gamma \in \mathcal{P}_{xy}$ that uses the edge $(a,b) = (i_0^*, \sigma_0, \dots, \sigma_{n'-1})$ as its first edge, and its second edge is $(i_1^*, \tau_0, \dots, \tau_{n'-1})$. We will count the number of valid second edges. The permutations $\pi_0, \dots, \pi_{n'-1}$ of (x,y) and $\sigma_0, \dots, \sigma_{n'-1}$ fully determine the permutations $\tau_0, \dots, \tau_{n'-1}$ of the second edge. Hence, for every edge (x,y), there exist $\Theta(n)$ valid second edges, where the degrees of freedom appear from the index i_1^*. The same happens if (a,b) is the second edge of γ, and thus the number of paths is at most doubled. Thus, the total number of paths that pass through an edge (a,b) is at most

$$\Theta(n) \cdot ((|\Sigma|^3)!)^{n/3}.$$

We proceed to bound the comparison constant of this flow.

$$\begin{aligned} A(F) &= \max_{(a,b) \in \widetilde{\mathcal{B}}_n^k} \left\{ |\mathsf{HighDist}_n^k| \cdot \Theta(n) \cdot ((|\Sigma|^3)!)^{n/3} \sum_{\gamma \in \mathcal{P}(a,b)} 2 \cdot \frac{1}{\Theta(n^2) \cdot ((|\Sigma|^3)!)^{2n/3} \cdot |\mathsf{HighDist}_n^k|} \right\} \\ &\leq |\mathsf{HighDist}_n^k| \cdot \Theta(n) \cdot ((|\Sigma|^3)!)^{n/3} \cdot \Theta(n) \cdot ((|\Sigma|^3)!)^{n/3} \cdot \frac{1}{\Theta(n^2) \cdot ((|\Sigma|^3)!)^{2n/3} \cdot |\mathsf{HighDist}_n^k|} \\ &= \Theta(1). \end{aligned}$$

The statement follows from Theorem 3.22. □

Corollary 6.11. *The spectral gap of the high-distance conditional brickwork circuit chain for constant k is at least a constant:*

$$\lambda(\widetilde{\mathcal{B}}_n^k) = \Theta(1).$$

Proof. We have just proved in Lemma 6.10 how the spectral gaps of the conditional brickwork and the brickwork over high-distance inputs are related:

$$\lambda(\widetilde{\mathcal{B}}_n^k) \geq \Theta(\lambda(\widetilde{\mathcal{R}}_n^k)).$$

Moreover, in Lemma 6.7 we show that the high-distance version of the brickwork chain has approximately the same spectral gap as the original brickwork chain, meaning:

$$\lambda(\widetilde{\mathcal{B}}_n^k) \geq \Theta(\lambda(\mathcal{R}_n^k)) - \mathrm{negl}(n).$$

Finally, Theorem 3.19 bounds the spectral gap of the brickwork Markov chain to be at least a constant. We conclude that

$$\lambda(\widetilde{\mathcal{B}}_n^k) = \Theta(\lambda(\mathcal{R}_n^k) = \Theta(1). \qquad \square$$

6.2 Proof of k-wise independence

Theorem 6.12. *For any k-tuple of distinct inputs $(x^1, \ldots, x^k) \in (\Sigma^n)^{(k)}$, the distribution of the output tuple*

$$(y^1, \ldots, y^k) = (\mathcal{P}^{n-2}(x^1), \ldots, \mathcal{P}^{n-2}(x^k))$$

is ε-far from the uniform distribution over $(\Sigma^n)^{(k)}$, where $\varepsilon = \mathrm{negl}(n)$.

Proof. According to Lemma 6.4, after the first $n_0 = O(\log^3 n)$ rounds of the block cipher, all but negligible probability mass is concentrated on k-tuples with special and high-distance strings.

It suffices thus to show that for any input $x = (x^1, \ldots, x^k)$ that consists of special and high-distance strings, $n_1 := n - 2 - n_0$ rounds suffice to reach approximate k-wise independence among k-tuples of high-distance strings. For the remainder of this proof, we will condition on the event that all of our inputs are either high-distance or special. Since the opposite happens with negligible probability in every round, applying the union bound inequality over all $k = \Theta(1)$ inputs and $O(n)$ rounds will only incur an additional negligible distance in n.

Formally, let us define $p_0, p_1, \ldots, p_{n_1}$ to be the probability distribution over k-tuples of distinct strings during the core rounds of the block cipher.

$$p_t = (\mathcal{P}^{n_0+t}(x^1), \ldots, \mathcal{P}^{n_0+t}(x^k))$$

Moreover, we use $\tilde{p}_0, \tilde{p}_1, \ldots, \tilde{p}_{n_1}$ to denote the probability distribution of the intermediate core rounds, when restricted to the tuples that are all either special or high-distance for all core rounds so far. Since only a negligible amount of probability is not in special or high-distance, we will use that

$$\|p_t - \tilde{p}_t\|_1 \leq \mathrm{negl}(n).$$

During each round t, there exists a subset of $[k]$ such that each k-tuple in the support of $\tilde{p}_t$ will contain high-distance strings in this subset, and special strings undergoing the deterministic counter steps in its complement. We will show that the conditional brickwork step $\mathcal{B}$ randomizes the high-distance subset of the inputs, whereas each special input will become almost truly random when it enters phase two in $\mathcal{W}$.

For each of the core rounds, let us define $S_0, \ldots, S_{n_1} \subseteq [k]$ to be the subset of the inputs that is special, where S_0 corresponds to the subset of the inputs that starts the core rounds in the special set.

During round t for $t \in [n_1]$:

$$p_t = \mathcal{W} \circ \mathcal{C} \circ \mathcal{B}^m \circ \mathcal{M} \cdot p_{t-1}.$$

We instead study

$$\tilde{p}_t = \widetilde{\mathcal{W}} \circ \widetilde{\mathcal{C}} \circ \widetilde{\mathcal{B}}^m \circ \widetilde{\mathcal{M}} \cdot \tilde{p}_{t-1},$$

where the rounds with the tilde correspond to a random process that follows the original step in the block cipher, but terminates if it maps a high-distance string to a non-high-distance (low-distance or special) string. Formally, the transition matrices that correspond to these procedures are called *sub-stochastic*, since they allow some of the probability mass to leave the system. In particular, we will not count the probability of being terminated in $\tilde{p}_t$.

Let us emphasize here that we are abusing notation: The Markov chain $\widetilde{\mathcal{B}}_n^k$ that we have defined in Definition 6.9 has additional probability on the self-loops to become stochastic. Fortunately, it is straightforward to argue that the same contraction property (Theorem 3.17) that holds for the normalized chain also holds for the sub-stochastic version. This is because one can apply the stochastic version and then remove the probability mass that corresponds to the self-loops. This can only further decrease the 2-norm.

We define the following intermediate probability distributions

$$\tilde{p}^{(a)}_{t-1} = \widetilde{\mathcal{M}} \cdot p_{t-1}, \qquad \tilde{p}^{(b)}_{t-1} = \widetilde{\mathcal{B}}^m \cdot \tilde{p}^{(a)}_{t-1}, \qquad \tilde{p}^{(c)}_{t-1} = \widetilde{\mathcal{C}} \cdot \tilde{p}^{(b)}_{t-1}.$$

It holds that applying a randomized permutation does not decrease the 2-norm of a distribution. This is still true for sub-stochastic matrices, and in particular for $\widetilde{\mathcal{M}}$:

$$\|\tilde{p}^{(a)}_{t-1}\|_2^2 \leq \|\tilde{p}_{t-1}\|_2^2. \tag{2}$$

Moreover, Theorem 3.17 implies that we can bound $\|\tilde{p}^{(b)}_{t-1}\|_2^2$ using the spectral gap of $\widetilde{\mathcal{B}}$. We note here that the conditional brickwork Markov chain is only applied to the $k_{t-1} = k - |S_{t-1}|$ inputs of the k-tuple that are currently not special. The remaining strings are determined from the input tuple $(x^1, \ldots, x^k)$. Fortunately, the spectral gaps satisfy

$$\lambda(\widetilde{\mathcal{B}}^k_n) \leq \lambda(\widetilde{\mathcal{B}}^{k'}_n) \tag{3}$$

for any $k' \leq k$, since restricting our attention to fewer inputs constitutes a projection of the Markov chain. Hence, the 2-norm of the current distribution is equal to the 2-norm of the distribution when restricted to the high-distance inputs, which live in $\mathsf{HighDist}_n^{k_{t-1}}$:

$$\|\tilde{p}^{(b)}_{t-1}\|_2^2 = \|\tilde{p}^{(b)}_{t-1} \mid \overline{S_{t-1}}\|_2^2.$$

Now the Markov chain $\widetilde{\mathcal{B}}^{k_{t-1}}_n$ acts on the space $\mathsf{HighDist}_n^{k_{t-1}}$ and its spectral gap implies that

$$\begin{aligned}
\|\tilde{p}^{(b)}_{t-1} \mid \overline{S_{t-1}}\|_2^2 &\leq (1 - \lambda(\widetilde{\mathcal{B}}^{k_{t-1}}_n))^m \cdot \|\tilde{p}^{(b)}_{t-1} \mid \overline{S_{t-1}}\|_2^2 + [1 - (1 - \lambda(\widetilde{\mathcal{B}}^{k_{t-1}}_n))^m] \cdot \frac{1}{|\mathsf{HighDist}_n^{k_{t-1}}|} \\
&\leq (1 - \lambda(\widetilde{\mathcal{B}}^{k}_n))^m \cdot \|\tilde{p}^{(b)}_{t-1} \mid \overline{S_{t-1}}\|_2^2 + [1 - (1 - \lambda(\widetilde{\mathcal{B}}^{k}_n))^m] \cdot \frac{1}{|\mathsf{HighDist}_n^{k_{t-1}}|}. \qquad \text{(Equation (3))}
\end{aligned}$$

Let us set $\rho = (1 - \lambda(\widetilde{\mathcal{B}}^k_n))^m$ and we will choose m later to be a large enough constant such that ρ becomes small enough and allows us to converge to k-wise independence in n_1 rounds. The last inequality can be rewritten as:

$$\|\tilde{p}^{(b)}_{t-1}\|_2^2 \leq \rho \cdot \|p_{t-1}\|_2^2 + (1 - \rho) \cdot \frac{1}{|\mathsf{HighDist}_n^{k_{t-1}}|}. \tag{4}$$

We again use the fact that the counter step cannot decrease our 2-norm:

$$\|\tilde{p}^{(c)}_{t-1}\|_2^2 \leq \|\tilde{p}^{(b)}_{t-1}\|_2^2. \tag{5}$$

Finally, the step $\mathcal{W}$ will cause $k_t - k_{t-1}$ inputs to become high-distance. Since these inputs were deterministic to start with, and we are mapping them to a random tuple that is negligibly far from a $(k_t - k_{t-1})$-wise independent one (Lemma 4.9), the 2-norm of the distribution decreases by a factor of at least $1/|\mathsf{HighDist}_n^{k_t - k_{t-1}}|$. For simplicity, let us use $u^1, \ldots, u^k$ to denote the current strings. Without loss of generality, let $u = (u^1, \ldots, u^{k_{t-1}})$ be the strings that are high-distance before $\widetilde{\mathcal{W}}$, and $u^{k_{t-1}+1}, \ldots, u^{k_t}$ be the strings that become high-distance right after $\widetilde{\mathcal{W}}$. Then

$$\begin{aligned}
\|\tilde{p}^{(c)}_{t-1}\|_2^2 &= \sum_{v=(v^1,\ldots,v^k)} \Pr\Big[\bigwedge_{i\in[k]} u^i = v^i\Big]^2 \\
&= \sum_{v} \Pr\Big[\bigwedge_{i\in[k_{t-1}]} u^i = v^i \mid \bigwedge_{i>k_{t-1}} u^i = v^i\Big]^2 \cdot \Pr\Big[\bigwedge_{k_{t-1}<i\leq k_t} u^i = v^i \mid \bigwedge_{i>k_t} u^i = v^i\Big]^2 \cdot \Pr\Big[\bigwedge_{i>k_t} u^i = v^i\Big]^2 \\
&= \sum_{v^1,\ldots,v^{k_{t-1}}} \Pr\Big[\bigwedge_{i\in[k_{t-1}]} u^i = v^i\Big]^2.
\end{aligned}$$

Applying the $\widetilde{\mathcal{W}}$ step will now change the above expression as follows:

$$\begin{aligned}
&\|\tilde{p}_t\|_2^2 = \|\widetilde{\mathcal{W}} \cdot \tilde{p}_{t-1}^{(c)}\|_2^2 \\
&= \sum_{v=(v^1,\ldots,v^k)} \Pr\Big[\bigwedge_{i\in[k]} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2 \\
&= \sum_{v} \Pr\Big[\bigwedge_{i\in[k_{t-1}]} \widetilde{\mathcal{W}}(u^i) = v^i \mid \bigwedge_{i>k_{t-1}} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2 \cdot \Pr\Big[\bigwedge_{k_{t-1}<i\le k_t} \widetilde{\mathcal{W}}(u^i) = v^i \mid \bigwedge_{i>k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2 \cdot \Pr\Big[\bigwedge_{i>k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2 \\
&= \sum_{v^{k_{t-1}+1},\ldots,v^{k_t}} \Pr\Big[\bigwedge_{k_{t-1}<i\le k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2 \sum_{v^1,\ldots,v^{k_{t-1}}} \Pr\Big[\bigwedge_{i\in[k_{t-1}]} \widetilde{\mathcal{W}}(u^i) = v^i \mid \bigwedge_{k_{t-1}<i\le k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2.
\end{aligned}$$

The final equalities for the two expressions above hold because the u^i are deterministic for $i > k_{t-1}$ and the $\widetilde{\mathcal{W}}(u^i)$ are deterministic for $i > k_t$. We now claim that for every $v^{k_{t-1}+1}, \ldots, v^{k_t}$ such that $\Pr\Big[\bigwedge_{k_{t-1}<i\le k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]$ is non-zero (i.e. the conditional probability below is defined), it holds that:

$$\sum_{v^1,\ldots,v^{k_{t-1}}} \Pr\Big[\bigwedge_{i\in[k_{t-1}]} \widetilde{\mathcal{W}}(u^i) = v^i \mid \bigwedge_{k_{t-1}<i\le k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2 \le \sum_{v^1,\ldots,v^{k_{t-1}}} \Pr\Big[\bigwedge_{i\in[k_{t-1}]} u^i = v^i\Big]^2.$$

This is because both expressions above correspond to the 2-norm of probability distributions. The condition on the left one may restrict the random coins of $\widetilde{\mathcal{W}}$, but it remains a permutation, and thus the 2-norm of a distribution cannot increase after applying a permutation.

Hence

$$\|\tilde{p}_t\|_2^2 = \|\widetilde{\mathcal{W}}\tilde{p}_{t-1}^{(c)}\|_2^2 \le \|\tilde{p}_{t-1}^{(c)}\|_2^2 \sum_{v^{k_{t-1}+1},\ldots,v^{k_t}} \Pr\Big[\bigwedge_{k_{t-1}<i\le k_t} \widetilde{\mathcal{W}}(u^i) = v^i\Big]^2.$$

On the other hand, $\widetilde{\mathcal{W}}$ takes special strings through phase two to almost a uniformly random $(k_t - k_{t-1})$-tuple of string of high-distance (the only difference is that these strings must have an index in which they all differ, which can only happen with $\mathrm{negl}(n)$ probability). Hence, the summation above is at most $(1 + \mathrm{negl}(n))/|\mathsf{HighDist}_n^{k_t - k_{t-1}}|$.

As a result,

$$\begin{aligned}
\|\tilde{p}_t\|_2^2 &\le \|\tilde{p}_{t-1}^{(c)}\|_2^2 \cdot \frac{1+\mathrm{negl}(n)}{|\mathsf{HighDist}_n^{k_t-k_{t-1}}|} \\
&\le \|\tilde{p}_{t-1}^{(c)}\|_2^2 \cdot \frac{(1+\mathrm{negl}(n))|\mathsf{HighDist}_n^{k_{t-1}}|}{|\mathsf{HighDist}_n^{k_t}|} \\
&\le \rho \cdot \|\tilde{p}_{t-1}\|_2^2 \cdot \frac{(1+\mathrm{negl}(n))|\mathsf{HighDist}_n^{k_{t-1}}|}{|\mathsf{HighDist}_n^{k_t}|} + (1-\rho)\cdot\frac{1+\mathrm{negl}(n)}{|\mathsf{HighDist}_n^{k_t}|}. \qquad \text{(Equations (4) and (5))}
\end{aligned}$$

Telescoping the above, we get that

$$\begin{aligned}
\|\tilde{p}_{n_1}\|_2^2 &\le \rho^{n_1} \cdot \|\tilde{p}_0\|_2^2 \cdot \frac{(1+\mathrm{negl}(n))|\mathsf{HighDist}_n^{k_0}|}{|\mathsf{HighDist}_n^{k}|} + (1-\rho^{n_1})\cdot\frac{1+\mathrm{negl}(n)}{|\mathsf{HighDist}_n^{k}|} \\
&\le \rho^{n_1}\cdot(1+\mathrm{negl}(n)) + (1-\rho^{n_1})\cdot\frac{1+\mathrm{negl}(n)}{|\mathsf{HighDist}_n^{k}|}.
\end{aligned}$$

To compute the statistical distance of p_{n_1} from the uniform distribution μ on $\mathsf{HighDist}_n^k$, we use the Cauchy-

Schwarz inequality:

$$\begin{aligned}\|\tilde{p}_{n_1} - \mu\|_1 &\leq \sqrt{|\mathsf{HighDist}_n^k|} \cdot \|\tilde{p}_{n_1} - \mu\|_2 \\ &\leq \sqrt{|\mathsf{HighDist}_n^k|} \cdot \sqrt{\|\tilde{p}_{n_1}\|_2^2 - \frac{1}{|\mathsf{HighDist}_n^k|}} \\ &\leq \sqrt{(1+\mathrm{negl}(n)) \cdot \rho^{n_1} \cdot |\mathsf{HighDist}_n^k| + \mathrm{negl}(n)}.\end{aligned}$$

It holds that $|\mathsf{HighDist}_n^k| = O(2^{n \cdot k \log|\Sigma|})$. Thus, let us choose m to be a constant large enough such that $\rho = 2^{-d}$, where $d > 2 \cdot k \log|\Sigma|$. Since the spectral gap of the conditional brickwork is at least a constant, this is always possible. Then

$$\begin{aligned}\|p_{n_1} - \mu\|_1 &\leq \sqrt{(1+\mathrm{negl}(n)) \cdot \rho^{n_1} \cdot |\mathsf{HighDist}_n^k| + \mathrm{negl}(n)} \\ &\leq \sqrt{(1+\mathrm{negl}(n)) \cdot \rho^{n/2} \cdot |\mathsf{HighDist}_n^k| + \mathrm{negl}(n)} \\ &\leq \sqrt{(1+\mathrm{negl}(n)) \cdot 2^{n \cdot (k \log|\Sigma| - d/2)} + \mathrm{negl}(n)} = \mathrm{negl}(n).\end{aligned}$$

This concludes the k-wise independence of our block cipher. □

7 Non-special strings reach high min-entropy quickly

In this section, we prove that any non-special string achieves almost maximal min-entropy after $O(\log^3 n)$ rounds. The main result of this subsection is in Lemma 7.13.

This section is organized as follows:

- In Lemma 7.2 we analyze the change in the distance D when a single wire is randomized.
- We use Lemma 7.2 to lower bound the growth of D when applying $\mathcal{M}$ in Propositions 7.3 to 7.5.
- In Lemma 7.7 we argue that for changes in D when applying $\mathcal{P}$ it is sufficient to analyze the changes caused by $\mathcal{M}$. Then we use the previous derived bounds from Propositions 7.3 to 7.5 to argue that after $\log(n)$ rounds, there is a constant probability that D is big and gets converted to high min-entropy.
- Having a high min-entropy is advantageous over keeping track of D because a distribution can not lose min-entropy after applying a permutation. Further, Lemma 7.12 establishes that high min-entropy distributions also have high D with overwhelming probability.
- In Lemma 7.13 we combine the results of this section to get the statement that applying $\mathcal{P}^{\log^3(n)}$ to a non-special string produces a close to full min-entropy distribution with overwhelming probability.

Throughout this section, we will often need to compute probability bounds for the sum of random samples without replacement from a finite population. The remark that follows will allow us to use well-known concentration inequalities that hold for independent variables, even in the case of sampling without replacement.

Remark 7.1. *Let $X_1, \ldots, X_N$ denote a random sample from a population without replacement, and $Y_1, \ldots, Y_N$ denote a random sample from the same population with replacement. Theorem 4 of Hoeffding [Hoe63] states that for any continuous and convex function f:*

$$\mathbf{E}\left[f\left(\sum_i X_i\right)\right] \leq \mathbf{E}\left[f\left(\sum_i Y_i\right)\right].$$

Using the above inequality for $f(x) = \exp(\lambda x)$ allows us to recover the statements of Theorems 3.5 and 3.6 for $X = \sum_i X_i$, even if the X_i's are not independent.

7.1 Distance growth after each mixing step

This section is mostly concerned with the effect of the mixing step $\mathcal{M}$. In particular, applying a round of $\mathcal{M}$ to a string v can be approximately thought of as applying linear shifts to random symbols of v with probability proportional to the distance $D(v)$ of v from a special string.

In particular, let us use $\mathcal{U}_j(\cdot)$ to represent the permutation that shifts the j-th symbol of its input by a random amount.

Algorithm 6 $\mathcal{U}_j(v)$

1: Sample $a \sim \Sigma$ uniformly at random.
2: Let $v_j \leftarrow v_j + a$.
3: **return** v

This random linear shift affects the distance of the string in a way that depends on the symbols at positions $j-1$ and $j+1$, if they exist. For simplicity, we define the random variables X and Y that satisfy

$$X = \begin{cases} 0 & \text{w.p. } 1/|\Sigma| \\ 1 & \text{w.p. } 1-2/|\Sigma| \\ -1 & \text{w.p. } 1/|\Sigma| \end{cases}, \qquad Y = \begin{cases} 0 & \text{w.p. } 1-2/|\Sigma| \\ -1 & \text{w.p. } 2/|\Sigma|. \end{cases} \tag{6}$$

We will use X, Y to lower bound the change in the distance $D(v)$, since X, Y are stochastically dominated by it, as shown in the lemma below.

Lemma 7.2. *Let the random variables X, Y be as above. Then for $v_j v_{j+1} \in \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$, it holds that*

$$X \leq_{\mathsf{st}} D(\mathcal{U}_j(v)) - D(v).$$

Otherwise, when $v_j v_{j+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$,

$$Y \leq_{\mathsf{st}} D(\mathcal{U}_j(v)) - D(v).$$

Proof. Fix arbitrary v, j. We analyze the expected behavior of $D(\mathcal{U}_j(v))$ where v is an arbitrary string. All probabilities in this analysis are strictly over the randomness of a.

First, we treat the case that $v_j v_{j+1} \in \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$. In this case, we claim that X is stochastically dominated by $D(\mathcal{U}_j(v)) - D(v)$, because

1. If $v_j + a = v_j$, then $D(\mathcal{U}_j(v)) - D(v) = 0$, which happens with the same probability that $X = 0$.
2. If $v_j + a \notin \{\heartsuit, \clubsuit\}$ then $D(\mathcal{U}_j(v)) - D(v) \geq 1$ because index j is now detectable and if index $j-1$ was detectable before the change then it is also detectable after the change, too. This happens with the same probability as $X = 1$.
3. If $v_j + a \in \{\heartsuit, \clubsuit\}$ and $v_j + a \neq v_j$, then $D(\mathcal{U}_j(v)) \geq D(v) - 1$ because either j or $j-1$ might not be detectable anymore. This happens with the same probability as $X = -1$.

Now, we treat the case that $v_j v_{j+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$. In this case, we claim that Y is stochastically dominated by $D(\mathcal{U}_j(v)) - D(v)$, because

1. If $v_j + a \notin \{\heartsuit, \clubsuit\}$ then $D(\mathcal{U}_j(v)) \geq D(v)$ because the indices j and $j+1$ are detectable irrespective of their prior value, which happens with the same probability as $Y = 0$.
2. If $v_j + a = \heartsuit$ then $D(\mathcal{U}_j(v)) \geq D(v) - 1$ because j might not be detectable anymore but $j-1$ stays detectable if it was before.
3. If $v_j + a = \clubsuit$ then $D(\mathcal{U}_j(v)) \geq D(v) - 1$ because $j-1$ might not be detectable anymore but j stays detectable if it was before.

The last two events together happen with the same probability as $Y = -1$. □

The above allows us to lower bound the expected increase (henceforth also referred to as "growth") of $D(v)$ by the expectations of X and Y.

Analysis of the mixing steps. The following Propositions 7.3 to 7.5 give us lower bounds on the growth of D when applying $\mathcal{M}$ to a string that has an invalid counter. First, we introduce some shared notation. Recall an index i is detectable in v if

- $i \in [2, n-3]$ and $v_i v_{i+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$,
- $i = 1$ and $v_i v_{i+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit\}$, or
- $i = n-2$ and $v_i v_{i+1} \notin \{\heartsuit\clubsuit, \clubsuit\clubsuit\}$.

By definition, $D(v)$ measures the number of detectable indices in v. We define the sets

$$\mathcal{I} = \{y_1, \ldots, y_{n/5}\}, \qquad \mathcal{J} = \{z_1, \ldots, z_{n/5}\},$$

where $y_1, \ldots, y_{n/5}, z_1, \ldots, z_{n/5}$ are the indices sampled in $\mathcal{M}$.

In each round of $\mathcal{M}$, we call an index i *detected* if it is detectable and if $i \in \mathcal{I}$. We use $\mathsf{Detected}$ to refer to the set of detected indices. Then, each round of $\mathcal{M}$ can be decomposed into $|\mathsf{Detected}|$ permutations $\mathcal{U}$ as in Algorithm 6. We will also say that the indices y_i and z_i are *connected* in $\mathcal{M}$, since they correspond to applying the same "gate". Let

$$\mathsf{ChangeHit} = \{z_i \mid y_i \in \mathsf{Detected}, z_i \text{ is detectable}\}$$

be the entries that are detectable and that are connected to an entry in $\mathsf{Detected}$. Furthermore, let

$$\mathsf{ChangeMiss} = \{z_i \mid y_i \in \mathsf{Detected}, z_i \text{ is not detectable}\}$$

be the entries that are *not* detectable and are connected to an entry in $\mathsf{Detected}$.

Having established our notation, the following lemma shows that for a string v that has a relatively high distance $D(v)$, the mixing step $\mathcal{M}$ will increase its distance with good probability.

Proposition 7.3. *Let $n \geq 32$ and $|\Sigma| \geq 10000$. If $200\,000 \leq D(v) \leq n/16$, then*

$$\Pr[D(\mathcal{M}(v)) \geq 1.011 D(v)] \geq 0.99.$$

Proof. Let $\mathsf{ManyDetected}$ be the event that $|\mathsf{Detected}|$ is greater than $D(v)/10$. Then

$$\Pr[\overline{\mathsf{ManyDetected}}] \leq \left(\frac{e^{-1/2}}{(1/2)^{1/2}}\right)^{D(v)/5} \leq 0.9^{D(v)/5} < 10^{-1830}$$

by the multiplicative Chernoff bound and Remark 7.1. For the rest of the proof, we assume $\mathsf{ManyDetected}$.

After selecting $\mathcal{I}$ there are at most $4n/5$ possible choices for $j \notin \mathcal{I}$ and at least $n - 1 - D(v) - 2n/5 \geq n/2$ of those choices are such that $v_j v_{j+1} \in \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$. Therefore, the expected number of $j \in \mathcal{J}$ that are connected to a detected index is at least $\frac{n/2}{4n/5} \cdot D(v)/10 = D(v)/16$. Let $\mathsf{ManyIncreasing}$ be the event that the number of such indices is greater than $D(v)/32$.

By the multiplicative Chernoff bound and Remark 7.1 we get

$$\Pr[\overline{\mathsf{ManyIncreasing}} | \mathsf{ManyDetected}] < \left(\frac{e^{-1/2}}{(1/2)^{1/2}}\right)^{D(v)/16} < 0.9^{D(v)/16} < 10^{-571}.$$

For the rest of the proof, we assume $\mathsf{ManyIncreasing}$.

For any a we can lower bound

$$\Pr[D(\mathcal{M}(v)) - D(v) \geq a | \mathsf{ManyDetected}, \mathsf{ManyIncreasing}]$$

by a sum of changes in D when applying $\mathcal{U}_j$ to $D(v)/32$ wires j such that $v_j v_{j+1} \in \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$ and applying $\mathcal{U}_j$ to $D(v) - D(v)/32$ wires j such that $v_j v_{j+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$. By Lemma 7.2 we can bound the growth by

$$D(\mathcal{M}(v)) - D(v) \geq_{\mathsf{st}} \sum_{i \in [D(v)/32]} X_i + \sum_{i \in [D(v) - D(v)/32]} Y_i,$$

where $X_1, \ldots, X_{D(v)/32}, Y_1, \ldots Y_{D(v)-D(v)/32}$ are independent copies of the random variables X and Y as defined in Equation (6).

The expected growth of D in this process is

$$\begin{aligned}\mathbf{E}[D(\mathcal{M}(v)) - D(v)] &\geq \mathbf{E}[\sum_{i\in[D(v)/32]} X_i + \sum_{i\in[D(v)-D(v)/32]} Y_i] \\ &\geq (1-3/|\Sigma|)D(v)/32 + (-2/|\Sigma|)D(v) \\ &\geq ((1-3/120)/32 + (-1/120))D(v) \\ &\geq 0.022D(v)\end{aligned}$$

for $|\Sigma| \geq 120$ by Lemma 7.2. By the Hoeffding bound, we get that the probability that this process produces a growth less than $0.011D(v)$ is at most

$$\exp\left(-\frac{2(0.011D(v))^2}{2^2 \cdot D(v)}\right) \leq \exp\left(-\frac{5D(v)}{200\,000}\right).$$

The claim follows by multiplying the probabilities $(1-10^{-1830})(1-10^{-571})(1-\exp\left(-\frac{5D(v)}{200\,000}\right))$ and replacing $D(v)$ by 200 000. □

Proposition 7.4. *Let $n \geq 7\,000\,000\,000\,000$ and $|\Sigma| \geq 10000$. If $1 \leq D(v) \leq 200\,000$, then*

$$\Pr_{\mathsf{r}}[D(\mathcal{M}(v)) \geq 1.02D(v)] \geq 0.466.$$

Proof. Let ManyDetected be the event that the number of detected indices is greater than $D(v)/10$. Then, by Markov's inequality, we get $\Pr[\mathsf{ManyDetected}] > 1/2$ because the expectation is $D(v)/5$. For the rest of the proof, we assume ManyDetected.

Let $3n/5 \leq t \leq 4n/5$ be the number of indices j such that $j \notin \mathcal{I}$ and $j-1 \notin \mathcal{I}$, i.e., the indices that are available to be chosen for $\mathcal{J}$. The probability that one of the at most $D(v)$ detected indices in $\mathcal{I}$ is connected to an index $j \in \mathcal{J}$ is at most $(D(v))^2/(3n/5) \leq 70\,000\,000\,000/n$ (i.e. $|\mathsf{ChangeHit}| > 0$). For $n \geq 7\,000\,000\,000\,000$ there is at most 0.01 probability that $|\mathsf{ChangeHit}| > 0$. For the rest of the proof, we assume that $|\mathsf{ChangeHit}| = 0$.

By Lemma 7.2 for each entry in $j \in \mathsf{ChangeMiss}$ we get

$$\mathbf{E}[D(\mathcal{U}_j(y)) - D(y)] \geq -1/|\Sigma| + (1-2/|\Sigma|) = 1 - 3/|\Sigma| > 0.999,$$

for $|\Sigma| \geq 10000$. We decompose $\mathcal{M}$ into $|\mathsf{ChangeHit}|$ many permutations $\mathcal{U}_j$ where j is detectable and $|\mathsf{ChangeMiss}|$ many permutations, where j is not detectable.

Because $|\mathsf{ChangeHit}| = 0$ and $|\mathsf{ChangeHit}| + |\mathsf{ChangeMiss}| \geq D(v)/10$, we get $|\mathsf{ChangeMiss}| \geq D(v)/10$. By Lemma 7.2 we get

$$\Pr[D(\mathcal{M}(v)) - D(v) \leq 0.02D(v)] \leq \Pr\left[\sum_{i\in[\lceil D(v)/10\rceil]} X_i \leq 0.02D(v)\right],$$

where $X_1, \ldots, X_{|\mathsf{ChangeMiss}|}$ are independent copies of the random variable X as defined in Equation (6).

For $D(v) \geq 400$ by Hoeffding's inequality:

$$\Pr[D(\mathcal{M}(v)) - D(v) \leq 0.02D(v)] \leq \exp\left(-\frac{(0.0799D(v))^2}{D(v)2^2/10}\right) < 0.06$$

because the expectation is $(D(v)/10) \cdot 0.999$.

For $|\mathsf{ChangeHit}| < 200$ we numerically verified (see Appendix A)that the probability that $D(\mathcal{M}(v)) < D(v)$ is maximized when $D(v) = 10$ for $|\Sigma| \geq 10000$. For $D(v) = 10$ the probability that $D(\mathcal{M}(v)) < 1.02D(v)$ is less than 0.0002.

Overall, we get the probability that $D(\mathcal{M}(v)) \geq 1.02D(v)$ is at least $0.94 \cdot 1/2 \cdot 0.993 \geq 0.466$. □

Proposition 7.5. *Let* $|\Sigma| \geq 10000$. *If* $D(v) \leq n/1000$, *then* $\Pr(D(\mathcal{M}(v)) \geq D(v)) \geq 0.999$.

Proof. If all $D(v) \leq n/1000$ detectable indices are in $\mathcal{J}$, then there are at least $n/5 - n/1000$ non-detectable indices in $\mathcal{J}$.

The probability that $|\mathsf{ChangeHit}| > |\mathsf{ChangeMiss}|$ is maximized when there is one detected index in $\mathcal{I}$ because a lower fraction of $j \in \mathcal{J}$ is detectable than not. The probability that $|\mathsf{ChangeHit}| > |\mathsf{ChangeMiss}|$ when there is one detected index in $\mathcal{I}$ is at most $(n/1000)/(n/5) = 1/200$.

By Lemma 7.2, we have for each entry in $\mathsf{ChangeHit}$ the expected change is $-2/|\Sigma|$. We also get for each entry in $\mathsf{ChangeMiss}$ the expected change is $1 - 3/|\Sigma|$. Together, each pair of one entry from $\mathsf{ChangeHit}$ and one entry from $\mathsf{ChangeMiss}$, we get the expected growth is at least $(1 - 5/|\Sigma|) \geq 0.9995$ for $|\Sigma| \geq 10000$.

We decompose $\mathcal{M}$ into $\mathsf{ChangeHit}$ many permutations $\mathcal{U}_j$ where j is detectable and $\mathsf{ChangeMiss}$ many permutations, where j is not detectable. By Lemma 7.2 we get

$$\Pr[D(\mathcal{M}(v)) - D(v) < 0] \leq \Pr\left[\sum_{i \in [|\mathsf{ChangeMiss}|]} X_i + \sum_{i \in [|\mathsf{ChangeHit}|]} Y_i < 0\right],$$

where $X_1, \ldots, X_{|\mathsf{ChangeMiss}|}, Y_1, \ldots Y_{|\mathsf{ChangeHit}|}$ are independent copies of the random variables X and Y as defined in Equation (6).

If $|\mathsf{ChangeHit}| \leq |\mathsf{ChangeMiss}|$ by $X \geq_{\mathsf{st}} Y$ and Hoeffding inequality we have the probability that $D(\mathcal{M}(v)) < D(v)$ is at most

$$\begin{aligned}\Pr[\sum_{i \in [|\mathsf{ChangeMiss}|]} X_i + \sum_{i \in [|\mathsf{ChangeHit}|]} Y_i < 0] &\leq \exp\left(-\frac{(0.9995|\mathsf{ChangeMiss}|)^2}{|\mathsf{ChangeMiss}| \cdot 3^2}\right) \\ &\leq \exp\left(-\frac{(0.9995^2|\mathsf{ChangeMiss}|)}{3^2}\right) \\ &\leq 0.0002\end{aligned}$$

for $|\mathsf{ChangeHit}| \geq 200$.

For $|\mathsf{ChangeHit}| < 200$ we numerically verified (see Appendix A)that the probability that $D(\mathcal{M}(v)) < D(v)$ is maximized when $|\mathsf{ChangeHit}| = |\mathsf{ChangeMiss}| = 1$ (i.e. two $\mathcal{U}_j$ where, j is detectable and two where it is not) for $|\Sigma| \geq 10000$. For $|\mathsf{ChangeHit}| = |\mathsf{ChangeMiss}| = 1$ we have the probability that $D(\mathcal{M}(v)) < D(v)$ is less than 0.0002.

Overall, we get the probability that $D(\mathcal{M}(v)) \geq D(v)$ is at least $0.9998 \cdot 9995/10000 \geq 0.999$. □

The permutation $\mathcal{M}$ affects the distance of the string in ways that depend on the specific structure of the string. For simplicity, we define the random variable Z that satisfies

$$Z = \begin{cases} 0.0157 & \text{w.p. } 0.466 \\ 0 & \text{w.p. } 0.533 \\ -1 & \text{w.p. } 0.001. \end{cases} \tag{7}$$

We will use Z to lower bound the change in the distance $D(v)$, since Z are stochastically dominated by it, as shown in the lemma below.

Lemma 7.6. *Let the random variable* Z *be as above. If the size of the alphabet* $|\Sigma| \geq 10000$, $n \geq 7\,000\,000\,000\,000$, *and* $1 \leq D(v) \leq n/1000$, *then*

$$Z \leq_{\mathsf{st}} \log(\mathcal{M}(D(v))) - \log(D(v)).$$

Proof. Follows directly from Propositions 7.3 to 7.5 and $\log(1.011) > 0.0157$. □

7.2 Min-entropy growth after mixing

Lemma 7.7. *Let $|\Sigma| \geq 10000$ and $n \geq 2\,000\,000$. There exists some constant c such that for all v with invalid counter, the bound $H_\infty(\mathcal{P}^t(v)) \geq n\log(|\Sigma|)/c$ holds with probability at least 0.662, when $t \geq \log(n)$.*

Proof. Note that if some distribution $\mathcal{D}$ has min-entropy h, then applying a permutation to it cannot decrease its min-entropy. Therefore, it is enough to analyze whether in the $\log(n)$ rounds there is one round that introduces at least $n\log|\Sigma|/1000$ min-entropy.

If during step $\mathcal{W}$ we have $v_{i^*} = \heartsuit$, $v_{n-2} = \heartsuit$, and $v_n \in [k{+}1]$ then applying the independent permutations $(\pi_{i,j})_{i\in[n-1],j\in\Sigma}$ to at least $n-3$ wires introduces a min-entropy of at least $(n-3)\log|\Sigma|$. Also, if step $\mathcal{B}$ applies one layer of brickwork to $n-4$ wires, it introduces at least $(n-4)\log|\Sigma|$ min-entropy.

Therefore, all that is left to prove is that min-entropy exceeds $n\log|\Sigma|/1000$ even in the case when during step $\mathcal{W}$ we have $v_{i^*} \neq \heartsuit$, $v_{n-2} \neq \heartsuit$, or $v_n \notin [k+1]$, and $\mathcal{B}$ does not apply a layer of brickwork.

Throughout this proof, we analyze the distance measure D (Definition 4.6). The counter step $\mathcal{C}$ does not affect this measure, and neither does $\mathcal{B}$ if it does not apply a layer of brickwork. Also, in $\mathcal{W}$ with $v_{i^*} \in \{\heartsuit, \clubsuit\}$ being distinct from the values $w^1_{i^*}, \ldots, w^{k+1}_{i^*}$, introduced in $\mathcal{W}$, and $v_{n-2} = \clubsuit$ there is no change to D. Therefore, it is sufficient to analyze the growth of D when we only apply rounds of $\mathcal{M}$.

Proposition 7.8. *For all v with an invalid counter, there exists an index $i < \log(n)$ such that $D(\mathcal{M}^i(v)) \geq n/(1000\cdot 2^{4\,000\,000})$ with probability at least 0.9999.*

Proof. For all v with an invalid counter, we have $D(v) \geq 1$. If $D(v) \geq n/(2^{40\,000\,010})$ the condition is already met. For all string v such that $D(v) < n/(2^{40\,000\,010})$ we can apply the previous analysis for the next $(\log(n) - 10 - \log(D(v))/\log(3))$ steps because with each application of $\mathcal{M}$ we have that $D(v)$ can grow at most by a factor of three. Therefore, in these steps, by Lemma 7.6 we have that

$$\begin{aligned}
&\mathbf{E}[\log(D(\mathcal{M}^{i+1}(v))) - \log(D(\mathcal{M}^i(v)))]\\
&\geq (0.001\cdot(-1)) + (0.533\cdot 0) + (0.466\cdot 0.0157)\\
&= 0.0063162.
\end{aligned}$$

Over all rounds we get

$$\begin{aligned}
&\mathbf{E}[\log(D(\mathcal{M}^{(\log(n)-10-\log(D(v)))/\log(3)}(v))) - \log(D(v))]\\
&\geq 0.0063162\cdot(\log(n) - 10 - \log(D(v)))/\log(3).
\end{aligned}$$

By Lemma 7.6 and Hoeffding's inequality, we get that the probability that

$$\begin{aligned}
&\log(D(\mathcal{M}^{(\log(n)-10-\log(D(v)))/\log(3)}(v))) - \log(D(v))\\
&\leq_{\mathsf{st}} \sum_{i\in[(\log(n)-10-\log(D(v)))/\log(3)]} Z_i
\end{aligned}$$

with $Z_1, \ldots, Z_{(\log(n)-10-\log(D(v)))/\log(3)}$ being independent copies of the random variable Z as defined in Equation (7) is less than half of its expectation is at most

$$\begin{aligned}
&\exp\left(-\frac{(0.0063162\cdot(\log(n) - 10 - \log(D(v)))/2\log(3))^2}{(1.0157)^2(\log(n) - 10 - \log(D(v)))/\log(3)}\right)\\
&\leq \exp\left(-\frac{0.0063162^2\cdot(\log(n) - 10 - \log(D(v)))/\log(3)}{4.12658596}\right)\\
&\leq \exp\left(-\frac{6.3162^2\cdot(\log(n) - 10 - \log(n) + 4\,000\,010)/\log(3)}{4\,126\,586}\right)\\
&\leq \exp(-35) \leq 10^{-15}.
\end{aligned}$$

□

Proposition 7.9. *If $D(v) \geq n/c'$ where c' is some constant then with probability at least $2/3$ we have that $H_\infty(\mathcal{M}(v)) \geq n\log(|\Sigma|)/15c'$.*

Proof. We have $\mathcal{M}$ introduces $|\mathsf{Detected}|\log(|\Sigma|)$ min-entropy because it independently turns $|\mathsf{Detected}|$ wires uniformly random. We have that the expected size of $\mathsf{Detected}$ is $(n/5c')$ By Markov's inequality, the claim follows. □

It follows that there exists a constant $c > 1$ such that for all v with an invalid counter, we have with probability at least $1 - 1/3 - 0.0001 \geq 0.6665$ there exists an index $i \leq \log(n)$ such that $H_\infty(\mathcal{M}^i(v)) \geq n\log(|\Sigma|)/c$.

The claim of Lemma 7.7 follows because min-entropy can not decrease when applying permutations. □

Lemma 7.10. *For all v with an invalid counter, $\mathcal{M}(v)$ has an invalid counter, too.*

Proof. For $\mathcal{M}$ to make any change to v, we need that $\mathsf{Detected} \geq 1$. However, $\mathsf{Detected} \cap \mathcal{J} = \emptyset$ implies that indices in $\mathsf{Detected}$ are not changed. Therefore, if $D(v) \geq 1$ then $D(\mathcal{M}(v)) \geq 1$. Further, a string v has a valid counter if and only if $D(v) = 0$. □

Corollary 7.11. *There exists some constant c such that for all $|\Sigma| \geq 10000$, $n \geq 2\,000\,000$, a v with invalid counter we have $H_\infty(\mathcal{P}^t(v)) \geq n\log(|\Sigma|)/c$ with probability at least $1 - \varepsilon$ where $t \geq \log(1/\varepsilon)\log(n)$.*

Proof. The claim follows applying Lemma 7.7. If $H_\infty(\mathcal{P}^{\log(n)}(v)) \geq n\log(|\Sigma|)/c$, we are done because applying a permutation can not reduce min-entropy. If, on the other hand $H_\infty(\mathcal{P}^{\log(n)}(v)) < n\log(|\Sigma|)/c$ We get by Lemma 7.10 that the counter of $\mathcal{P}^{\log(n)}(v)$ is still invalid. Then we apply Lemma 7.7 for the next $\log(n)$ repetitions of $\mathcal{P}^{\log(n)}$. We repeat this process $\log(1/\varepsilon)$ times.

Because the probabilities of achieving high min-entropy in each set of $\log(n)$ permutations $\mathcal{P}$ is independent we get that the probability that $H_\infty(\mathcal{P}^t(v)) \geq n\log(|\Sigma|)/c$ is at least $1-\varepsilon$ where $t = \log(1/\varepsilon)\log(n)$. □

Lemma 7.12. *Let $c > 1$ be some constant and sample v from a distribution with min-entropy at least $\log(|\Sigma|)n/c$ with $|\Sigma| \geq 2^{9c}$ then $D(v) \leq n/2c$ with probability 2^{-n}.*

Proof. First, we establish that $H_\infty(v) \leq H_\infty(\mathcal{M}(v))$ because $\mathcal{M}$ is a permutation and can, therefore, not reduce min-entropy.

We define

$$\mathsf{LowDistance} = \{x \mid x \in \Sigma^n, D(x) \leq n/2c\}$$

be the strings that have distance D lower than a $2c$ fraction. Further, we define

$$\Delta(x) = \{i \mid i \in [n-2], x_i \notin \{\heartsuit, \clubsuit\}\}$$

a Hamming-like distance and

$$\mathsf{LowHamming} = \{x \mid x \in \Sigma^n, \Delta(x) \leq n/2c\}.$$

We have $|\mathsf{LowDistance}| \leq |\mathsf{LowHamming}|$ because $D(x) \geq \Delta(x)$ which in turn is caused by the fact that if $x_i \notin \{\heartsuit, \clubsuit\}$ then also $x_i x_{i+1} \notin \{\heartsuit\heartsuit, \heartsuit\clubsuit, \clubsuit\clubsuit\}$. We upper bound $|\mathsf{LowHamming}|$.

Fix some $n/2c$-sized subset U of $[n-2]$. Now we want for all $u \in U$ that $v_u \in \{\heartsuit, \clubsuit\}$ and for all $u \in [n]\setminus U$ that $v_u \in \Sigma$. So, in total we get at most $2^{n-2-n/2c}|\Sigma|^{n/2c+2}$ different strings if all wires that are distinct from $\heartsuit$ and $\clubsuit$ are in U. With a union bound over all possible U and the Stirling approximation, we get

$$\begin{aligned}
|\mathsf{LowHamming}| &\leq \binom{n-2}{n-2-n/2c} 2^{n-2-n/2c}|\Sigma|^{n/2c+2} \\
&\leq \frac{(n-2)e}{n-2-n/2c}^{n-2-n/2c} 2^{n-2-n/2c}|\Sigma|^{n/2c+2} \\
&\leq (4e)^{n-2-n/2c}|\Sigma|^{n/2c+2}
\end{aligned}$$

if $n/c < (n-2)$. Therefore, $|\mathsf{LowDistance}| \leq (4e)^{n-2-n/2c}|\Sigma|^{n/2c+2}$.

Because we have $H_\infty(v) \geq \log(|\Sigma|)n/c$ each of these string of $|\mathsf{LowDistance}|$ can only appear with probability $|\Sigma|^{-n/c}$. By union bound, we get the probability that $v \in \mathsf{LowDistance}$ is

$$\leq \frac{(4e)^{n-2-n/2c}|\Sigma|^{n/2c+2}}{|\Sigma|^{n/c}} \leq \frac{(4e)^n}{|\Sigma|^{n/2c}}$$

for $|\Sigma| \geq 4e$ and

$$\leq 2^{\log(4e)n - \log(|\Sigma|)n/2c} \leq 2^{-n}$$

for $\log(|\Sigma|) \geq 9c > 2c(\log(4e)+1)$. □

Lemma 7.13. *There exist constants $c, c', n^* > 1$ such that for all non-special $v \in \Sigma^n$, $|\Sigma| \geq c'$, and $n \geq n^*$ we have $H_\infty(\mathcal{P}^t(v)) \geq (n-7)\log(|\Sigma|)$ with overwhelming probability where $t \in \Omega(\log^3(n))$.*

Proof. By Corollary 7.11 for $\varepsilon = 2^{-\log^2(n)}$ and v with an invalid counter, we get that there exists some constant such that with overwhelming probability

$$H_\infty(\mathcal{P}^{\Omega(\log^3(n))}(v)) \geq n/c^*.$$

By union bound on Lemma 7.12, we get that within the next $\log^2(n)$ rounds of $\mathcal{P}$ that D is greater than $n/2c^*$ with overwhelming probability. In these $\log^2(n)$ rounds are at least $\log^2(n)$ layers of $\mathcal{B}$. Each layer has at least $1/2c^*$ probability of applying a layer of random brickwork on $n-6$ wires. One layer of this brickwork ensures the min-entropy is at least $(n-7)\log(|\Sigma|)$.

The only way for v to be non-special is to have an invalid counter or if $v_n \notin [5]$. We have already covered the first case. For the second case, we have that in any application of $\mathcal{P}$, we have that $\mathcal{M}$ and $\mathcal{C}$ do not change v_n, and $\mathcal{B}$ may only change n if it introduces $(n-3)\log(|\Sigma|)$ bits of min entropy. Further, if for some string x with $x_n \notin [5]$ then $H_\infty(x) \geq (n-3)\log(|\Sigma|)$ because $\mathcal{W}$ chooses $n-3$ wire-values uniformly at random. □

8 Discussion

Our result is a *formal* separation. We believe understanding why block cipher constructions could be pseudorandom permutations is a fundamental question. If it were the case that limited independence would be sufficient for the composition of simple permutations to be pseudorandom, this would have been a remarkable insight into our understanding of pseudorandomness. Our result shows that this cannot be true in its most general form. However, that does not mean that proving k-wise independence for block ciphers is meaningless, as our construction is highly artificial, and limited independence, when combined with high degree/resistance to algebraic attacks, can still be sufficient to exclude all possible attacks.

Although we refute Conjecture 1 in its broadest form, our counterexample has apparent limitations. To further encourage the study of pseudorandom permutations, we posit the following conjectures. Each is a weakened variant of Conjecture 1 not refuted by our work. Understanding their veracity may lead to a sharper perspective on the hardness of pseudorandom permutations.

Conjecture 3 (Logarithmic Independence)**.** *Let $\mathcal{P}$ be a randomized permutation $\{0,1\}^n \to \{0,1\}^n$ such that each output bit only depends on a constant number of input bits. If $\mathcal{P}^T$ is negligibly far from $\log(n)$-wise independence, then $\mathcal{P}^T$ is also a pseudorandom permutation.*

Our counterexample crucially relies on the ability to distinguish a constant number of invariants using only a constant number of wires. This approach is not viable when $\omega(1)$-many invariants have to be maintained.

Conjecture 4 (Increased Pseudorandomness Threshold)**.** *Let $\mathcal{P}$ be a randomized permutation $\{0,1\}^n \to \{0,1\}^n$ such that each output bit only depends on a constant number of input bits. If $\mathcal{P}^T$ is negligibly far from 4-wise independence, then there exists a polynomial p such that $\mathcal{P}^{2T}$ is a pseudorandom permutation.*

We use that on specific inputs, the permutation is not random at all for all rounds but the last. This pattern is not possible when the permutation needs to be 4-wise independent in intermediate rounds.

Alternatively, we can also put tighter restrictions on the kinds of allowed round functions

Conjecture 5 (Constant IO Locality)**.** *Let $\mathcal{P} : \{0,1\}^n \to \{0,1\}^n$ be a randomized permutation such that $\mathcal{P}$ has constant input and output locality, meaning that each input wire can only influence a constant number of output wires and each output wire can only be influenced by a constant number of input wires. If $\mathcal{P}^T$ is negligibly far from 4-wise independence, then $\mathcal{P}^T$ is also a pseudorandom permutation.*

These restrictions are further tightened in the following conjecture.

Conjecture 6 (Independently Sampled Gates)**.** *Let $\mathcal{P}$ be a randomized permutation $\{0,1\}^n \to \{0,1\}^n$ such that $\mathcal{P}$ only acts on a constant number of wires. If $\mathcal{P}^T$ is negligibly far from 4-wise independence, then $\mathcal{P}^T$ is also a pseudorandom permutation.*

The ability to switch from an entirely predictable string to a random one in one round is fundamental to our counterexample. This technique is not possible when each round can only change a constant number of wires.

This list would not be complete without reiterating Gower's conjecture. It was put forth by Gowers [Gow96], made explicit by Barak [Bar17], and has interesting connections to obfuscation candidates by Canetti et al. [CCMR24]. This conjecture is weaker than Conjectures 3 to 6, that is, to refute the following conjecture, one would also refute Conjectures 3 to 6.

Conjecture 7 (Conjectured by [Gow96])**.** *Let $\mathcal{P} : \{0,1\}^n \to \{0,1\}^n$ be a randomized permutation such that $\mathcal{P}$ picks three random wires and applies a uniformly random permutation on them. Then there exists a polynomial p such that $\mathcal{P}^{p(n)}$ is a pseudorandom permutation.*

None of these conjectures is likely to be proven any time soon unconditionally, as such a proof would also imply the existence of one-way functions and $\mathsf{P} \neq \mathsf{NP}$. We hope that attempts to prove them under non-trivial conditions (such as the existence of one-way functions) or refute them lead to new insights about the pseudorandomness of random reversible circuits.

Acknowledgments

We thank Antoine Joux and Vinod Vaikuntanathan for the helpful discussions. We also thank the anonymous Eurocrypt 2026 reviewers for their comments and suggestions.

Dujmovic is supported by the European Union (ERC, LACONIC, 101041207). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. Pelecanos is supported by DARPA under Agreement No. HR00112020023. Tessaro's research was partially supported by NSF grants CNS-2026774, CNS-2154174, CNS-2426905, a gift from Microsoft, and a Stellar Development Foundation Academic Research Award.

References

[AES01] Advanced Encryption Standard (AES). National Institute of Standards and Technology, NIST FIPS PUB 197, U.S. Department of Commerce, November 2001.

[Bar17] Boaz Barak. The complexity of public-key cryptography. In *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, pages 45–77. Springer, 2017.

[BH08] Alex Brodsky and Shlomo Hoory. Simple permutations mix even better. *Random Struct. Algorithms*, 32(3):274–289, 2008.

[BHH16] Fernando GSL Brandao, Aram W Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, 2016.

[BKL+12] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, François-Xavier Standaert, John P. Steinberger, and Elmar Tischhauser. Key-alternating ciphers in a provable setting: Encryption using a small number of public permutations - (extended abstract). In *EUROCRYPT*, volume 7237 of *Lecture Notes in Computer Science*, pages 45–62. Springer, 2012.

[BKR11] Andrey Bogdanov, Dmitry Khovratovich, and Christian Rechberger. Biclique cryptanalysis of the full AES. In Dong Hoon Lee and Xiaoyun Wang, editors, *ASIACRYPT 2011*, volume 7073 of *LNCS*, pages 344–371, Seoul, South Korea, December 4–8, 2011. Springer Berlin Heidelberg, Germany.

[BLS25] Tim Beyne, Gregor Leander, and Immo Schütt. Pairwise independence of AES-like block ciphers. Cryptology ePrint Archive, Paper 2025/1495, 2025.

[BP14] Abhishek Banerjee and Chris Peikert. New and improved key-homomorphic pseudorandom functions. In Juan A. Garay and Rosario Gennaro, editors, *CRYPTO 2014, Part I*, volume 8616 of *LNCS*, pages 353–370, Santa Barbara, CA, USA, August 17–21, 2014. Springer Berlin Heidelberg, Germany.

[BPR12] Abhishek Banerjee, Chris Peikert, and Alon Rosen. Pseudorandom functions and lattices. In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 719–737, Cambridge, UK, April 15–19, 2012. Springer Berlin Heidelberg, Germany.

[BS91] Eli Biham and Adi Shamir. Differential cryptanalysis of des-like cryptosystems. *J. Cryptol.*, 4(1):3–72, 1991.

[CCMR24] Ran Canetti, Claudio Chamon, Eduardo R. Mucciolo, and Andrei E. Ruckenstein. Towards general-purpose program obfuscation via local mixing. In Elette Boyle and Mohammad Mahmoody, editors, *TCC 2024, Part IV*, volume 15367 of *LNCS*, pages 37–70, Milan, Italy, December 2–6, 2024. Springer, Cham, Switzerland.

[CDK+18] Benoît Cogliati, Yevgeniy Dodis, Jonathan Katz, Jooyoung Lee, John P. Steinberger, Aishwarya Thiruvengadam, and Zhe Zhang. Provable security of (tweakable) block ciphers based on substitution-permutation networks. In Hovav Shacham and Alexandra Boldyreva, editors, *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, volume 10991 of *Lecture Notes in Computer Science*, pages 722–753. Springer, 2018.

[CHH+24] Chi-Fang Chen, Jeongwan Haah, Jonas Haferkamp, Yunchao Liu, Tony Metger, and Xinyu Tan. Incompressibility and spectral gaps of random circuits, 2024.

[CL18] Benoît Cogliati and Jooyoung Lee. Wide tweakable block ciphers based on substitution-permutation networks: Security beyond the birthday bound. *IACR Cryptol. ePrint Arch.*, 2018:488, 2018.

[CP02] Nicolas Courtois and Josef Pieprzyk. Cryptanalysis of block ciphers with overdefined systems of equations. In Yuliang Zheng, editor, *ASIACRYPT 2002*, volume 2501 of *LNCS*, pages 267–287, Queenstown, New Zealand, December 1–5, 2002. Springer Berlin Heidelberg, Germany.

[CS14] Shan Chen and John P. Steinberger. Tight security bounds for key-alternating ciphers. In *EUROCRYPT*, volume 8441 of *Lecture Notes in Computer Science*, pages 327–350. Springer, 2014.

[Din25] Itai Dinur. New techniques for analyzing differentials with application to AES. Cryptology ePrint Archive, Paper 2025/1326, 2025.

[DKS+17] Yevgeniy Dodis, Jonathan Katz, John P. Steinberger, Aishwarya Thiruvengadam, and Zhe Zhang. Provable security of substitution-permutation networks. *IACR Cryptol. ePrint Arch.*, 2017:16, 2017.

[DKW22] Yevgeniy Dodis, Harish Karthikeyan, and Daniel Wichs. Small-box cryptography. In Mark Braverman, editor, *ITCS 2022*, volume 215, pages 56:1–56:25, Berkeley, CA, USA, January 31 – February 3, 2022. LIPIcs.

[DSC93] Persi Diaconis and Laurent Saloff-Coste. Comparison Theorems for Reversible Markov Chains. *The Annals of Applied Probability*, 3(3):696 – 730, 1993.

[DSSL16] Yevgeniy Dodis, Martijn Stam, John P. Steinberger, and Tianren Liu. Indifferentiability of confusion-diffusion networks. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II*, volume 9666 of *Lecture Notes in Computer Science*, pages 679–704. Springer, 2016.

[GGM84] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions (extended abstract). In *25th Annual Symposium on Foundations of Computer Science, West Palm Beach, Florida, USA, 24-26 October 1984*, pages 464–479. IEEE Computer Society, 1984.

[GHKO25] William Gay, William He, Nicholas Kocurek, and Ryan O'Donnell. Pseudorandomness properties of random reversible circuits. In Yael Tauman Kalai and Seny F. Kamara, editors, *CRYPTO 2025, Part I*, volume 16000 of *LNCS*, pages 651–678, Santa Barbara, CA, USA, August 17–21, 2025. Springer, Cham, Switzerland.

[GHP25] Lucas Gretta, William He, and Angelos Pelecanos. More efficient approximate k-wise independent permutations from random reversible circuits via log-sobolev inequalities. In Yossi Azar and Debmalya Panigrahi:, editors, *36th SODA*, pages 5582–5598, New Orleans, LA, USA, January 12–15, 2025. ACM-SIAM.

[Gow96] W. T. Gowers. An almost m-wise independed random permutation of the cube. *Comb. Probab. Comput.*, 5:119–130, 1996.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[HMMR04] Shlomo Hoory, Avner Magen, Steven Myers, and Charles Rackoff. Simple permutations mix well. In Josep Díaz, Juhani Karhumäki, Arto Lepistö, and Donald Sannella, editors, *ICALP 2004*, volume 3142 of *LNCS*, pages 770–781, Turku, Finland, July 12–16, 2004. Springer Berlin Heidelberg, Germany.

[HMMR05] Shlomo Hoory, Avner Magen, Steven A. Myers, and Charles Rackoff. Simple permutations mix well. *Theor. Comput. Sci.*, 348(2-3):251–261, 2005.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[HT16] Viet Tung Hoang and Stefano Tessaro. Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In *CRYPTO (1)*, volume 9814 of *Lecture Notes in Computer Science*, pages 3–32. Springer, 2016.

[JK97] Thomas Jakobsen and Lars R. Knudsen. The interpolation attack on block ciphers. In *FSE*, volume 1267 of *Lecture Notes in Computer Science*, pages 28–40. Springer, 1997.

[Knu94] Lars R. Knudsen. Truncated and higher order differentials. In Bart Preneel, editor, *Fast Software Encryption: Second International Workshop. Leuven, Belgium, 14-16 December 1994, Proceedings*, volume 1008 of *Lecture Notes in Computer Science*, pages 196–211. Springer, 1994.

[Knu98] Lars Knudsen. Deal - a 128-bit block cipher. In *NIST AES Proposal*, 1998.

[KW02] Lars R. Knudsen and David A. Wagner. Integral cryptanalysis. In *FSE*, volume 2365 of *Lecture Notes in Computer Science*, pages 112–127. Springer, 2002.

[LPTV23] Tianren Liu, Angelos Pelecanos, Stefano Tessaro, and Vinod Vaikuntanathan. Layout graphs, random walks and the t-wise independence of SPN block ciphers. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part III*, volume 14083 of *LNCS*, pages 694–726, Santa Barbara, CA, USA, August 20–24, 2023. Springer, Cham, Switzerland.

[LR86] Michael Luby and Charles Rackoff. How to construct pseudo-random permutations from pseudo-random functions (abstract). In Hugh C. Williams, editor, *CRYPTO'85*, volume 218 of *LNCS*, page 447, Santa Barbara, CA, USA, August 18–22, 1986. Springer Berlin Heidelberg, Germany.

[LR88] Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Journal on Computing*, 17(2):373–386, 1988.

[LS14] Rodolphe Lampe and Yannick Seurin. Security analysis of key-alternating feistel ciphers. In *FSE*, volume 8540 of *Lecture Notes in Computer Science*, pages 243–264. Springer, 2014.

[LTV21] Tianren Liu, Stefano Tessaro, and Vinod Vaikuntanathan. The t-wise independence of substitution-permutation networks. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part IV*, volume 12828 of *LNCS*, pages 454–483, Virtual Event, August 16–20, 2021. Springer, Cham, Switzerland.

[Mau92] Ueli Maurer. A simplified and generalized treatment of luby-rackoff pseudorandom permutation generator. In *International Conference on the Theory and Application of Cryptographic Techniques*, 1992.

[MY92] Mitsuru Matsui and Atsuhiro Yamagishi. A new method for known plaintext attack of FEAL cipher. In *EUROCRYPT*, volume 658 of *Lecture Notes in Computer Science*, pages 81–91. Springer, 1992.

[NR97] Moni Naor and Omer Reingold. Number-theoretic constructions of efficient pseudo-random functions. In *38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997*, pages 458–467. IEEE Computer Society, 1997.

[NRR02] Moni Naor, Omer Reingold, and Alon Rosen. Pseudorandom functions and factoring. *SIAM J. Comput.*, 31(5):1383–1404, 2002.

[Sin92] Alistair Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.

[Ste12] John P. Steinberger. Improved security bounds for key-alternating ciphers via hellinger distance. *IACR Cryptol. ePrint Arch.*, 2012:481, 2012.

[Vau03] Serge Vaudenay. Decorrelation: A theory for block cipher security. *J. Cryptol.*, 16(4):249–286, 2003.

A Code

This section contains the python code for numerically verifying statements in Propositions 7.4 and 7.5.

```
from fractions import Fraction # Rationals to avoid rounding inaccuracy
alphabet_size = 10000
X = dict([(−1, Fraction(1,alphabet_size)),(0, Fraction(1,alphabet_size)),(1, 1 − Fraction(2,alphabet_size))])
Y = dict([(−1, Fraction(2,alphabet_size)),(0, 1 − Fraction(2,alphabet_size))])

def sum_dist(X: dict, Y: dict): # Sum of two independent random variables
    output = dict()
    for vx,px in X.items():
        for vy,py in Y.items():
            if vx+vy not in output:
                output[vx+vy] = px*py
            else:
                output[vx+vy] += px*py
    return output
def geq(X: dict, a): # Computes the probability that a sample from X is ≥a
    output = 0
    for vx,px in X.items():
        if vx >= a:
            output += px
    return output
def greater(X: dict, a): # Computes the probability that a sample from X is > a
    output = 0
    for vx,px in X.items():
        if vx > a:
            output += px
    return output

# Prop 7.4
s = dict([(0, Fraction(1, 1))])
minimum_prob = 1
minimum_prob_index = 0
#i = ⌈D(v)/10⌉
for i in range(1, 41):
    s = sum_dist(s,X)
    tmp = greater(s, 0.02*(i*10))
    if tmp < minimum_prob:
        minimum_prob = tmp
        minimum_prob_index = i
print("The smallest possible value for Pr[Σ_i X_i > 0.02D(v)] is", float(minimum_prob)," at D(v) =", 10 * minimum_prob_index)

# Prop 7.5
s = dict([(0, Fraction(1, 1))])
minimum_prob = 1
minimum_prob_index = 0
for i in range(1, 201):
    s = sum_dist(sum_dist(s,X),Y)
    tmp = geq(s, 0)
    if tmp < minimum_prob:
        minimum_prob = tmp
        minimum_prob_index = i
print("The smallest possible value for Pr[Σ_i X_i + Σ_i Y_i ≥0] is", float(minimum_prob)," at i =", minimum_prob_index)
```