The t-wise Independence of Substitution-Permutation Networks

Tianren Liu
University of Washington
tianrenl@uw.edu

Stefano Tessaro
University of Washington
tessaro@cs.washington.edu

Vinod Vaikuntanathan
MIT
vinodv@mit.edu

July 22, 2021

Abstract

Block ciphers such as the Advanced Encryption Standard (Rijndael) are used extensively in practice, yet our understanding of their security continues to be highly incomplete. This paper promotes and continues a research program aimed at *proving* the security of block ciphers against important and well-studied classes of attacks. In particular, we initiate the study of (almost) t-wise independence of concrete block-cipher construction paradigms such as substitution-permutation networks and key-alternating ciphers. Sufficiently strong (almost) pairwise independence already suffices to resist (truncated) differential attacks and linear cryptanalysis, and hence this is a relevant and meaningful target. Our results are two-fold.

Our first result concerns substitution-permutation networks (SPNs) that model ciphers such as AES. We prove the almost pairwise-independence of an SPN instantiated with concrete S-boxes together with an appropriate linear mixing layer, given sufficiently many rounds and independent sub-keys. Our proof relies on a *characterization* of S-box computation on input differences in terms of sampling output differences from certain subspaces, and a new randomness extraction lemma (which we prove with Fourier-analytic techniques) that establishes when such sampling yields uniformity. We use our techniques in particular to prove almost pairwise-independence for sufficiently many rounds of both the AES block cipher (which uses a variant of the patched inverse function $x \mapsto x^{-1}$ as the S-box) and the MiMC block cipher (which uses the cubing function $x \mapsto x^3$ as the S-box), assuming independent sub-keys.

Secondly, we show that instantiating a key-alternating cipher (which can be thought of as a degenerate case of SPNs) with most permutations gives us (almost) t-wise independence in $t+o(t)$ rounds. In order to do this, we use the probabilistic method to develop two new lemmas, an *independence-amplification lemma* and a *distance amplification lemma*, that allow us to reason about the evolution of key-alternating ciphers.

Contents

1 Introduction **1**
- 1.1 Our Results and Techniques . . . 2
- 1.2 Perspectives and Open Problems . . . 4
- 1.3 Related Work . . . 5

2 Preliminaries **7**
- 2.1 Almost t-wise Independent Permutations and Cryptanalysis . . . 8
- 2.2 Key-Alternating Ciphers and Substitution Permutation Networks . . . 9
- 2.3 Trace in Fields of Characteristic Two . . . 11
- 2.4 Basics of Discrete Fourier Analysis . . . 13
- 2.5 Tail Bounds . . . 14

3 Pairwise Independence of SPNs **15**
- 3.1 The S-box Extraction Lemma . . . 15
- 3.2 Properties of Mixing Functions . . . 21
- 3.3 Proofs of Pairwise Independence . . . 22
- 3.4 AES is Almost Pairwise-Independent . . . 26
- 3.5 Multi-round SPNs and AES . . . 29

4 t-wise Independence of KAC **29**
- 4.1 Definitions and Notations . . . 30
- 4.2 Existential Results for Key Alternating Ciphers . . . 31
- 4.3 Independence Amplification . . . 32
- 4.4 Distance Amplification . . . 35

A Pairwise Independence – Two-round Convergence **43**

B Attack against 4-wise Independence **44**
- B.1 Warm-up: Pointwise t-wise Independence of Generalized KAC . . . 45

C Alternative Proof of the Extraction Lemma **47**

1 Introduction

Block ciphers are among the most fundamental building blocks in cryptography, and applications demand strong pseudorandomness properties from them. However, the simplicity of widely adopted designs, such as Substitution-Permutation Networks (SPNs), which underlie AES, is inherently at odds with the reductionist approach of provable security, as there are no clear underlying hard mathematical problems upon which security can be based. Instead, the security validation of block ciphers has gone through cryptanalysis, and considered a number of different techniques, including *linear* [MY92] and *differential* [BS91] cryptanalysis, higher-order [Lai94] and truncated [Knu94] differential attacks, impossible differential attacks [Knu98], algebraic attacks [JK97], integral cryptanalysis [KW02], biclique attacks [BKR11], and so on.

Lacking full proofs of security, the next best thing is to prove that certain relevant *classes* of attacks *cannot* possibly succeed. The more "concrete" and less "asymptotic" such a proof is, the better, and the class of attacks should be as large as possible. The most successful such effort has developed provable bounds for linear and differential cryptanalysis, starting with the seminal work of Nyberg and Knudsen [NK95], and culminating with fairly precise estimates for concrete block ciphers like AES (see e.g. [KMT01b, KMT01a, PSC+02, PSLL03, Kel04, KS07]).

t-wise independence. In this paper, we move one step forward and study the (almost) t-*wise independence* of concrete block ciphers – namely, for a block cipher $E : \{0,1\}^s \times \{0,1\}^n \to \{0,1\}^n$, we demand that for any distinct t inputs $x_1, \ldots, x_t$ and a random key S, the distribution of

$$E(S, x_1), \ldots, E(S, x_k)$$

is statistically close to that of t uniform, but distinct, n-bit strings.

This property is attractive for two reasons. First and foremost, it is potentially achievable unconditionally by a concrete design, as long as $s \geq t \cdot n$. For example, a variant of AES-128 with 11 independent round keys[1] can (potentially) be 11-wise independent. Second, t-wise independence already implies resilience against a large class of attacks that have been previously studied. Indeed, the case $t = 2$ (i.e., almost pairwise independence) already implies resilience to linear and differential cryptanalysis but also to truncated differential attacks and any other attack that exploits statistical deviations of pairs of outputs. Similarly, t-wise independence implies resilience to order $\log_2(t)$ differential attacks. One caveat with this view point is that actual cipher instances typically have fixed-length keys which do not grow with t – however, similar to prior works on analyzing simpler properties of block ciphers, and in particular expected differential probabilities, we promote the heuristic angle that properties which are true for independent keys (possibly, unconditionally) remain true (computationally) when these keys are derived via a suitable key-scheduling algorithms from a short, single key.[2]

We note that existing bounds on differential probabilities for ciphers such as AES *could* imply pairwise independence, if good enough, but unfortunately, the current state of the art (cf. e.g. [PSLL03]) proves *upper* bounds of the order 2^{-111} for 128-bit outputs which does not imply anything about (almost) pairwise independence. Without a finer grained understanding of the difference distribution, this *could* well imply a large distance of pairs of outputs from the uniform distribution.

[1] Such an "independence assumption" is common across block cipher analyses. For more, see Section 1.2.

[2] We note that the impact of key-schedules on cryptographic attacks is mostly not well understood.

Scope: Substitution-Permutation Networks. Our focus in this paper is on concrete block cipher designs (which likely benefit from other security properties, such as resilience to algebraic attacks), and in particular *Substitution-Permutation Networks* (SPNs), a class for which AES is a special instance, and a generalization thereof called *Key-Alternating Ciphers* (KACs). SPNs alternate *computationally simple* rounds as follows, starting from the state being equal to the block cipher input:

1. A key-mixing step which consists of XORing the keys bit-wise with the current state;
2. A *local* non-linear step where each bit of the output depends only on a few bits of the input; Concretely, this proceeds by partitioning the n-bit state into k b-bit blocks, and applying a non-linear permutation $S : \{0,1\}^b \to \{0,1\}^b$ (a so-called "S-box") to each block in parallel;
3. A linear *mixing step* is then applied to the state.

We will refer to k as the *width* and to the important special case where $b = n$ (i.e., $k = 1$) as a *Key-Alternating Cipher* (or KAC, for short). (For this case, we can omit the mixing step without loss of generality.) Most modern ciphers are SPNs (or KACs). For example, AES uses an S-box obtained from the patched inverse $x \mapsto x^{2^b-2}$ and a mixing layer alternating two simple operations (ShiftRows and MixColumns). The MiMC cipher [AGR+16] is a KAC applying the permutation $x \mapsto x^3$ to its state.

A similar viewpoint to ours was already taken by Vaudenay's *decorrelation theory* [Vau03], but we are unaware of any application of decorrelation to SPNs with concrete S-boxes. (In fact, this was left as an open problem.) Similarly, Hoory *et al.* [HMMR05] also suggested the use and analysis of t-wise independence, but the resulting constructions, while very elegant and simple, are far from existing practical designs, and better fit in the general *theoretical* pursuit of building t-wise independent permutations [KNR05, BH08, AL13].

Our Program. This raises the following questions: If we take t-wise independence as our security goal, what are good choices for the non-linear (resp. linear) step? Which choices *provably* work and which do not? Again, we stress that our goal is to find concrete, fixed choices of these layers, without modeling the S-box as a random permutation oracle.

Our results come in two forms:

1. Results about concrete SPN instantiations of SPNs with S-boxes such as the patched inversion function, and where we prove *pairwise* independence of the resulting construction. In particular, one of our results applies to the round structure of AES, without any simplifications or idealized assumptions.
2. Existential results, which hold for most choices of P, where we prove almost t-wise independence for KACs with a number of rounds that grows with t.

Next, we provide a detailed overview of our results, and the underlying techniques. Then, we give an overview of the most relevant related work.

1.1 Our Results and Techniques

This section gives an overview of our results, and the underlying techniques.

Pairwise independence of SPNs. Our first result deals with SPNs of width k with a concrete S-box $S : \mathbb{F}_{2^b} \to \mathbb{F}_{2^b}$ (thus, $n = b \cdot k$ is the block size here). In particular we focus on the case where the S-box is $S(x) = x^{-1}$ (patched so that $0^{-1} = 0$), though the results extend to other S-boxes. Our main theorem here can be cast as follows.

> **Theorem (Informal).** For a suitably instantiated mixing layer,[3] and as long as $\frac{2k+8}{2^b} + \sqrt{k/2^b} < \frac{1}{2}$, the r-round SPN with S-box $S(x) = x^{-1}$ of width k is δ-close to pairwise independent for sufficiently large $r = r(\delta)$. In particular, if $\frac{2k+8}{2^b} + \sqrt{k/2^b} = C/2$, then $r = O(\frac{\log(1/\delta)}{\log(1/C)})$.

We briefly highlight the main ideas behind the proof and note that we will focus in particular on showing that a *three* round SPN is $O(\sqrt{k/2^b})$-close to pairwise independent – this result will rely on a new extraction lemma, which we explain below. We then resort to an amplification result by Maurer, Pietrzak, and Renner [MPR07] to conclude that the $(r/3)$-fold sequential composition of the SPN is δ-close to pairwise independent, as desired.

Our analysis of the output distribution of a three-round SPN for any two distinct inputs $x \neq x'$ will take the standard (and essentially equivalent) approach of studying the distribution of the *difference* of the outputs of the two evaluation. To this end, we start with a (fixed) input difference $\Delta = x \oplus x' \neq 0^n$. Then, our first step is to show, using (mostly) algebraic properties of the field $\mathbb{F}_{2^b}$, that after ignoring some corner cases that happen with probability no more than $O(k/2^b)$, the input differences to the third round – denoted by $V_1, \ldots, V_k$ – satisfy jointly a very strong distributional property, namely:

$$\text{any subset of them of size } k' \leq k \text{ has (jointly) min-entropy at least } k'(b-1).$$

For this to true, we only need mild assumptions on the linear mixing layer. We merely require it to be described by a full-rank $k \times k$ matrix whose entries are all non-zero.

To understand the effect of the third round, at last, we resort to our extraction lemma – we want to show in particular that the distribution of the differences $Z_1, \ldots, Z_k$, which we obtain after applying the final round of S-boxes with input differences $V_1, \ldots, V_k$, is very close to uniform.[4] Imagine first that the differences Z_i are not sampled via the S-box, but rather each Z_i is sampled independently from the $(n-1)$-dimensional sub-space orthogonal to $\{0^b, V_i\}$. (We interpret the latter as a linear subspace of $\mathbb{F}_2^b$, and V_i as a vector in this space.) Our extraction lemma shows that in this case, the Z_i's are very close to uniform – the proof uses Fourier-analytic techniques.

Of course, the Z_i's are not sampled this way – by applying the S-boxes to inputs with differences V_i – yet, the key insight is that this is almost equivalent to our sub-space representation, in that by applying a lemma of Nyberg [Nyb93] we can show that there exist permutations π, π' such that $\pi'(Z_i)$ is $O(k/2^b)$ close to a random vector sampled orthogonal to $\{0^b, \pi(V_i)\}$.

We also give a proof of a weaker bound for a two-round SPN of order $\sqrt{2^{k-b}}$. This bound could be interesting in some parameter regimes.

[3] Our requirement is very mild, and is in particular implied by having maximum branch number, as it is in the case in many SPN analysis.

[4] The last mixing stage does not affect the argument – it will merely preserve uniformity by virtue of being a permutation.

The AES case. Unfortunately, we cannot apply the above theorem directly to the AES round structure or the AES parameters. First off, the AES S-box combines the inverse with a $\mathbb{F}_2$-affine function – it turns out this is not particularly difficult to handle (the affine function can be cast as part of the mixing). But we encounter other problems, in that the mixing layers does not satisfy the assumptions needed for the theorem to work, and the theorem does not apply when $k = 16$ and $b = 8$. Still, we can adapt our techniques to obtain a refined analysis which tells us that *six* AES rounds (with independent sub-keys) are ϵ-close to pairwise independent, for some $\epsilon < 1/2$. Then, using the MPR result in the iteration, we obtain the following result:

> **Theorem.** $6r$-round AES is $2^{r-1}(0.472)^r$-close to pairwise independence.

The bound is likely far from tight, as we expect much better, but non-trivial further work seems required to obtain a substantial improvement. However, we do stress that barring the use of independent keys (which again, are common in analyses of expected differential probabilities for AES), this theorem applies to the *actual* AES structure.

The width-1 case. The special case with $k = 1$ is interesting, and can be derived from our general result above. For pedagogical reason, we give a direct proof which shows that after *two* rounds we obtain already $(5/2^b)$-closeness to pairwise independent. Interestingly, we also conclude that the construction is *not* 4-wise independent when using the inverse S-box, which makes it likely not to be a good block cipher candidate. For lack of space, these are both discussed in Appendices A and B.

Existential Results. All of the above results are about pairwise independence. It is interesting to extend them to t-wise independence for $t \geq 3$. While we leave this important question open for SPNs and concrete S-boxes, we investigate the general question whether (almost) t-wise independent constructions exist in the first place.

To this end, we employ the probabilistic method to show that there exist permutations to instantiate a $(t + 1)$-round key-alternating cipher so that it is (almost) t-wise independent. We stress that while our probabilistic argument picks such permutations at random to show their existence, these permutations can then be *fixed*.

Our probabilistic argument is quite involved and requires the study of martingale sequences and their concentration. Our result follows by showing two new lemmas, and employing a careful alternation between them. The first is an *independence amplification lemma* that shows how to take a KAC that is very close to t-wise independent and by adding an additional round, obtain a KAC that is somewhat close to a $t + 1$-wise independent distribution. The second is a *distance amplification lemma* that shows how to get from a somewhat close to t-wise independent KAC to a very close to t-wise independent KAC, again by adding one round.

1.2 Perspectives and Open Problems

On Independent Keys and Other Such "Ideal" Assumptions. We remark that, to date, *all* analyses of block ciphers make ideal assumptions such as the independence of round keys and/or ideal components. For example, analyses of (iterated) Even-Mansour ciphers assume that both

the construction and the adversary have *oracle access* to a random permutation P, and that P remains unqueried on an exponential number of points. This is a *highly idealized model*: a random permutation would take exponentially many bits to write down, and indeed, in the real world, P is instantiated with a concrete permutation. The proofs say nothing about what happens to the pseudorandomness of such a cipher when P is instantiated with any concrete permutation. And moreover, analyses of multi-round constructions *all* assume independent keys.

In contrast, our work continues a research program that aims to avoid such "oracle access" assumptions. This line of work, which has its roots in the work of Nyberg in the 1990s, treats the component permutations and mixing functions as concrete functions (indeed, ones that are used in block ciphers such as AES and MiMC). While proving computational pseudorandomness is way out of reach, this line of research aims to understand the security of these constructions against concrete practical attacks.

The "independent round keys" assumption is very common and rooted in the model of Markov Ciphers of Lai, Massey, and Murphy [LMM91], and adopted by Nyberg [Nyb93] and follow-up works. The expectation is that t-wise independence becomes t-wise pseudorandomness with an appropriate instantiation of the key schedule; nevertheless, understanding the precise role of key schedules is an important open problem.

On Algebraic and Other Attacks. The research program we undertake is to study several classes of concrete, powerful, attacks against block ciphers. In particular, t-wise independence rules out an important attack vector, but the program does not stop at just t-wise independence. In particular, the two outstanding open problems that come of this work are (a) to prove t-wise independence of multi-round AES with independent round keys, for $t > 2$; and (b) to formalize and prove security against algebraic attacks. We view solving these problems as an important quest that will likely require importing analytic techniques from mathematics and TCS, as well as inventing new ones.

On Differential Attacks vs. Almost Pairwise Independence. We note that meaningful differential probabilities need to be *very* close to 2^{-n}, or else, they do not rule out distinguishers. For example, in the case of AES-128, a 2^{-127} bound on the expected differential probability (see Section 2 for the definition) does not rule out the first bit of the output being always the same as the first bit of the input. In this case, there is a distinguisher that always works!

We note that our analysis can make the statistical distance as small as we want with sufficiently many rounds, and in particular, make the differential probabilities arbitrarily close to the ideal 2^{-128}. We note that ours is the first such result; in particular, our result for AES is the first such optimal bound for the AES design. Showing a tighter tradeoff between the number of rounds and the statistical distance is an interesting open question. Showing a direct bound on the differential probability without going through statistical distance would be interesting as well.

1.3 Related Work

Coppersmith and Grossman [CG75] and Kaliski, Rivest and Sherman [JRS88] analyzed the groups generated by transition functions of the DES block cipher. [CVS09] show that the group generated

by the round functions of a cipher similar to AES is the alternating group. On the other hand, [MPW94] provide a cautionary tale where guarantees on the group generated by the round functions does not guarantee security.

Bounds on Linear and Differential Probabilities. There is an extensive body of literature on *provable* bounds for linear [MY92] and differential cryptanalysis [BS91] of block ciphers. We note that while sufficiently strong bounds on the differential probability – say $(1+\epsilon)2^{-n}$ for block size n and $\epsilon = o(1)$ - would imply almost pairwise independence, these works fall short of proving such strong guarantees.

Adopting the formal framework of Lai, Massey, and Murphy [LMM91], Nyberg and Knudsen [NK95] prove bounds on the differential probability for Feistel ciphers as a function of the underlying non-linear function. Several works have been devoted to studying the differential properties of fixed functions to instantiate these results – relevant to this work, [Nyb93] is the first work to show properties of differentials of the inverse permutation $x \mapsto x^{-1}$ in a finite field (these were later revisited by Daemen and Rijmen [DR06]). We also refer to [BN15] for a comprehensive survey on the progress in designing non-linear functions suitable for cryptography.

Much effort has also been devoted to provable bounds on linear and differential probabilities for AES and (more abstractly) SPNs. Hong et al. [HLL+00] gave the first analysis of two-round SPNs where the mixing layer has optimal branch number. This result was further generalized to arbitrary branch number by Kang et al. [KHL+02]. Very concrete bounds for the specific case of AES were then given via refined methods in several works [KMT01b, KMT01a, PSC+02, PSLL03, Kel04, KS07]. The best known result here shows that the maximum expected differential probability is at most 1.144×2^{-111} for four rounds of AES. Miles and Viola [MV15] also provide generic bounds (i.e., these bounds only depend on the S-box and the number of rounds) for linear and differential attacks against multi-round SPNs – however, the quality of their bounds decreases with a higher number of rounds.

Baignères and Vaudenay [BV05] proved optimal resilience to differential cryptanalysis whenever the S-boxes are chosen uniformly at random and secret (i.e., their description is part of the key). Later, Miles and Viola [MV15] improves this result (implicitly) by showing that SPNs with random S-boxes are effectively a pseudorandom function when the number of queries is smaller than the input size of the S-box.

Stronger Differentials. Strong notions of differential attacks have been proposed. For example, Lai [Lai94] introduced the notion of *higher order differentials*, which consider the k-th derivative (as opposed to the simple derivative of a function), whereas Knudsen [Knu94] introduced *truncated differentials*, which only consider a subset of the bits of the output. We note that security against k-th order differential cryptanalysis is implied by the k-wise independence, whereas pairwise independence implies resistance to truncated differential cryptanalysis. Another attack technique introduced by Knudsen is that of "impossible differential attacks" [Knu98], which leverage differences which occur with probability 0 – once again, sufficiently strong pairwise independence implicitly guarantees that differences occur with sufficiently large probability.

Decorrelation theory. Vaudenay [Vau03] takes a similar position to ours, proving properties of block cipher constructions on a bounded number of inputs, and inferring a number of properties

from these statements. The work also naturally exploits a natural connection with t-wise independence, like ours. Interestingly, Vaudenay considers a number of different distance measures for the resulting distributions, and use their properties to derive a number of results. However, we are not aware of any use of decorrelation theory about the security of SPNs or KACs with concrete permutations. Still, it would be interesting to considering distance measures from decorrelation theory in the context of our paper to improve tightness.

Analyses with Public Ideal Permutations. A substantial body of works considers analyses in models where the rounds of a KAC are (public) random permutation $P : \{0,1\}^n \to \{0,1\}^n$ given to the adversary. In particular, since the adversary is query-bounded, she cannot obtain the entire truth table of P and therefore, this is an idealized model. (This model is effectively capturing *generic* attacks that treat these components as a black box.) Increasingly tighter bounds for security as a pseudorandom permutation have been developed by several works [BKL+12, Ste12, LS14, CS14, HT16] which assume the permutations *and* the keys are independent. Other works consider identical permutations and/or identical keys [CLL+14, WYCD20]. The model was also considered to prove the stronger version of *indifferentiability* for key-alternating ciphers (cf.]e.g. [ABD+13, GL15a, GL15b]).

The model was then adapted to SPNs by assuming that the individual S-boxes are public random permutations $\{0,1\}^b \to \{0,1\}^b$ [DSSL16, DKS+17, CDK+18, CL18]. Crucially, these results assume that the number of queries to the S-box is smaller than 2^b, which is rather unrealistic for small values of b (e.g., $b = 8$ as in AES).

2 Preliminaries

Notational Conventions. When n is a positive integer, let $[n]$ denote the set $\{1, 2, \ldots, n\}$. When p is a prime or prime power, let $\mathbb{F}_p$ denote the finite field of size p. The logarithm function log uses base 2 by default. Probability distributions are typically denoted by calligraphic letters, e.g., $\mathcal{D}$. Sampling an element from $\mathcal{D}$ is denoted by $d \leftarrow \mathcal{D}$. For any finite set S, sampling x uniformly from S is denoted by $x \leftarrow S$.

Definition 2.1 (Entropy). *For a distribution over domain Ω whose probability mass function is p.*

- *Its Shannon entropy is* $H(p) = -\sum_{x\in\Omega} p(x)\log(p(x))$.
- *Its Min-entropy is* $H_\infty(p) = -\log\big(\max_{x\in\Omega} p(x)\big)$.
- *Its Rényi entropy of order 2, also known as the collision entropy, is* $H_2(p) = -\log\big(\sum_{x\in\Omega} p^2(x)\big)$.

Lemma 2.2. *For any random variable X with probability mass function p, we have $2^{-H_2[X]} = \|p\|_2^2 \le 2^{-H_\infty[X]}$.*

Proof. For any $0 \le u \le 2^{-H_\infty[X]}$, we have $u^2 \le u \cdot 2^{-H_\infty[X]}$. Thus

$$2^{-H_2[X]} = \|p\|_2^2 = \sum_x p^2(x) \le \sum_x p(x) \cdot 2^{-H_\infty[X]} = 2^{-H_\infty[X]}. \qquad \square$$

2.1 Almost t-wise Independent Permutations and Cryptanalysis

We review notions of almost t-wise independence, and state some connections with standard notions from the cryptanalytic literature.

Definition 2.3. *The statistical distance (or total variation distance) between two probability distributions* p *and* q *with domain* Ω *is* $d_{TV}(p, q) := \frac{1}{2} \cdot \sum_{x \in \Omega} |p(x) - q(x)|$.

Proposition 2.4. *The following is a different formula to compute the total variation distance.*

$$d_{TV}(p, q) := \sum_{x \in \Omega : p(x) > q(x)} p(x) - q(x) \ .$$

For a two argument function $F : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^\ell$ we often write $F_K(x) = F(K, x)$, and refer to F as a *function family*. (Alternatively, we use the set notation $\mathcal{F} = \{F_K\}_{K \in \{0,1\}^m}$ whenever more convenient.) We will be considering mostly *permutation* families, where $\ell = n$, and F_K is one-to-one for each K.

Definition 2.5 (close to t-wise independence)**.** *We say that a permutation family* $F : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n$ *is* ϵ-close to t-wise independent *if for all* distinct $x_1, \ldots, x_t \in \{0,1\}^n$*, and a uniformly random* m*-bit string* K*, the distribution of* $(F_K(x_1), \ldots, F_K(x_t))$ *has statistical distance at most* ϵ *from that of* t *uniformly sampled* distinct n*-bit values (i.e., sampled* without *repetition).*

We will use the following amplification lemma, which is due to Maurer, Pietrzak, and Renner [MPR07].

Lemma 2.6 (MPR Amplification Lemma)**.** *Let* F *and* G *be* ϵ*- and* δ*-close to* t*-wise independent permutation families. Then, the permutation family* $F \circ G$ *such that* $(F \circ G)_{K_1 \| K_2}(x) = F_{K_1}(G_{K_2}(x))$ *is* $2\epsilon\delta$*-close to* t*-wise independent.*

In particular, this implies that the permutation family F^r obtained by sequential r-fold composition of an ϵ-close to t-wise independent permutation family F is $2^{r-1}\epsilon^r$-close to t-wise independent. We point out that for a meaningful application of this lemma, we require that $\epsilon < 1/2$.

Differential and linear cryptanalysis. For a permutation family $F : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n$, we define the *expected differential probability* (EDP) for a given pair Δ and Δ' of non-zero input- and output-differences, as

$$\mathrm{EDP}_F(\Delta, \Delta') = \Pr_{K,X}[F_K(X \oplus \Delta) \oplus F_K(X) = \Delta'] \ ,$$

where K and X are independent and uniformly distributed over the m-bit and n-bit strings, respectively. We also define $\mathrm{MEDP}_F = \max_{\Delta, \Delta' \neq 0} \mathrm{EDP}_F(\Delta, \Delta')$. It is easy to see that if F is ϵ-close to pairwise independent, then $\mathrm{MEDP}_F \leq \epsilon + \frac{1}{2^n - 1}$. We note that a similar result extends to any subset of n output bits, and hence to so-called *truncated* differential probabilities.

We note that higher-order differential cryptanalysis [Lai94, Knu94] generalizes differential cryptanalysis to look at higher order derivatives. It is not hard to see that almost t-wise independence

will imply resistance to order-$\log_2 t$ differential cryptanalysis, as the property relies on the evaluation of the cipher on at most t inputs. We note that while (almost) t-wise independence refers to attacks that look at an arbitrary set of t inputs, an order-$\log_2 t$ differential attack looks at all inputs that lie in some $\log_2 t$-dimensional hypercube, so a total of t inputs but they are not arbitrary.

Linear cryptanalysis. The connection between pairwise independence and linear cryptanalysis is slightly less obvious. In particular, linear cryptanalysis considers the correlation

$$\mathrm{CORR}_F(\Delta, \Delta') = \mathop{\mathbb{E}}_K \left[\left(2 \cdot \mathop{\Pr}_X [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle] - 1 \right)^2 \right] ,$$

where X, K are uniform for any non-zero Δ, Δ', and where $\langle \cdot, \cdot \rangle$ denotes inner product mod two. Note that

$$\mathop{\mathbb{E}}_K \left[\left(2 \cdot \mathop{\Pr}_X [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle] - 1 \right)^2 \right] =$$
$$4 \mathop{\mathbb{E}}_K \left[\mathop{\Pr}_X [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle]^2 \right] - 4 \mathop{\mathbb{E}}_K \left[\mathop{\Pr}_X [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle] \right] + 1 .$$

But then, if F is ϵ-close to pairwise independent,

$$\mathop{\mathbb{E}}_K \left[\mathop{\Pr}_X [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle]^2 \right]$$
$$= \mathop{\Pr}_{K,X,X'} [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle \wedge \langle \Delta, X' \rangle = \langle \Delta', F_K(X') \rangle] \leq \frac{1}{2^n} + \left(\frac{1}{4} + \epsilon \right) ,$$

and also $\mathbb{E}_K \left[\Pr_X [\langle \Delta, X \rangle = \langle \Delta', F_K(X) \rangle] \right] \geq 1/2 - \epsilon$. Thus, $\mathrm{CORR}_F(\Delta, \Delta') \leq 8\epsilon + \frac{4}{2^n}$.

2.2 Key-Alternating Ciphers and Substitution Permutation Networks

A *Key Alternating Cipher* (KAC) (cf. Figure 1) is parameterized by a block size n, number of rounds r, and a fixed permutation $P : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$. A KAC is a family of functions indexed by $r+1$ sub-keys $K_0, K_1, \ldots, K_r$, and defined recursively as follows:

$$F_P^{(0)}(x) = x \oplus K_0$$
$$F_{P,K_0,\ldots,K_i}^{(i)}(x) = P(F_{P,K_0,\ldots,K_{i-1}}^{(i-1)}(x)) \oplus K_i .$$

The family of functions is $\mathcal{F}_P := \left\{ F_{P,K_0,\ldots,K_r}^{(r)}(x) : K_i \in \mathbb{F}_2^n \right\}$. One can also naturally extend this to have different permutations in each round.

A *Substitution-Permutation Network* (SPN) (cf. Figure 2) can be seen as a special case of a KAC, where $n = k \cdot b$ (we refer to k as the *width*), and the permutation P is obtained from an S-box $S : \mathbb{F}_{2^b} \to \mathbb{F}_{2^b}$ and a linear *mixing layer*, described by a matrix $M \in \mathbb{F}_{2^b}^{k \times k}$. In particular, P splits its input x into k b-bit blocks $x_1, \ldots, x_k$, and computes first $y_i = S(x_i)$ for each i, and finally outputs $M \cdot (y_1, \ldots, y_k)$. One can of course instead think of a KAC as a special of an SPN with width $k = 1$.

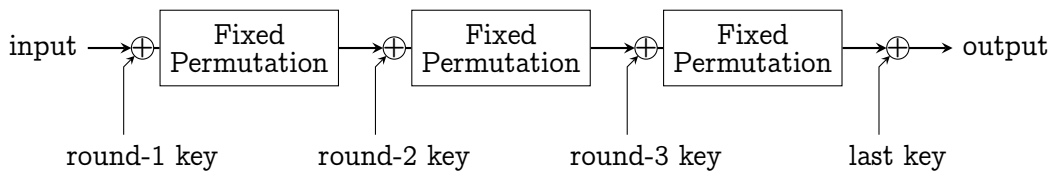

Figure 1: Illustration of Key Alternating Cipher

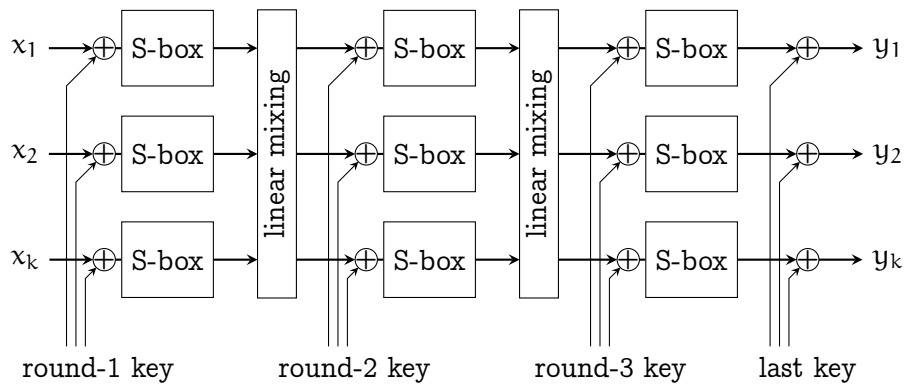

Figure 2: Illustration of Substitution Permutation Network

A fact that we will use repeatedly is that in order to bound how close to pairwise independent an SPN or KAC is, it is enough to analyze the distribution of the non-zero difference of outputs of the SPN/KAC, and its distance from the uniform distribution over non-zero strings.

Analyzing Pairwise Independence of KACs and SPNs. We will use the following lemma to reduce the analysis of pairwise independence to analyzing the distribution of differences.

Lemma 2.7. *Assume that the KAC (resp. SPN) $\mathcal{F}_P$ (resp. $\mathcal{F}_{P,M}$) has the property that for any input difference $\Delta \neq 0$, the distribution of*

$$\Delta' := F_K(x) \oplus F_K(x \oplus \Delta)$$

is ϵ-close to uniform (where the randomness of the distribution is taken over x and K). Then, the KAC (resp. SPN) is ϵ-close to pairwise independent.

Proof. Take any two distinct inputs x_1, x_2, and distinct y_1, y_2, let $\Delta = x_1 \oplus x_2$ and $\Delta' = y_1 \oplus y_2$. Then,

$$\begin{aligned}
\Pr_K\left[F_K(x_1) = y_1, F_K(x_2) = y_2\right] &= \Pr_K\left[F_K(x_1) = y_1 \wedge F_K(x_1) \oplus F_K(x_2) = y_1 \oplus y_2\right] \\
&= \frac{1}{2^n} \Pr_K\left[F_K(x_1) \oplus F_K(x_2) = y_1 \oplus y_2\right] \\
&= \frac{1}{2^n} \Pr_{K,x}\left[F_K(x) \oplus F_K(x \oplus \Delta) = \Delta'\right] ,
\end{aligned}$$

where the middle inequality follows from the fact that $F_K(x_1) = y_1$ occurs with probability $1/2^n$ even condition on $F_K(x_1) \oplus F_K(x_2) = \Delta'$, as the latter event is independent of the final round key(s). The latter inequality follows from the fact that adding a uniform input shift does not change the input distributions.Thus, the statistical distance d_{TV} from uniform is

$$\begin{aligned} d_{TV} &= \frac{1}{2} \sum_{y_1 \neq y_2} \left| \Pr\left[F_K(x_1) = y_1, F_K(x_2) = y_2\right] - \frac{1}{2^n(2^n-1)} \right| \\ &= \frac{1}{2 \cdot 2^n} \sum_{y_1 \neq y_2} \left| \Pr_{K,x}\left[F_K(x) \oplus F_K(x \oplus \Delta) = y_1 \oplus y_2\right] - \frac{1}{2^n - 1} \right| \\ &= \frac{1}{2} \sum_{\Delta' \neq 0} \left| \Pr_{K,x}\left[F_K(x) \oplus F_K(x \oplus \Delta) = \Delta'\right] - \frac{1}{2^n - 1} \right| \leq \epsilon \ . \end{aligned}$$

□

Advanced Encryption Standard. The mostly widely used block cipher is the world is Advanced Encryption Standard (AES), which is based on the SPN framework. The block size is 128 bits, width is 16, i.e. $n = 128, k = 16, b = 8$. AES is a family of ciphers which have 10, 12 or 14 rounds.

The S-box is instantiated by $S(x) = A(x^{2^8-2})$, where $x \mapsto x^{2^8-2}$ is the patched inverse function over $\mathbb{F}_{2^8}$, A is an invertible affine function over $\mathbb{F}_2^8$. The exact form of A is irrelevant for this paper (as shown by Lemma 3.4).

The linear mixing function is instantiated by the composition of ShiftRows and MixColumns. The domain of AES is typically represented as a 4×4 matrix of bytes. ShiftRows permutes bytes, it shifts the i-th row cyclically to the left by an offset of $i - 1$ coordinates.

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} & x_{16} \end{bmatrix} \xrightarrow{\text{ShiftRows}} \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_6 & x_7 & x_8 & x_5 \\ x_{11} & x_{12} & x_9 & x_{10} \\ x_{16} & x_{13} & x_{14} & x_{15} \end{bmatrix}$$

MixColumns acts on each column. In each column, the four bytes is mixed by an invertible linear transformation. All coefficients in the linear transformation are non-zero.

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} & x_{16} \end{bmatrix} \xrightarrow{\text{MixColumns}} \begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 3 \\ 3 & 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} & x_{16} \end{bmatrix}$$

2.3 Trace in Fields of Characteristic Two

We describe a number of facts related to the finite field $\mathbb{F}_{2^n}$ of characteristic 2 and the trace function over it. For proofs of the claims below, we refer the reader to any standard text on the subject, e.g. [LN86].

Definition 2.8. *The trace function* $\mathrm{Tr} : \mathbb{F}_{2^n} \to \mathbb{F}_2$ *is defined as* $\mathrm{Tr}(x) = \sum_{i=0}^{n-1} x^{2^i}$.

Lemma 2.9. *For every* $x \in \mathbb{F}_{2^n}$, $\mathsf{Tr}(x^2) = \mathsf{Tr}(x)$.

Lemma 2.10. *For every* $x, y \in \mathbb{F}_{2^n}$, $\mathsf{Tr}(x+y) = \mathsf{Tr}(x)+\mathsf{Tr}(y)$. *In particular, the set of elements* $x \in \mathbb{F}_{2^n}$ *with* $\mathsf{Tr}(x) = 0$ *form an* $\mathbb{F}_2$*-subspace of dimension* $n-1$.

Lemma 2.11. *Let* $\alpha \in \mathbb{F}_{2^n}$. *The equation* $y(y \oplus 1) = \alpha$ *over* $\mathbb{F}_{2^n}$ *has two solutions if* $\mathsf{Tr}(\alpha) = 0$ *and no solutions otherwise.*

Corollary 2.12. *Let* $a, b, c \in \mathbb{F}_{2^n}$ *and* a, b *are non-zero. The equation* $ax^2 + bx + c = 0$ *has two solutions over* $\mathbb{F}_{2^n}$ *if* $\mathsf{Tr}(ac/b^2) = 0$ *and no solutions otherwise.*

Proof. Define $y = ax/b$, by a change of variable, the equation becomes $y^2 + y + ac/b^2 = 0$. □

Lemma 2.13. *For every* $x \neq y \in \mathbb{F}_{2^n}$, *let* $S_x := \{z : \mathsf{Tr}(xz) = 0\}$ *and* $S_y := \{z : \mathsf{Tr}(yz) = 0\}$. *Then,* $S_x \neq S_y$. *Indeed, since these are* $(n-1)$*-dimensional subspaces, they intersect at exactly* 2^{n-2} *elements.*

Proof. Let $\alpha \in \mathbb{F}_{2^n}$ be such that $\mathsf{Tr}(\alpha) = 1$. Then, we know that $\beta := \alpha \cdot (x \oplus y)^{-1}$ cannot be in both sets S_x and S_y. For, if it were, we would have

$$0 = \mathsf{Tr}(\beta x) \oplus \mathsf{Tr}(\beta y) = \mathsf{Tr}(\beta(x \oplus y)) = \mathsf{Tr}(\alpha) = 1$$

giving us the necessary contradiction. □

We also need the following Lemma from Nyberg's work [Nyb93], which we reprove for completeness.

Lemma 2.14 ([Nyb93])**.** *Let* $P : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ *be the patched inversion function* $P(x) = x^{2^n-2}$. *For every* $\delta, \gamma \neq 0$, *let* $p_{\delta,\gamma} := \Pr_{x \leftarrow \mathbb{F}_{2^n}}[P(x) \oplus P(x \oplus \delta) = \gamma]$. *Then, for even* n,

$$p_{\delta,\gamma} = \begin{cases} 4/2^n & \text{if } \delta\gamma = 1 \\ 2/2^n & \text{if } \delta\gamma \neq 1 \text{ and } \mathsf{Tr}((\delta\gamma)^{-1}) = 0 \\ 0 & \text{if } \delta\gamma \neq 1 \text{ and } \mathsf{Tr}((\delta\gamma)^{-1}) = 1 \end{cases}$$

and for odd n,

$$p_{\delta,\gamma} = \begin{cases} 2/2^n & \text{if } \delta\gamma = 1 \text{ or } \mathsf{Tr}((\delta\gamma)^{-1}) = 0 \\ 0 & \text{if } \delta\gamma \neq 1 \text{ and } \mathsf{Tr}((\delta\gamma)^{-1}) = 1 \end{cases}$$

Proof. First, $x = 0$ and $x = \delta$ are both solutions to this equation if and only if $\delta\gamma = 1$. If $x \neq 0, \delta$, we need

$$x^{-1} \oplus (x \oplus \delta)^{-1} = \delta \cdot (x(x \oplus \delta))^{-1} = \gamma.$$

By Corollary 2.12, this has two solutions if $\mathsf{Tr}((\delta\gamma)^{-1}) = 0$ and no solutions otherwise.

To summarize, when n is even. If $\delta\gamma = 1$, we have four solutions: two solutions (namely $x = 0$ and $x = \delta$) and two additional solutions since $\mathsf{Tr}(1) = 0$ (which happens if and only if n is even). If $\delta\gamma \neq 1$, the claim in the lemma follows by the discussion above.

To summarize when n is odd: if $\delta\gamma = 1$, we have two solutions (namely $x = 0$ and $x = \delta$) and no more since $\mathsf{Tr}(1) = 1$. If $\delta\gamma \neq 1$, we have two solutions when $\mathsf{Tr}(\delta) = 0$ and none when $\mathsf{Tr}(\delta) = 1$. □

The following corollary is an immediate consequence.

Corollary 2.15. *For any non-zero $\delta \in \mathbb{F}_{2^n}$, let*

$$p(\gamma) := \Pr_{x \leftarrow \mathbb{F}_{2^n}} [P(x) \oplus P(x \oplus \delta) = \gamma] \ .$$

Let $\mathcal{D}_\delta$ denote the distribution with probability mass function p and let $\mathcal{D}'_\delta$ denote the distribution with probability mass function $p'(\gamma) = p(\gamma^{-1})$, we have:

- *$\mathcal{D}'_\delta$ is $(2/2^b)$-close to the uniform distribution on a subspace of dimension $b-1$.*
- $H_2(\mathcal{D}_\delta) \geq -\log_2\left(\frac{2}{2^b} + \frac{8}{2^{2b}}\right)$.

2.4 Basics of Discrete Fourier Analysis

The characters of the group $\mathbb{F}_2^n$ are functions $\{\chi_x : \mathbb{F}_2^n \to \mathbb{R}\}_{x \in \mathbb{F}_2^n}$ defined by

$$\chi_x(y) = (-1)^{\langle x, y \rangle}$$

The functions $\{\chi_x\}_{x \in \mathbb{F}_2^n}$ are orthonormal under the inner product[5]

$$\langle \chi_x, \chi_{x'} \rangle := \frac{1}{2^n} \sum_{y \in \mathbb{F}_2^n} \chi_x(y) \chi_{x'}(y) \ .$$

Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be a real-valued function on $\mathbb{F}_2^n$. Writing $f = \sum_{x \in \mathbb{F}_2^n} \widehat{f}(x) \chi_x$, we have the Fourier (inversion) formulas

$$f(y) = \sum_{x \in \mathbb{F}_2^n} \widehat{f}(x) \chi_x(y) \quad \text{and} \quad \widehat{f}(x) = \langle f, \chi_x \rangle = \frac{1}{2^n} \sum_{y \in \mathbb{F}_2^n} f(y) \chi_x(y)$$

We need the following two facts. For proofs, we refer the reader to [O'D14].

Lemma 2.16 (Parseval's Theorem). $\frac{1}{2^n} \sum_{y \in \mathbb{F}_2^n} f(y)^2 = \sum_{x \in \mathbb{F}_2^n} \widehat{f}(x)^2$.

If S is a subspace of $\mathbb{F}_2^n$, let $S^\perp = \{y : \langle x, y \rangle = 0 \text{ for all } x \in S\}$ denote its dual subspace. If S is k-dimensional, $S^\perp$ is $(n-k)$-dimensional.

Lemma 2.17. *Let $S \subseteq \mathbb{F}_2^n$ be a subspace and f_S denote the uniform probability distribution on S. That is,*

$$f_S(y) = \begin{cases} \frac{1}{|S|} & \text{if } y \in S \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$\widehat{f}_S(x) = \begin{cases} \frac{1}{2^n} & \text{if } x \in S^\perp \\ 0 & \text{otherwise} \end{cases}$$

[5] Note that there are two inner products at play here, one over $\mathbb{F}_2^n$ and the other over $\mathbb{R}^{2^n}$, and we are abusing notation by denoting them both as $\langle \cdot, \cdot \rangle$.

In particular, let $S \subseteq \mathbb{F}_2^n$ be an $(n-1)$-dimensional subspace which can equivalently be denoted as (the dual subspace) $S = \{0, v\}^\perp$ for some $v \in \mathbb{F}_2^n$. Thenby Lemma 2.17,

$$\hat{f}_S(y) = \begin{cases} \frac{1}{2^n}, & \text{if } y \in \{0, v\} \\ 0, & \text{otherwise} \end{cases}$$

Let $f : \mathbb{F}_2^n \to \mathbb{R}, g : \mathbb{F}_2^{n'} \to \mathbb{R}$ be two real-valued functions on $\mathbb{F}_2^n$ and $\mathbb{F}_2^{n'}$ respectively. Their tensor product $f \otimes g : \mathbb{F}_2^{n+n'} \to \mathbb{R}$ is a real-valued function on $\mathbb{F}_2^{n+n'}$ such that

$$(f \otimes g)(x, y) := f(x) \cdot g(y) \text{ for all } x \in \mathbb{F}_2^n, y \in \mathbb{F}_2^{n'}.$$

Assume X, Y are two independent random variables on $\mathbb{F}_2^n$ and $\mathbb{F}_2^{n'}$ respectively, and f, g are the probability mass functions of X, Y. Then $f \otimes g$ is the probability mass function of (X, Y), as

$$\Pr[(X, Y) = (x, y)] = \Pr[X = x] \cdot \Pr[Y = y] = f(x) \cdot g(y) = (f \otimes g)(x, y).$$

The Fourier transform of the tensor equals the tensor of the Fourier transforms.

Lemma 2.18 (Fourier transform of a Tensor). *For any* $f : \mathbb{F}_2^n \to \mathbb{R}, g : \mathbb{F}_2^{n'} \to \mathbb{R}$, $\widehat{f \otimes g} = \hat{f} \otimes \hat{g}$.

2.5 Tail Bounds

Theorem 2.19 (Bernstein's Bound). *Let* $X_1, \ldots, X_N$ *be a sequence of independent bounded random variables such that* $\left|X_i - \mathbb{E}[X_i]\right| \leq b$ *for all* $i \in [N]$. *Let* $X = \sum_{i=1}^N X_i$, *then*

$$\Pr\left[X - \mathbb{E}[X] > \tau\right] \leq \exp\left(-\frac{\frac{1}{2}\tau^2}{\mathrm{Var}[X] + \frac{1}{3}b\tau}\right),$$
$$\Pr\left[X - \mathbb{E}[X] < -\tau\right] \leq \exp\left(-\frac{\frac{1}{2}\tau^2}{\mathrm{Var}[X] + \frac{1}{3}b\tau}\right).$$

Theorem 2.20 (Martingale version of Bernstein's Bound [CL06]). *Let* $X_1, \ldots, X_N$ *be real value random variables such that*

- *There exists a sequence of random variables* $\mathfrak{F}_1, \ldots, \mathfrak{F}_N$, *called filters*
- $\mathfrak{F}_i$ *determines* X_i *and* $\mathfrak{F}_{i-1}$
- $\mathrm{Var}\left[X_i \mid \mathfrak{F}_{i-1}\right] \leq \sigma_i^2$, *for any* $i \in [N]$
- $\mathbb{E}\left[X_i \mid \mathfrak{F}_{i-1}\right] = 0$, *for any* $i \in [N]$
- $X_i \leq b$ *for all* $i \in [N]$.

Then, we have

$$\Pr\left[\sum_{i=1}^N X_i > \tau\right] \leq \exp\left(-\frac{\frac{1}{2}\tau^2}{\sum_{i=1}^N \sigma_i^2 + \frac{1}{3}b\tau}\right).$$

Corollary 2.21. *Let* $X_1, \ldots, X_N$ *be real value random variables such that*

- *There exists a sequence of random variables* $\mathfrak{F}_1, \ldots, \mathfrak{F}_N$*, called filters*
- $\mathfrak{F}_i$ *determines* X_i *and* $\mathfrak{F}_{i-1}$
- $\mathrm{Var}\big[X_i \mid \mathfrak{F}_{i-1}\big] \leq \sigma_i^2$*, for any* $i \in [N]$
- $\mathbb{E}\big[X_i \mid \mathfrak{F}_{i-1}\big] \leq a_i$*, for any* $i \in [N]$
- $X_i \leq a_i + b$ *for all* $i \in [N]$*.*

Then, we have

$$\Pr\Big[\sum_{i=1}^{N} X_i - \sum_{i=1}^{N} a_i > \tau\Big] \leq \exp\Big(-\frac{\frac{1}{2}\tau^2}{\sum_{i=1}^{N} \sigma_i^2 + \frac{1}{3}b\tau}\Big).$$

3 Pairwise Independence of SPNs

The main result of this section is a proof of pairwise independence of the 3-round substitution-permutation network (see Figure 2) where the non-linear S-box is the patched inverse function over $\mathbb{F}_{2^n}$, used in the AES block cipher. We will show that the 3-round SPN is ϵ-close to pairwise independent for a constant $\epsilon < 1/2$, and note that an application of the MPR amplification lemma (Lemma 2.6) gives us $2^{-\Omega(r)}$-closeness to pairwise independence in $3r$ rounds.

In Section 3.1, we start with our main technical result, an *S-box extraction lemma*, which says that when the input difference of a single round of SPN has *sufficient Rényi entropy*, the output difference is close to uniformly random. We follow this up by describing mixing functions and their properties in Section 3.2. In Section 3.3, we then use the S-box extraction lemma and properties of mixing functions to show our main result, namely the pairwise independence of 3-round SPN. The reader is encouraged to refer back to Section 2.4 for relevant facts about discrete Fourier analysis as and when necessary.

3.1 The S-box Extraction Lemma

Before we state the S-box extraction lemma, we describe how it will be used to show the pairwise independence of SPNs. As noted in Lemma 2.7, it is sufficient to show that the distribution of *output differences* on any two inputs is close to uniformly random.

Consider the scenario in the last round of a substitution-permutation network, as illustrated in Figure 3. Before the last round, we will show that the input difference already has high (Rényi) entropy. Indeed, we will show that if there is one round of S-boxes and mixing before the last round, Δ_i has large entropy for any $i \in [k]$; and if there are two rounds of S-boxes and mixing before the last round, the joint distribution of $(\Delta_1, \ldots, \Delta_k)$ has (proportionally) high entropy. The question we ask then is, is the output (difference) vector $(\Delta'_1, \ldots, \Delta'_k)$ close to uniform? The extraction lemma provides an affirmative answer to this question.

Lemma 3.1 (The S-Box Extraction Lemma)**.** *Let* k, b *be positive integers and* $n = bk$*. Let* $\mathcal{D}$ *be a distribution over* $(\mathbb{F}_2^b)^k$ *and consider the following probabilistic process called* $\mathsf{Samp}_{\mathcal{D}}$*.*

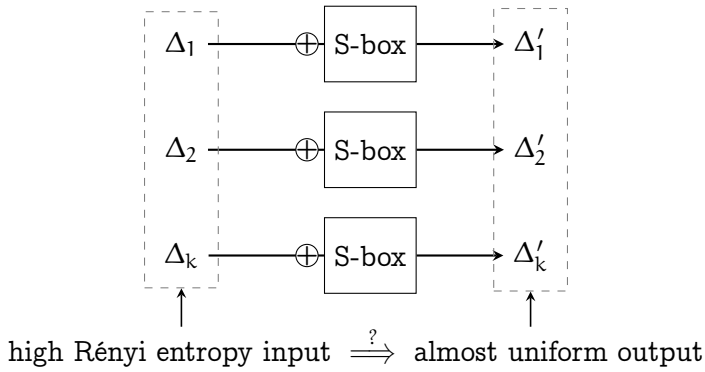

Figure 3: Application Scenario of the Extraction Lemma

1. *Sample* $(v_1, \ldots, v_k) \leftarrow \mathcal{D}$. *Let* $S_1, \ldots, S_k$ *be* $(b-1)$*-dimensional subspaces where each* $S_i = \{0, v_i\}^{\perp}$ *is the subspace orthogonal to* v_i.

2. *For each* $i \in [k]$, *sample* $x_i \leftarrow S_i$ *independently at random, and output* $(x_1, \ldots, x_k)$.

For any $T \subseteq [k]$, *let* v_T *denote the concatenation of* $(v_i)_{i \in T}$, *let* $\mathcal{D}_T$ *denote the distribution of* v_T, *let* $H_2[\mathcal{D}_T]$ *denote its Rényi entropy. Then, the statistical distance between the joint distribution of* $(x_1, \ldots, x_k)$ *and the uniform distribution over* $\mathbb{F}_2^{bk}$ *is at most*

$$\frac{1}{2}\sqrt{\sum_{T \subseteq [k], T \neq \varnothing} 2^{-H_2[\mathcal{D}_T]}}\ .$$

In particular, we have:

- **Weak Extraction:** *Assume that for all* $i \in [k]$, $H_2[v_i] \geq h$ *for a fixed real* $h \leq b$. *Then the statistical distance between the joint distribution of* $(x_1, \ldots, x_k)$ *and the uniform distribution over* $\mathbb{F}_2^{bk}$ *is at most* $\frac{1}{2} \cdot \sqrt{\frac{2^k - 1}{2^h}}$.

- **Strong Extraction:** *Assume that for any* $T \subseteq [k]$, $H_2[v_T] \geq h \cdot |T|$ *where* v_T *denotes the concatenation of* $(v_i)_{i \in T}$. *Then the statistical distance between the joint distribution of* $(x_1, \ldots, x_k)$ *and the uniform distribution over* $\mathbb{F}_2^{bk}$ *is at most*

$$\frac{1}{2} \cdot \sqrt{\left(1 + \frac{1}{2^h}\right)^k - 1}$$

  which, in turn, is at most $\sqrt{\frac{k}{2^{h+1}}}$ *assuming* $k \leq 2^h$.

Proof. Let f denote the probability mass function of $\mathsf{Samp}_{\mathcal{D}}$. That is, $f(x_1, \ldots, x_k)$ is the probability that $\mathsf{Samp}_{\mathcal{D}}$ outputs $(x_1, \ldots, x_k)$. Let $p(v_1, \ldots, v_k)$ denote the probability assigned by the distribution $\mathcal{D}$ to $(v_1, \ldots, v_k)$ and let ϕ_S denote the probability mass function of the uniform distribution over the subspace $S \subseteq \mathbb{F}_2^b$. Then,

$$f(x_1, \ldots, x_k) = \sum_{v_1, \ldots, v_k \in \mathbb{F}_2^b} p(v_1, \ldots, v_k) \cdot \phi_{S_1}(x_1) \cdot \phi_{S_2}(x_2) \cdot \ldots \cdot \phi_{S_k}(x_k)$$

where $S_i = \{0, v_i\}^\perp$ is an implicit function of v_i, as before. We will write this as

$$f = \sum_{v_1,\ldots,v_k \in \mathbb{F}_2^b} p(v_1,\ldots,v_k) \cdot \left(\phi_{S_1} \otimes \phi_{S_2} \otimes \ldots \otimes \phi_{S_k}\right)$$

We are interested in the statistical distance $d_{TV}(f,u) = \frac{1}{2}\|f-u\|_1$, where u is the uniform distribution over $\mathbb{F}_2^{bk}$. It suffices to bound $\|\hat{f}-\hat{u}\|_2^2$ since

$$\|f-u\|_1^2 \le 2^{kb}\|f-u\|_2^2 = 2^{2kb}\|\hat{f}-\hat{u}\|_2^2. \tag{1}$$

where the inequality comes from Cauchy-Schwartz and the equality comes from Parseval's theorem (Lemma 2.16).

The Fourier transform of f equals

$$\hat{f}(y_1,\ldots,y_k) = \sum_{v_1,\ldots,v_k \in \mathbb{F}_2^b} p_{(v_1,\ldots,v_k)} \cdot \prod_{i\in[k]} \hat{\phi}_{S_i}(y_i)$$

Observe that by Lemma 2.17, $\hat{\phi}_{S_i}$ is 0 everywhere except for $\hat{\phi}_{S_i}(v_i) = \hat{\phi}_{S_i}(0) = 1/2^b$. Thus the only inputs $(y_1,\ldots,y_k)$ on which $\hat{f}(y_1,\ldots,y_k) \neq 0$ are those in the set $\{0,v_1\}\times\{0,v_2\}\times\ldots\times\{0,v_k\}$. Thus,

$$\hat{f}(y_1,\ldots,y_k) = \frac{1}{2^{bk}} \cdot \Pr[v_i = y_i \text{ for all } i \text{ s.t. } y_i \neq 0]. \tag{2}$$

The ℓ_2-norm of the Fourier transform of $f-u$ can then be computed as

$$\begin{aligned}
\left\|\hat{f}-\hat{u}\right\|_2^2 &= \sum_{\substack{y_1,\ldots,y_k \in \mathbb{F}_2^b \\ (y_1,\ldots,y_k)\neq\vec{0}}} \hat{f}^2(y_1,\ldots,y_k) \\
&= \sum_{\substack{T\subseteq[k] \\ T\neq\varnothing}} \sum_{\substack{y_1,\ldots,y_k \in \mathbb{F}_2^b \\ y_i\neq 0 \text{ iff } i\in T}} \hat{f}^2(y_1,\ldots,y_k) \\
&= \sum_{\substack{T\subseteq[k] \\ T\neq\varnothing}} \sum_{\substack{y_1,\ldots,y_k \in \mathbb{F}_2^b \\ y_i\neq 0 \text{ iff } i\in T}} \frac{1}{2^{2bk}} \cdot \Pr[v_i = y_i \text{ for all } i \in T]^2.
\end{aligned} \tag{3}$$

Let $v_T := (v_i)_{i\in T}$ denote the vector v restricted to indices in T, let $\mathcal{D}_T$ denote the distribution of v_T, and let f_T denote the probability mass function of $\mathcal{D}_T$. Then,[6]

$$\left\|\hat{f}-\hat{u}\right\|_2^2 \le \frac{1}{2^{2bk}} \sum_{\substack{T\subseteq[k] \\ T\neq\varnothing}} \|f_T\|_2^2 = \frac{1}{2^{2kb}} \sum_{\substack{T\subseteq[k] \\ T\neq\varnothing}} 2^{-H_2[\mathcal{D}_T]}. \tag{4}$$

Combining with equation (1) concludes the proof of the general case.

$$d_{TV}(f,u) \le \frac{1}{2} \cdot 2^{kb} \cdot \|\hat{f}-\hat{u}\|_2 \le \frac{1}{2}\sqrt{\sum_{\substack{T\subseteq[k] \\ T\neq\varnothing}} 2^{-H_2[\mathcal{D}_T]}}. \tag{5}$$

[6] The first inequality symbol in the equation is tight, if $V_1,\ldots,V_k$ are always non-zero.

Setting 1: Weak Extraction. Assume for any $i \in [k]$, $H_2[\mathcal{D}_{\{i\}}] \geq h$. Then, for any non-empty set $T \subseteq [k]$, we have $H_2[\mathcal{D}_T] \geq h$. Therefore, combining with equation (5),

$$d_{TV}(f, u) \leq \frac{1}{2}\sqrt{\sum_{\substack{T \subseteq [k] \\ T \neq \emptyset}} 2^{-H_2[\mathcal{D}_T]}} \leq \frac{1}{2} \cdot \sqrt{\frac{2^k - 1}{2^h}}.$$

Setting 2: Strong Extraction. Assume for any $T \subseteq [k]$, $H_2[\mathcal{D}_T] \geq h \cdot |T|$. Then

$$\sum_{\substack{T \subseteq [k] \\ T \neq \emptyset}} 2^{-H_2[\mathcal{D}_T]} \leq \sum_{\substack{T \subseteq [k] \\ T \neq \emptyset}} \left(\frac{1}{2^h}\right)^{|T|} = \left(1 + \frac{1}{2^h}\right)^k - 1$$

using the binomial expansion. Combining with equation (5), we have

$$d_{TV}(f, u) \leq \frac{1}{2}\sqrt{\sum_{\substack{T \subseteq [k] \\ T \neq \emptyset}} 2^{-H_2[\mathcal{D}_T]}} \leq \frac{1}{2} \cdot \sqrt{\left(1 + \frac{1}{2^h}\right)^k - 1}.$$

If we additionally assume that $k \leq 2^h$, then

$$d_{TV}(f, u) \leq \frac{1}{2} \cdot \sqrt{\left(1 + \frac{1}{2^h}\right)^k - 1} \leq \frac{1}{2}\sqrt{e^{\frac{k}{2^h}} - 1} \leq \frac{1}{2}\sqrt{\frac{2k}{2^h}}.$$

The last inequality symbol holds only if $\frac{k}{2^h} \leq 1.256\ldots$, which follows from the condition $k \leq 2^h$. □

We remark that Fourier analysis can be bypassed here. The above proof uses Fourier analysis to bound the collision probability. There is an alternative proof of the extraction lemma in Section C that bounds the collision probability using "elementary" non-Fourier methods.

Comparing Figure 3 with the statement of the extraction lemma. The outstanding contrast is that the extraction lemma assumes a very specific linear algebra structure. That is, consider the domain as vector space $\mathbb{F}_2^b$, the output (difference) vector is sampled as a random vector orthogonal to the input (difference) vector. While in each round of SPN, the input is subtracted by the random key and then feed into the S-box. The output difference *is not* sampled uniformly from a subspace.

However, we hope the two can be bridged by change of variables. Say we start with two inputs differing Δ, let Δ' denote the difference after key-subtraction and S-box. We hope there exist 1-to-1 mappings $\pi_{in}, \pi_{out} : \mathbb{F}_{2^b} \to \mathbb{F}_2^b$ such that $\pi_{out}(\Delta')$ is a random vector orthogonal to $\pi_{in}(\Delta)$.

Figure 4 illustrates the property we are looking for. Although it cannot be exactly satisfied by any S-box — we know $\pi_{out}(\Delta')$ doesn't equal x by distribution, because $\Delta = 0 \iff \Delta' = 0$ — we show that pragmatic S-boxes almost satisfy the property.

Assuming the S-box is the patched inverse function, the following lemma shows that $\pi_{out}(\Delta')$ is statistically close to a random vector orthogonal to $\pi_{in}(\Delta)$, as long as $\Delta \neq 0$.

Lemma 3.2. *Assume S-box is the patched inverse* $P(x) = x^{2^b - 2}$. *There exist 1-to-1 mappings* $\pi_{in}, \pi_{out} : \mathbb{F}_{2^b} \to \mathbb{F}_2^b$ *such that for any non-zero* $\Delta \in \mathbb{F}_{2^b}$, *letting* Δ' *denotes a random variable defined by*

$$\Delta' := P(r) - P(r + \Delta)$$

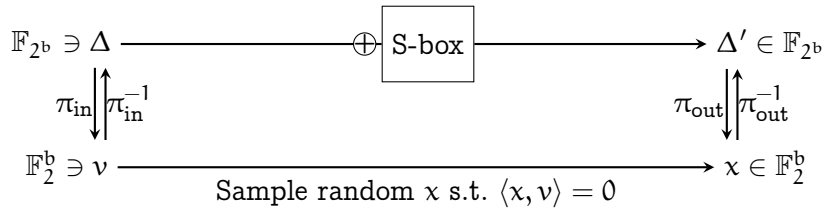

Figure 4: Subtracting Key followed by S-box $\approx$ Subspace Sampling, modulo Change of Variables

for a uniformly random $r \in \mathbb{F}_{2^b}$*, the statistical distance between* $\pi_{out}(\Delta')$ *and the uniform distribution over* $\{0, \pi_{in}(\Delta)\}^{\perp}$ *is no more than* $\frac{2}{2^b}$*.*

Proof. As shown in Lemma 2.14 (from [Nyb93]),

$$\Pr[\Delta' = \delta] = \begin{cases} \frac{2}{2^b}, & \text{if } \delta = \frac{1}{\Delta} \\ 0, & \text{o.w.} \end{cases} + \begin{cases} \frac{2}{2^b}, & \text{if } \mathrm{Tr}(\frac{1}{\delta\Delta}) = 0 \\ 0, & \text{o.w.} \end{cases}$$

Define $\pi_{out}(x) = x^{2^b-2}$ to be the patched inverse as well. Then

$$\Pr[\pi_{out}(\Delta') = x] = \begin{cases} \frac{2}{2^b}, & \text{if } x = \Delta \\ 0, & \text{o.w.} \end{cases} + \begin{cases} \frac{2}{2^b}, & \text{if } x \neq 0 \text{ and } \mathrm{Tr}(\frac{x}{\Delta}) = 0 \\ 0, & \text{o.w.} \end{cases}$$

As show in Lemma 2.10, $x \mapsto \mathrm{Tr}(\frac{x}{\Delta})$ is linear function over $\mathbb{F}_2$. Define $\pi_{in}(\Delta)$ as the coefficient vector of $x \mapsto \mathrm{Tr}(\frac{x}{\Delta})$. Then

$$\Pr[\pi_{out}(\Delta') = x] = \begin{cases} \frac{2}{2^b}, & \text{if } x = \Delta \\ 0, & \text{o.w.} \end{cases} + \begin{cases} \frac{2}{2^b}, & \text{if } x \neq 0 \text{ and } \langle \pi_{in}(\Delta), x\rangle = 0 \\ 0, & \text{o.w.} \end{cases}$$

Apparently, the statistical distance between $\pi_{out}(\Delta')$ and the uniform distance over $\{0, \pi_{in}(\Delta)\}^{\perp}$ is $\frac{2}{2^b}$. □

The following lemma shows the analogous statement for the cube function. The proof is deferred to the full version [LTV21].

Lemma 3.3. *Assume S-box is the cube function* $P(x) = x^3$ *over* $\mathbb{F}_{2^b}$ *where* b *is odd*[7]*. There exist 1-to-1 mappings* $\pi_{in}, \pi_{out} : \mathbb{F}_{2^b} \to \mathbb{F}_2^b$ *such that for any non-zero* $\Delta \in \mathbb{F}_{2^b}$*, letting* Δ' *denote a random variable defined by*

$$\Delta' := P(r) - P(r + \Delta)$$

for a uniformly random $r \in \mathbb{F}_{2^b}$*,* $\pi_{out}(\Delta')$ *is the uniform distribution over* $\{0, \pi_{in}(\Delta)\}^{\perp}$*.*

[7] The condition on b being odd is necessary to ensure that P is a permutation.

Proof. We want to compute the probability

$$\Pr[\Delta' = \delta] = \Pr_{r \leftarrow \mathbb{F}_{2^b}}[r^3 + (r+\Delta)^3 = \delta].$$

The equation can be rewritten as $\Delta r^2 + \Delta^2 r + \Delta^3 + \delta = 0$, a quadratic equation in r. By Corollary 2.12, this has two solutions for r when

$$\mathsf{Tr}(\Delta(\Delta^3+\delta)/\Delta^4) = \mathsf{Tr}(1) + \mathsf{Tr}(\delta/\Delta^3) = 0,$$

and no solution otherwise. Note that $\mathsf{Tr}(1) = 0$ if b is odd, therefore,

$$\Pr[\Delta' = \delta] = \begin{cases} \frac{2}{2^b}, & \text{if } \mathsf{Tr}(\delta/\Delta^3) = 0; \\ 0, & \text{otherwise.} \end{cases}$$

As show in Lemma 2.10, $x \mapsto \mathsf{Tr}(x/\Delta^3)$ is linear function over $\mathbb{F}_2$. Define $\pi_{\text{in}}(\Delta)$ as the coefficient vector of $x \mapsto \mathsf{Tr}(x/\Delta^3)$, define π_{out} as the identical function. Then

$$\Pr[\pi_{\text{out}}(\Delta') = x] = \begin{cases} \frac{2}{2^b}, & \text{if } \langle \pi_{\text{in}}(\Delta), x \rangle = 0 \\ 0, & \text{o.w.} \end{cases}$$

In other words, $\pi_{\text{out}}(\Delta')$ is the uniform distribution over $\{0, \pi_{\text{in}}(\Delta)\}^{\perp}$. □

In Section 3.4 we are going to analyze AES. The S-box in AES is called Rijndael S-box, which is not exactly the patched inverse function. Rijndael S-box is the composition of the patched inverse function and an affine transformation. The following lemma shows that the additional affine transformation makes little difference.

Lemma 3.4. *Assume S-box is $P(x) = A(x^{2^b-2})$, where A is an affine permutation over $\mathbb{F}_2^b$. There exist 1-to-1 mappings $\pi_{\text{in}}, \pi_{\text{out}} : \mathbb{F}_{2^b} \to \mathbb{F}_2^b$. For any non-zero $\Delta \in \mathbb{F}_{2^b}$, let Δ' denote a random variable defined by*

$$\Delta' := P(r) - P(r+\Delta)$$

for a uniformly random $r \in \mathbb{F}_{2^b}$. The statistical distance between $\pi_{\text{out}}(\Delta')$ and the uniform distribution over $\{0, \pi_{\text{in}}(\Delta)\}^{\perp}$ is no more than $\frac{2}{2^b}$.

Proof. As we are analyzing the differences, any additive constant in the affine function A has no effect. Thus we can safely assume A is a linear permutation.

When input difference is Δ, the output difference is

$$\Delta' = P(r) - P(r+\Delta) = A(r^{2^b-2}) - A((r+\Delta)^{2^b-2}) = A(r^{2^b-2} - (r+\Delta)^{2^b-2}).$$

Define $\Delta^* = r^{2^b-2} - (r+\Delta)^{2^b-2}$, then $\Delta' = A(\Delta^*)$.

Lemma 3.2 shows that there exists $\pi_{\text{in}}, \pi_{\text{out}}$ such that $\pi_{\text{out}}(\Delta^*)$ is close to uniform distribution over $\{0, \pi_{\text{in}}(\Delta)\}$. Define $\pi'_{\text{out}}(x) := \pi_{\text{out}}(A^{-1}(x))$. Then $\pi'_{\text{out}}(\Delta') = \pi_{\text{out}}(A^{-1}(\Delta')) = \pi_{\text{out}}(\Delta^*)$, which is close to uniform distribution over $\{0, \pi_{\text{in}}(\Delta)\}$. Thus $\pi_{\text{in}}, \pi'_{\text{out}}$ are what we need. □

3.2 Properties of Mixing Functions

Before proceeding to show the almost-pairwise independence of SPN constructions using the extraction lemma, we describe properties that we need the mixing functions to satisfy. We define two such properties below and prove some elementary statements about them.

The first property that we call *diffusion* requires that if one of the input blocks of the (typically linear) function $M : (\mathbb{F}_{2^b})^k \to (\mathbb{F}_{2^b})^k$ has sufficient entropy and the distribution of the k input blocks are independent, then *each output block* has large entropy. It is not hard to see that both the sufficient entropy condition and the independence condition on the input are necessary for such a statement to be true. Looking ahead, this property will turn out to be useful in the first layer (or the first few layers) of the SPN where we wish to propagate differences in one input block to differences in all of them.

Property 1 (Diffusion). *Let* $M : (\mathbb{F}_{2^b})^k \to (\mathbb{F}_{2^b})^k$ *be a function. Let* $H_\alpha \in \{H_2, H_\infty\}$ *be an entropy function.*[8] *Let* $X_1, \ldots, X_k$ *be independent random variables over* $\mathbb{F}$ *such that there exists an* i *for which* $H_\alpha(X_i) \geq h$ *for a real* h*, and let* $(Y_1, \ldots, Y_k) := M(X_1, \ldots, X_k)$*.* M *is* diffusing *if*

$$\textit{for all } i \in [k],\ H_\alpha(Y_i) \geq h.$$

We now show a sufficient condition for a function to be diffusing.

Lemma 3.5. *If* $M \in (\mathbb{F}_{2^b})^{k \times k}$ *is a matrix with no zero entry, the linear mapping* $x \mapsto Mx$ *is diffusing* (i.e. satisfies Property 1).

Proof. Without loss of generality, assume that $H_\alpha(X_1) \geq h$. Then for any $j \in [k]$, Y_j can be written as

$$Y_j = m_{j,1}X_1 + \Big(\sum_{2 \leq j' \leq k} m_{j,j'}X_{j'}\Big)$$

Observe that X_1 has entropy at least h bits conditioned on any fixed values of $X_{j'} = x_{j'}$ (due to independence of the input blocks). Thus

$$H_\alpha(Y_i) \geq H_\alpha(m_{i,1}X_1) = H_\alpha(X_1). \qquad \square$$

The second property that we call *entropy-preservation* requires that if all of the input blocks of the (typically linear) function $M : (\mathbb{F}_{2^b})^k \to (\mathbb{F}_{2^b})^k$ have sufficient entropy and the distribution of the k blocks are independent, then *each collection of output blocks* have large joint entropy. Looking ahead, this property will turn out to be useful in the subsequent layers of the SPN to ensure that the mixing layers do not *reduce* the entropy. As one might expect, this property comes for free if M is an invertible linear map.

Property 2 (Entropy Preservation). *A function* $M : (\mathbb{F}_{2^b})^k \to (\mathbb{F}_{2^b})^k$ *is* entropy preserving *if for any entropy function* $H_\alpha \in \{H_2, H_\infty\}$*, for any real* h*, for any independent random variables*

[8] We only consider min-entropy and Rényi entropy in this work. While the results can be extended to other entropy functions as well.

$X_1, \ldots, X_k$ *over* $\mathbb{F}_{2^b}$ *such that* $H_\alpha(X_i) \geq h$ *for all* $i \in [k]$*, letting* $(Y_1, \ldots, Y_k) := M(X_1, \ldots, X_k)$*, we have*

$$H_\alpha(Y_{i_1}, \ldots, Y_{i_s}) \geq s \cdot h$$

for any $\{i_1, \ldots, i_s\} \subseteq [k]$*.*

Lemma 3.6. *If* $M \in (\mathbb{F}_{2^b})^{k \times k}$ *is an invertible matrix, the mapping* $x \mapsto Mx$ *is entropy-preserving* (i.e. satisfies Property 2)*.*

Proof. W.l.o.g., it suffices to show $(Y_1, \ldots, Y_s)$ has large min-entropy (or Rényi entropy) for any $s \in [k]$.

Let $M_{1:s,1:k}$ denotes the sub-matrix of M that consists the first s rows. Then $(Y_1, \ldots, Y_s) = M_{1:s,1:k}(X_1, \ldots, X_k)$. Because M is full-rank, its sub-matrix $M_{1:s,1:k}$ has full row-rank. Thus $M_{1:s,1:k}$ must has an $s \times s$ full-rank sub-matrix. Without loss of generality, assume the first s columns of $M_{1:s,1:k}$ form a full-rank matrix, denoted by $M_{1:s,1:s}$.

$$(Y_1, \ldots, Y_s) = M_{1:s,1:s}(X_1, \ldots, X_s) + M_{1:s,s+1:k}(X_{s+1}, \ldots, X_k).$$

Thus $H_\alpha(Y_1, \ldots, Y_s) \geq H_\alpha(M_{1:s,1:s}(X_1, \ldots, X_s)) = H_\alpha(X_1, \ldots, X_s) \geq s \cdot h$. □

Connection to Branch Number. The branch number of a matrix $M \in (\mathbb{F}_{2^b})^{k \times k}$ is defined to be

$$\mathsf{br}(M) = \max_{\alpha \in (\mathbb{F}_{2^b})^k} (\mathsf{wt}(\alpha) + \mathsf{wt}(M\alpha))$$

where wt denotes the Hamming weight. Having an optimal branch number is considered a desirable feature for mixing functions [Dae95, KHL+02]. An observation by Miles and Viola [MV15] says that any matrix with the maximal branching number of $k + 1$ also satisfies properties 1 and 2, although the converse does not necessarily hold.

3.3 Proofs of Pairwise Independence

In this section, we show several proofs of pairwise independence of SPNs using the patched inverse function $P(x) = x^{2^b-2}$ over the finite field $\mathbb{F}_{2^b}$. The first result (Theorem 3.7) applies in a regime where $k \leq b$ is relatively small; here, the result says that a 2-round SPN is close to pairwise independent. The second result (Theorem 3.8) is much more general and applies to large k as long as $k \leq 2^{b-4}$; here, the result says that a 3-round SPN is close to pairwise independent.

Theorem 3.7. *Assume the S-box is* $P(x) = x^{2^b-2}$ *over* $\mathbb{F}_{2^b}$ *assume the mixing function is diffusing, that is, it satisfies Property 1. Then a 2-round SPN with* k *blocks each of which has* b *bits is* ϵ*-close to 2-wise independent where*

$$\epsilon \leq \frac{2 + 4k}{2^b} + \sqrt{\frac{2^k - 1}{2^{b+1}}}.$$

Proof. Naming the variables as in Figure 5, fix any set of input differences $\Delta_{1,1}, \ldots, \Delta_{1,k}$ which are not all zero. We wish to show that the distribution of $(\Delta'_{2,1}, \Delta'_{2,2}, \ldots, \Delta'_{2,k})$ is ϵ-close to uniform. By Lemma 2.7, this implies ϵ-closeness to pairwise independence. We proceed via a hybrid argument.

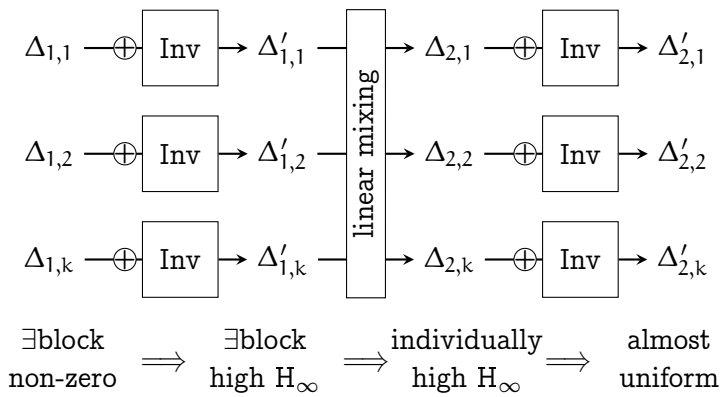

Figure 5: Illustration of the proof of Theorem 3.7. Inv denotes the patched inversion function over $\mathbb{F}_{2^b}$, that is $\mathsf{Inv}(x) = x^{2^b-2}$.

Hybrid 0. $(\Delta'_{2,1}, \Delta'_{2,2}, \ldots, \Delta'_{2,k})$ are as in the SPN construction illustrated in Figure 5.

Hybrid 1. Pick some j where $\Delta_{1,j} \neq 0$, and note that the distribution of $\Delta'_{1,j}$ is $(2/2^b)$-close to uniformly random over a subset of size 2^{b-1} (Corollary 2.15). Call this uniform distribution $\mathcal{D}'_{1,j}$. We have $H_2(\mathcal{D}'_{1,j}) \geq H_\infty(\mathcal{D}'_{1,j}) = b - 1$.

Hybrid 1 is the same as hybrid 0 except that we replace $\Delta'_{1,j}$ by a vector drawn from the distribution $\mathcal{D}'_{1,j}$. The statistical distance from Hybrid 0 is at most $\frac{2}{2^b}$.

Claim Assume that the mixing function satisfies Property 1. In Hybrid 1, $H_2[\Delta_{2,i}] \geq H_\infty[\Delta_{2,i}] \geq b - 1$ for all $i \in [k]$.

Hybrid 2. In this hybrid, we change the way $\Delta'_{2,i}$ is sampled based on $\Delta_{2,i}$. In particular:

- When $\Delta_{2,i} = \delta \neq 0$, the distribution of $\pi_{\mathsf{out}}(\Delta'_{2,i})$ conditioning on $\Delta_{2,i} = \delta$ is $\frac{2}{2^b}$-close to uniform distribution over $\{0, \pi_{\mathsf{in}}(\delta)\}^\perp$.
- When $\Delta_{2,i} = 0$, $\Delta'_{2,i}$ is chosen to be uniformly random.

Let us calculate the statistical distance between hybrids 1 and 2. The first bullet introduces a statistical distance of at most $2k/2^b$. The probability that a fixex coordinate $\Delta_{2,i}$ is 0 is at most $2/2^b$, and therefore, the probability that some coordinate is 0 is at most $2k/2^b$. In total, the statistical distance is at most $\frac{4k}{2^b}$.

By applying our extraction lemma (Lemma 3.1), we know that, in hybrid 2, the joint distribution of $\Delta'_{2,1}, \ldots, \Delta'_{2,k}$ is at most $\sqrt{\frac{2^k-1}{2^{b+1}}}$-away from uniform.

Putting everything together, the statistical distance between $(\Delta'_{2,1}, \ldots, \Delta'_{2,k})$ and the uniform distribution is at most $\frac{2+4k}{2^b} + \sqrt{\frac{2^k-1}{2^{b+1}}}$. □

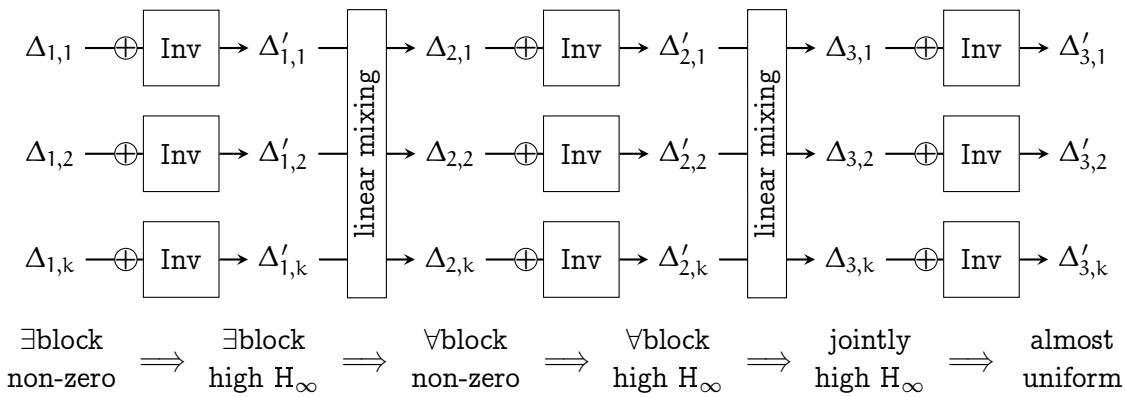

Figure 6: Illustration of the proof of Theorem 3.8 and Lemma 3.9

Theorem 3.8. *Assume the S-box is patched inverse* $P(x) = x^{2^b-2}$*, assume the mixing function satisfies Property 1 and Property 2. Then 3-round SPN is* ϵ*-close to 2-wise independent where*

$$\epsilon \leq \frac{2+8k}{2^b} + \sqrt{\frac{k}{2^b}}.$$

Proof. Name the variables as in Figure 6, fix any set of input differences $\Delta_{1,1}, \ldots, \Delta_{1,k}$ which are not all zero. We wish to show that the distribution of $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ is ϵ-close to uniform. By Lemma 2.7, this implies ϵ-closeness to pairwise independence. We proceed via a hybrid argument.

Hybrid 0. Hybrid 0 is the real world hybrid that is illustrated in Figure 6.

Hybrid 1. Pick some j where $\Delta_{1,j} \neq 0$. W.l.o.g., assume $\Delta_{1,1} \neq 0$. Note that the distribution of $\Delta'_{1,1}$ is $(2/2^b)$-close to uniformly random over a subset of size 2^{b-1} (Corollary 2.15). Call this uniform distribution $\mathcal{D}'_{1,1}$. We have $H_\infty(\mathcal{D}'_{1,1}) = b - 1$.

Hybrid 1 is the same as hybrid 0 except that we replace $\Delta'_{1,1}$ by a random sample from the distribution $\mathcal{D}'_{1,1}$. The statistical distance from Hybrid 0 is at most $\frac{2}{2^b}$.

Claim Assume that the mixing function satisfies Property 1. In Hybrid 1, $H_\infty[\Delta_{2,j}] \geq b - 1$ for all $j \in [k]$.

Hybrid 2. In this hybrid, we ensure $\Delta_{2,j} \neq 0$ for all $j \in [k]$. Formally, hybrid 2 is the same as hybrid 1 except that we replace $\Delta_{2,j}$ by 1 if $\Delta_{2,j} = 0$ in hybrid 1. The statistical distance from Hybrid 1 is at most $\frac{2k}{2^b}$.

Lemma 3.9 shows that the joint distribution of $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ is $\left(\frac{6k}{2^b} + \sqrt{\frac{k}{2^b}}\right)$-close to uniform in hybrid 2.

Putting everything together, the statistical distance between $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ and the uniform distribution is at most $\frac{2+8k}{2^b} + \sqrt{\frac{k}{2^b}}$. □

Lemma 3.9. *Assume the S-box is patched inverse* $P(x) = x^{2^b-2}$*, assume the mixing function satisfies Property 2. Starting with a pair of inputs, whose difference is entry-wise-nonzero, after a 2-round SPN, the statistical distance between the output difference and the uniform distribution is no more than* $\frac{6k}{2^b} + \sqrt{\frac{k}{2^b}}$.

Proof. Name the variables as the last two rounds in Figure 6, fix any set of input differences $\Delta_{2,1}, \ldots, \Delta_{2,k}$ which are all non-zero. We wish to show that the distribution of $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ is ϵ-close to uniform. We proceed via a hybrid argument.

Hybrid 0. Hybrid 0 is the real world hybrid.

Hybrid 1. Since $\Delta_{2,j} \neq 0$ for all $j \in [k]$, the distribution of $\Delta'_{2,j}$ is $(2/2^b)$-close to uniformly random over a subset of size 2^{b-1} (Corollary 2.15). Call this uniform distribution $\mathcal{D}'_{2,j}$. We have $H_\infty(\mathcal{D}'_{2,j}) = b - 1$.

Hybrid 1 is the same as hybrid 0 except that we replace $\Delta'_{2,j}$ by a vector drawn from the distribution $\mathcal{D}'_{2,j}$ for each $j \in [k]$. The statistical distance from Hybrid 0 is at most $\frac{2k}{2^b}$.

Claim Assume that the mixing function satisfies Property 2. In Hybrid 1, $H_\infty[\Delta_{3,j}] \geq b - 1$ for all $j \in [k]$.

Hybrid 2. In this hybrid, we change the way $\Delta'_{3,j}$ is sampled based on $\Delta_{3,j}$. In particular:

- When $\Delta_{3,j} = \delta \neq 0$, the distribution of $\pi_{\text{out}}(\Delta'_{3,j})$ conditioning on $\Delta_{3,j} = \delta$ is $\frac{2}{2^b}$-close to uniform distribution over $\{0, \pi_{\text{in}}(\delta)\}^\perp$. Let $\pi_{\text{out}}(\Delta'_{3,j})$ sampled uniformly from $\{0, \pi_{\text{in}}(\delta)\}^\perp$ in hybrid 2.
- When $\Delta_{3,j} = 0$, $\Delta'_{3,j}$ is chosen to be uniformly random in hybrid 2.

Let us calculate the statistical distance between hybrids 1 and 2. The first bullet introduces a statistical distance of at most $2k/2^b$. The probability that a fixed coordinate $\Delta_{3,j}$ is 0 is at most $2/2^b$, and therefore, the probability that some coordinate is 0 is at most $2k/2^b$. In total, the statistical distance is at most $\frac{4k}{2^b}$.

By applying our extraction lemma[9] (Lemma 3.1), we know that, in hybrid 2, the joint distribution of $\Delta'_{3,1}, \ldots, \Delta'_{3,k}$ is at most $\sqrt{\frac{k}{2^b}}$-away from uniform.

Counting them together, the statistical distance between $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ and the uniform distribution is at most $\frac{6k}{2^b} + \sqrt{\frac{k}{2^b}}$. □

[9] Our extraction lemma also requires $k \leq 2^{b-1}$. In the case $k > 2^{b-1}$, Lemma 3.9 can be trivially proved as $d_{TV} \leq 1 \leq \frac{6k}{2^b} + \sqrt{\frac{k}{2^b}}$.

3.4 AES is Almost Pairwise-Independent

Good asymptotic bounds have been shown in Theorem 3.7 and 3.8, but the analysis there is way too loose on AES parameter ($k = 16$, $b = 8$). This section emphasizes on better concrete bound. Comparing with Section 3.3, the concrete bound is improved by the following tricks.

- Lemma 3.1 shows that the statistical distance is less than $\frac{1}{2} \cdot \sqrt{\left(1 + \frac{1}{2^h}\right)^k - 1}$, which is less than $\sqrt{\frac{k}{2^{h+1}}}$. The former is tighter. In particular, when $k = 16$, $b = 8$, $h = -\log_2\left(\frac{2}{2^b} + \frac{8}{2^{2b}}\right)$, the former shows $d_{TV} \leq 0.18357\ldots \leq \frac{47}{256}$, and the latter shows $d_{TV} \leq 0.25$.

- Lemma 3.10 is the strengthening of Lemma 3.9. Besides using the tighter bound from Lemma 3.1, it also considers Rényi entropy instead of min-entropy.

- Theorem 3.12 is the strengthening of Theorem 3.8. The proof of Theorem 3.12 (resp. Theorem 3.8) shows that after two rounds of AES (resp. one round of SPN), all block differences are non-zero with high probability. Then ignoring the rare event, Lemma 3.10 (resp. Lemma 3.9) will conclude the proof.

  The proof of Theorem 3.12 also carefully analyzes the rare event that some block difference is zero after 2 rounds of AES. It observes that, given the rare event happens, after two more rounds, all block differences will be non-zero with high probability.

Lemma 3.10 (Strengthening of Lemma 3.9)**.** *Assume the S-box is patched inverse* $P(x) = x^{2^b-2}$*, assume the mixing function satisfies Property 2. Starting with a pair of inputs, whose difference is entry-wise-nonzero, after a 2-round SPN, the statistical distance between the output difference and the uniform distribution is no more than* $\frac{4k}{2^b} + \frac{1}{2}\sqrt{(1+2^{-h})^k - 1}$*, where* $h = -\log_2\left(\frac{2}{2^b} + \frac{8}{2^{2b}}\right)$.

In particular, when $k = 16$*,* $b = 8$*, we have* $d_{TV} \leq \frac{64+47}{256}$.

Proof. The proof is mostly the same of Lemma 3.9. Name the variables as the last two rounds in Figure 6, fix any set of input differences $\Delta_{2,1}, \ldots, \Delta_{2,k}$ which are all non-zero. We wish to show that the distribution of $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ is ϵ-close to uniform. We proceed via a hybrid argument.

Hybrid 0. Hybrid 0 is the real world hybrid.

The input difference $\Delta_{2,1}, \ldots, \Delta_{2,k}$ are all non-zero, therefore by Corollary 2.15, the Rényi entropy of $\Delta'_{2,i}$ is at least $h = -\log_2\left(\frac{2}{2^b} + \frac{8}{2^{2b}}\right)$ for any $i \in [k]$.

The mixing function of AES satisfies Property 2, so for any $s \leq k$ and for any $\{i_1, \ldots, i_s\} \subseteq [k]$, we have $H_2[\Delta_{3,i_1}, \ldots, \Delta_{3,i_s}] \geq s \cdot h$.

Hybrid 1. In this hybrid, we change the way $\Delta'_{3,j}$ is sampled based on $\Delta_{3,j}$. In particular:

- When $\Delta_{3,j} = \delta \neq 0$, the distribution of $\pi_{out}(\Delta'_{3,j})$ conditioning on $\Delta_{3,j} = \delta$ is $\frac{2}{2^b}$-close to uniform distribution over $\{0, \pi_{in}(\delta)\}^\perp$ in hybrid 0. Let $\pi_{out}(\Delta'_{3,j})$ sampled uniformly from $\{0, \pi_{in}(\delta)\}^\perp$ in hybrid 1.

- When $\Delta_{3,j} = 0$, $\Delta'_{3,j}$ is chosen to be uniformly random in hybrid 1.

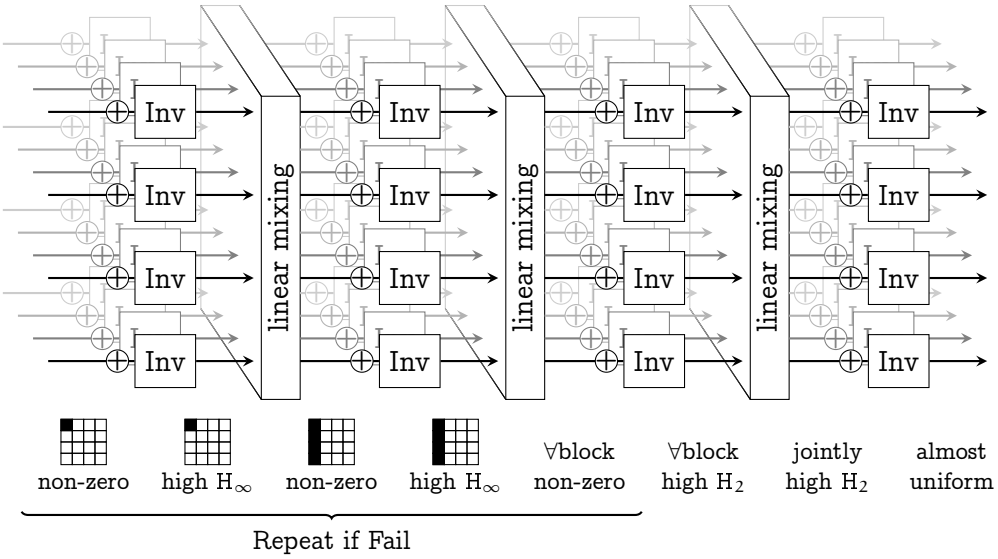

Figure 7: Illustration of the proof of Theorem 3.12 and Lemma 3.11

Let us calculate the statistical distance between hybrids 1 and 2. The first bullet introduces a statistical distance of at most $2k/2^b$. The probability that a fixed coordinate $\Delta_{3,j}$ is 0 is at most $2/2^b$, and therefore, the probability that some coordinate is 0 is at most $2k/2^b$. In total, the statistical distance is at most $\frac{4k}{2^b}$.

By applying our extraction lemma (Lemma 3.1), we know the jointly distribution of $\pi_{out}(\Delta'_{3,1}), \ldots, \pi_{out}(\Delta'_{3,k})$ is at most $\frac{1}{2}\sqrt{(1+2^{-h})^k - 1}$-away from uniform in hybrid 1.

Counting them together, the statistical distance between $(\Delta'_{3,1}, \ldots, \Delta'_{3,k})$ and the uniform distribution is at most $\frac{4k}{2^b} + \frac{1}{2}\sqrt{(1+2^{-h})^k - 1}$. □

Lemma 3.11. *Starting with a pair of distinct inputs, after 2-round of AES, including a tailing linear mixing, the output difference has zero entry with probability no more than* $\frac{25}{2^7}$.

Proof. Comparing with the analysis proving Theorem 3.8, an extra round is required so that every output block (difference) is non-zero with high probability. Because the mixing function in AES does not satisfy Property 1.

The proof is illustrated in Figure 7. Fix any set of input differences $\Delta_{1,1}, \ldots, \Delta_{1,16}$ which are not all zero. After the i-th round of S-boxes, $\Delta'_{i,j}$ is generated based on $\Delta_{i,j}$. After applying the i-th mixing, $(\Delta'_{i,1}, \ldots, \Delta'_{i,16})$ is mapped to $(\Delta_{i+1,1}, \ldots, \Delta_{i+1,16})$. We wish to show that $\Delta_{3,1}, \ldots, \Delta_{3,16}$ are not all non-zero with high probability. We proceed via a hybrid argument.

Hybrid 0. The real world hybrid.

Hybrid 1. Pick some j where $\Delta_{1,j} \neq 0$. W.l.o.g., assume $\Delta_{1,1} \neq 0$. Note that the distribution of $\Delta'_{1,1}$ is $(2/2^b)$-close to uniformly random over a subset of size 2^7 (Corollary 2.15). Call this uniform distribution $\mathcal{D}'_{1,1}$. We have $H_\infty(\mathcal{D}'_{1,1}) = 7$.

Hybrid 1 is the same as hybrid 0 except that we replace $\Delta'_{1,1}$ by a random sample from the distribution $\mathcal{D}'_{1,1}$. The statistical distance from Hybrid 0 is at most $\frac{1}{2^7}$.

Hybrid 2. The AES domain is typically represented as a 4×4 matrix of bytes. The mixing function of AES will carries the entropy of the first entry to every entry in the first column. Since $H_\infty(\Delta'_{1,1}) = 7$ in hybrid 1, after applying the mixing function, we have $H_\infty(\Delta_{2,j}) \geq 7$ for $j \in \{1, 5, 9, 13\}$ (i.e., the coordinates in).

In Hybrid 2, we ensure $\Delta_{2,j}$ to be non-zero for $j \in \{1, 5, 9, 13\}$ (i.e.,). Formally, hybrid 2 is the same as hybrid 1 except that we replace $\Delta_{2,j}$ by 1 if $\Delta_{2,j} = 0$ in hybrid 1 for each $j \in \{1, 5, 9, 13\}$. The statistical distance from Hybrid 0 is at most $\frac{4}{2^7}$.

Hybrid 3. Until further notice, the following analysis are conditioning on $(\Delta_{2,1}, \ldots, \Delta_{2,16}) = (\delta_{2,1}, \ldots, \delta_{2,16})$. The conditioning simplifies the analysis as it removes the dependency.

Note that, for each $j \in \{1, 5, 9, 13\}$ (i.e.,), $\Delta_{2,j} = \delta_{2,j} \neq 0$, thus the distribution of $\Delta'_{2,j}$ is $(2/2^b)$-close to uniformly random over a subset of size 2^7 (Corollary 2.15). Call this uniform distribution $\mathcal{D}'_{2,j}$. We have $H_\infty(\mathcal{D}'_{2,j}) = 7$.

Hybrid 3 is the same as hybrid 2 except that we replace $\Delta'_{2,j}$ by a random sample from the distribution $\mathcal{D}'_{2,j}$ for each $j \in \{1, 5, 9, 13\}$ (i.e.,). The statistical distance from Hybrid 2 is at most $\frac{4}{2^7}$.

The mixing function of AES will carries the entropy of the first entry (13-th entry, 9-th entry, 5-th entry, resp.) to every entry in the first column (second column, third column, fourth column, resp.). Since $H_\infty(\Delta'_{2,j}) = 7$ for $j \in \{1, 5, 9, 13\}$ (i.e., the coordinates in) in hybrid 3, after applying the mixing function, we have $H_\infty(\Delta_{3,j}) \geq 7$ for all $j \in [16]$. Thus, $\Pr[\exists j, \Delta_{3,j} = 0] \leq \frac{16}{2^7}$.

The analysis so far is conditioning on $(\Delta_{2,1}, \ldots, \Delta_{2,16}) = (\delta_{2,1}, \ldots, \delta_{2,16})$. By the law of total probability, we have $\Pr[\exists j, \Delta_{3,j} = 0] \leq \frac{16}{2^7}$ in hybrid 3, without conditioning.

Putting everything together, the probability that $\Delta_{3,1}, \ldots, \Delta_{3,16}$ has zero entry is no more than $\frac{1+4+4+16}{2^7}$ in the real world. □

Theorem 3.12. *6-round of AES is 0.472-close to pairwise independence.*

Proof. The proof is illustrated in Figure 7. Fix any set of input differences $\Delta_{1,1}, \ldots, \Delta_{1,16}$ which are not all zero. After the i-th round of S-boxes, $\Delta'_{i,j}$ is generated based on $\Delta_{i,j}$. After applying the i-th mixing, $(\Delta'_{i,1}, \ldots, \Delta'_{i,16})$ is mapped to $(\Delta_{i+1,1}, \ldots, \Delta_{i+1,16})$. We wish to show that $\Delta'_{6,1}, \ldots, \Delta'_{6,16}$ are not all non-zero with high probability. By Lemma 2.7, this implies 0.472-closeness to pairwise independence.

Lemma 3.11 shows that after 2-round of AES (namely, Sbox-mix-Sbox-mix), $\Delta_{3,1}, \ldots, \Delta_{3,16}$ are all non-zero with probability at least $1 - \frac{25}{2^7}$. Combining with Lemma 3.10, we have 4-round AES is $\frac{151}{256}$-close to pairwise independence. We need the statistical distance to be smaller than a half.

Let C denotes the event that $\exists j, \Delta_{3,j} = 0$. Then Lemma 3.11 shows that $\Pr[C] \leq \frac{25}{2^7}$. We consider the statistical distance between the output distance and uniform distance, conditioning on C and $\neg C$ respectively.

Conditioning on C, since there are still 4 rounds left, the conditional output distribution is $\frac{151}{256}$-close to pairwise independence.

Conditioning on $\neg C$, Lemma 3.10 shows that the (conditional) distribution of $\Delta'_{4,1}, \ldots, \Delta'_{4,16}$ is $\frac{111}{256}$-close from uniform in statistical distance. The last two rounds of AES will not increase the statistical distance from uniform.

The (unconditional) statistical distance can be bounded by the convex combination

$$d_{TV} \leq \frac{111}{256} \cdot \Pr[\neg C] + \frac{151}{256} \cdot \Pr[C] \leq 0.472. \qquad \square$$

3.5 Multi-round SPNs and AES

We now combine the bounds from Theorems 3.7, 3.8, and 3.12 with the MPR amplification lemma (Lemma 2.6) to obtain the following theorems.

Theorem 3.13. *Assume the S-box is $P(x) = x^{2^b-2}$ over $\mathbb{F}_{2^b}$ assume the mixing function is diffusing, that is, it satisfies Property 1. Then a $(2r)$-round SPN with k blocks each of which has b bits is ϵ-close to 2-wise independent where*

$$\epsilon \leq 2^{r-1} \left(\frac{2+4k}{2^b} + \sqrt{\frac{2^k-1}{2^{b+1}}} \right)^r .$$

Further, if the mixing function additionally satisfies Property 2, then $(3r)$-round SPN is ϵ-close to 2-wise independent where

$$\epsilon \leq 2^{r-1} \left(\frac{2+8k}{2^b} + \sqrt{\frac{k}{2^b}} \right)^r .$$

Theorem 3.14. *$6r$-round AES is $2^{r-1}(0.472)^r$-close to pairwise independence.*

4 t-wise Independence of KAC

In this section, we consider a key-alternating cipher whose i^{th} round consists of applying a *public, fixed* permutation p_i to the current state followed by adding a (private) round-key s_i. The main result of this section is that for every r, there exist public permutations $p_1, \ldots, p_r$ such that r rounds of KAC using these permutations gets us close to $(r - o(r))$-wise independence. We achieve a strong notion of *pointwise* closeness (see definition 4.2) much stronger than the statistical distance measures considered in previous sections. Furthermore, it is easy to see that a t-round KAC can at best be (close to) t-wise independence, due to a simple entropy argument, meaning that our result is nearly optimal and entropy-preserving.

We remark that this is an *existential result*: namely, we do not explicitly construct the fixed permutations used by the KAC, but merely show that they exist. Indeed, we show that most permutations work, as is typical of probabilistic arguments. We also remark that the permutations $p_1, \ldots, p_r$ are fixed and known to the adversary, thus the only secret randomness in the construction comes from the round keys s_i.

We start with some new notations. We encourage the reader to consult Section 2.5 for tail bounds that are extensively used in our analysis.

4.1 Definitions and Notations

Let $\mathfrak{D}$ denote the domain and let $2^n = N := |\mathfrak{D}|$. Throughout this report, we will consider many distribution of permutations over $\mathfrak{D}$. Permutation distributions will be denoted by calligraphic letters (e.g. $\mathcal{F}, \mathcal{G}, \mathcal{H}$). A random choice of a permutation from such a distribution will act as a key for the KAC. Here are two simple examples of permutation distributions:

Example 1 (Shift permutations). *Denoted by $\mathcal{S}$, the uniform distribution over*

$$\{\sigma_s : x \mapsto x + s \mid s \in \mathfrak{D}\},$$

which consists of all shift permutations σ_s that additively shifts the input by s. The definition assumes $\mathfrak{D}$ to be a group. The support of $\mathcal{S}$ is of size N.

Example 2 (Linear permutations). *Denoted by $\mathcal{L}$, the uniform distribution over*

$$\{\ell_{a,b} : x \mapsto ax + b \mid a \in \mathfrak{D} \setminus \{0\}, b \in \mathfrak{D}\},$$

which consists of all linear permutations over $\mathfrak{D}$. The definition assumes $\mathfrak{D}$ to be a field. The support of $\mathcal{L}$ is of size $N(N-1)$.

We now define a notation for composition of permutations, the cornerstone of the KAC construction.

Definition 4.1 (Composition). *Let $\mathcal{F}, \mathcal{G}$ be distributions over permutations, and let p be a permutation over $\mathfrak{D}$. Their compositions are defined as*

$$\begin{aligned}
&\mathcal{F} \circ p \text{ is the distribution of } f \circ p \text{ where } f \leftarrow \mathcal{F},\\
&p \circ \mathcal{G} \text{ is the distribution of } p \circ g \text{ where } g \leftarrow \mathcal{G},\\
&\mathcal{F} \circ \mathcal{G} \text{ is the distribution of } f \circ g \text{ where } f \leftarrow \mathcal{F}, g \leftarrow \mathcal{G} \text{ independently.}
\end{aligned}$$

As a quick example of composition, $\mathcal{S} \circ \mathcal{L} = \mathcal{L} \circ \mathcal{S} = \mathcal{L}$.

Key Alternating Cipher. Given the language of permutation distributions from above, we can give an alternative definition of key-alternating ciphers (KACs). A t-round KAC is parametered by fixed permutations $p_1, \ldots, p_{t-1}$, and is the composition

$$\mathcal{S} \circ p_1 \circ \mathcal{S} \circ p_2 \circ \mathcal{S} \circ p_3 \circ \cdots \circ p_{t-1} \circ \mathcal{S}.$$

In words, this means picking t round-keys $s_1, \ldots, s_t \leftarrow \mathfrak{D}$ and letting

$$f_{s_1,\ldots,s_t}(x) = s_t + \underbrace{p_{t-1}(s_{t-1} + p_{t-2}(s_{t-2} + \ldots))}_{\text{repeated } t-1 \text{ times}}$$

as illustrated in Figure 1.

Pointwise Closeness to t-wise Independence. Finally, we define the notion of being pointwise close to t-wise independent which we achieve. The lemma that follows asserts that this is a stronger notion than being close to t-wise independent (Definition 2.5), a notion that we worked with in Section 4. This only makes the results of this section stronger.

Definition 4.2 (pointwise close to t-wise independence)**.** *Let $\mathcal{F}$ be a distribution over permutations. $\mathcal{F}$ is pointwise ϵ-close to t-wise independence if for any distinct $x_1, \ldots, x_t \in \mathfrak{D}$ and any distinct $y_1, \ldots, y_t \in \mathfrak{D}$,*

$$\Pr_{f \leftarrow \mathcal{F}}\Big[f(x_1) = y_1 \wedge f(x_2) = y_2 \wedge \cdots \wedge f(x_t) = y_t\Big] \in \Big(\frac{1-\epsilon}{N^{\underline{t}}}, \frac{1+\epsilon}{N^{\underline{t}}}\Big).$$

Lemma 4.3. *If a permutation distribution $\mathcal{F}$ is pointwise ϵ-close to t-wise independence, then $\mathcal{F}$ is also ϵ-close to t-wise independence.*

We also observe that composition preserves closeness to t-wise independence.

Lemma 4.4. *Let $\mathcal{F}$ be a permutation distribution that is pointwise ϵ-close to t-wise independence. For any permutation distribution $\mathcal{G}$, $\mathcal{F} \circ \mathcal{G}$ is pointwise ϵ-close to t-wise independence.*

4.2 Existential Results for Key Alternating Ciphers

In this section, we will prove our main existential result, that is, for some $r = t + o(t) + s$, there exist permutations $p_1, \ldots, p_r$ such that a r-round KAC using these permutations is $\exp(-s)$-close to t-wise independent.

The result is proved by a careful induction that combines two steps.

- *Independence Amplification:* Lemma 4.6 shows that if $\mathcal{F}$ is pointwise ε-close to t-wise independent, then $\mathcal{S} \circ p \circ \mathcal{F}$ is pointwise $(c(1+\varepsilon)t^2 \log N)$-close to $(t+1)$-wise independent, for most permutations p and for some constant $c > 1$. In other words, one more KAC round takes you from *very* t-*wise independent* to *somewhat* $(t+1)$-*wise independent*. It is important to note that even though the distance of the resulting permutation is $c(1+\varepsilon)t^2 \log N \gg 1$, this is still a non-trivial pointwise guarantee.

  In fact, one can inductively apply Lemma 4.6 and conclude that t-round KAC is pointwise $((t!)^2(c \log N)^{t-1})$-close to t-wise independence, starting from just 1-wise independence. As mentioned before, although the distance is much larger than 1, this is a non-trivial statement, because it is about pointwise closeness.

- *Distance Amplification:* Lemma 4.7 will reduce the distance to t-wise independence by adding more rounds. Say $\mathcal{F}$ is pointwise ε-close to t-wise independent *and* is pointwise ε'-close to $(t+1)$-wise independent, where $\varepsilon' \gg \varepsilon$. I.e., $\mathcal{F}$ is very close to t-wise independent and somewhat close to $(t+1)$-wise independent. Lemma 4.7 shows that adding one more round makes it much closer to $(t+1)$-wise independent. More formally, $\mathcal{S} \circ p \circ \mathcal{F}$ is pointwise $\big(\varepsilon + \tilde{O}(\frac{\varepsilon' t}{\sqrt[3]{N}})\big)$-close to $(t+1)$-wise independent, for most permutations p.

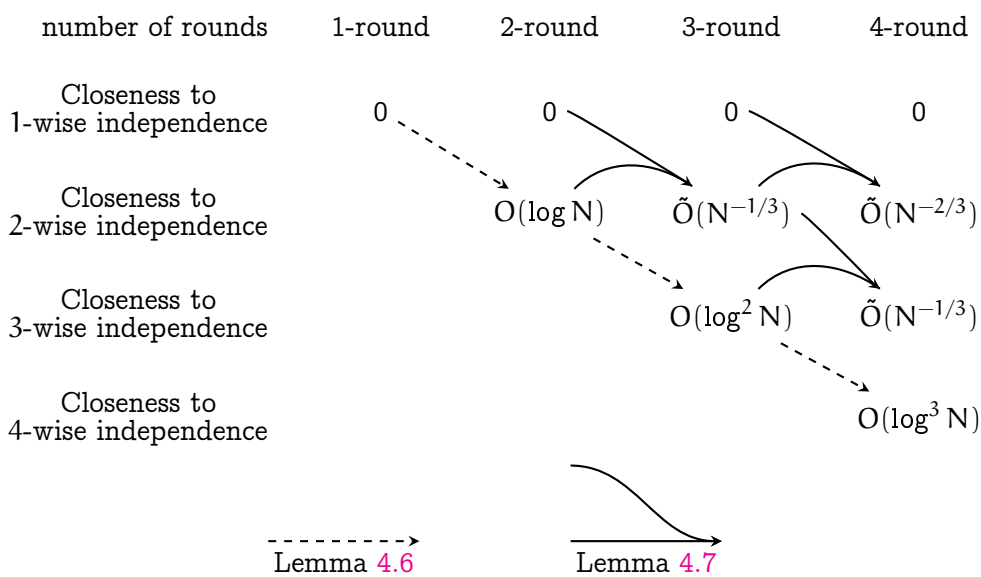

Figure 8: Illustration of the Inductive Proof using Lemmas 4.6, 4.7.

Iterated applications of Lemmas 4.6 and 4.7 takes us very close to t-wise independence in 2t rounds. Indeed, it is not hard to see that one can do even better: between any two successive applications of distance amplification, one can afford to do a large number ($\approx \log N/\log\log N$ many) of iterations of independence amplification. Therefore, to get to t-wise independence, it suffices to work with a $(t + o(t))$-round KAC.

For example, 1-round KAC is 1-wise independent. Then, 2-round KAC is $O(\log N)$-close to 2-wise independent, due to Lemma 4.6. By adding one more round, Lemma 4.7 shows that 3-round KAC is $O(\frac{\log N}{N})$-close to 2-wise independent. Figure 8 illustrates the progression of the inductive argument.

More generally, we show:

Theorem 4.5 (Main KAC Theorem). *For every t, let $r = t + o(t)$. There exist fixed permutations $p_1, \ldots, p_r$ such that the r-round key-alternating cipher is $1/N^{\Omega(1)}$-close to t-wise independent.*

The theorem follows from Lemma 4.6 and Lemma 4.7 below whose proofs are deferred to the full version [LTV21]. Before proceeding to the proof, we remark that the proof shows more: that an overwhelming fraction of choices of permutations $p_1, \ldots, p_r$ give us a t-wise independent KAC.

4.3 Independence Amplification

We first show how to go from a very t-wise independent permutation to a somewhat $(t + 1)$-wise independent permutation by adding one round of the key-alternating cipher.

Lemma 4.6. *Let $\mathcal{F}$ be a distribution which is pointwise ε-close to ℓ-wise independence. At least $1 - 1/N^{t+1}$ of the possible permutations p satisfy the property that $\mathcal{S} \circ p \circ \mathcal{F}$ is pointwise $O((1 + \varepsilon)(t + 1)^2 \log N)$-close to $(t + 1)$-wise independence.*

Proof. We need to show that for most existence of some fixed p, such that for any distinct $x_1, \ldots, x_{t+1} \in \mathfrak{D}$ and for any distinct $y_1, \ldots, y_{t+1} \in \mathfrak{D}$, the probability

$$\Pr_{s \leftarrow \mathfrak{D}, f \leftarrow \mathcal{F}}\Big[(\sigma_s \circ p \circ f)(x_1) = y_1, \ldots, (\sigma_s \circ p \circ f)(x_{t+1}) = y_{t+1}\Big]$$

is close to $\frac{1}{N^{\underline{t+1}}}$.

We prove the statement using a probabilistic argument. Sample the permutation p uniformly at random. Define random variable X_s (whose randomness comes merely from p) as the probability

$$X_s := \Pr_{f \leftarrow \mathcal{F}}\Big[(\sigma_s \circ p \circ f)(x_1) = y_1, \ldots, (\sigma_s \circ p \circ f)(x_{t+1}) = y_{t+1}\Big].$$

Then obviously,

$$\Pr_{s \leftarrow \mathfrak{D}, f \leftarrow \mathcal{F}}\Big[(\sigma_s \circ p \circ f)(x_1) = y_1, \ldots, (\sigma_s \circ p \circ f)(x_{t+1}) = y_{t+1}\Big] = \frac{1}{N}\sum_{s \in \mathfrak{D}} X_s.$$

We need to analyze the random variable X_s. In particular, we have to find out the support and expectation of X_s.

$$\begin{aligned}
X_s &:= \Pr_{f \leftarrow \mathcal{F}}\Big[(\sigma_s \circ p \circ f)(x_1) = y_1, \ldots, (\sigma_s \circ p \circ f)(x_{t+1}) = y_{t+1}\Big] \\
&= \Pr_{f \leftarrow \mathcal{F}}\Big[f(x_1) = p^{-1}(y_1 - s), \ldots, f(x_{t+1}) = p^{-1}(y_{t+1} - s)\Big] \\
&\leq \Pr_{f \leftarrow \mathcal{F}}\Big[f(x_1) = p^{-1}(y_1 - s), \ldots, f(x_t) = p^{-1}(y_t - s)\Big] \\
&\leq \frac{1+\varepsilon}{N^{\underline{t}}}.
\end{aligned}$$

$$\begin{aligned}
\mathop{\mathbb{E}}_{p}[X_s] &= \mathop{\mathbb{E}}_{p} \Pr_{f \leftarrow \mathcal{F}}\Big[(\sigma_s \circ p \circ f)(x_1) = y_1, \ldots, (\sigma_s \circ p \circ f)(x_{t+1}) = y_{t+1}\Big] \\
&= \Pr_{p,f}\Big[f(x_1) = p^{-1}(y_1 - s), \ldots, f(x_{t+1}) = p^{-1}(y_{t+1} - s)\Big] \\
&= \frac{1}{N^{\underline{t+1}}}.
\end{aligned}$$

If $\{X_s\}_{s \in \mathfrak{D}}$ are independent, then the concentration of $\frac{1}{N}\sum_{s \in \mathfrak{D}} X_s$ follows from Chernoff bound. Unfortunately, $\{X_s\}_{s \in \mathfrak{D}}$ are not independent. Instead, they are almost independent, and can be analyzed by martingales.

Sort all elements in $\mathfrak{D}$ into $2(t+1)$ sequences $(s_{1,1}, \ldots, s_{1,N_1}), (s_{2,1}, \ldots, s_{2,N_2}), \ldots, (s_{2(t+1),1}, \ldots, s_{2(t+1),N_{2(t+1)}})$ such that

- for any $j \in [2(t+1)], \ell \in [t]$ and for any $1 \leq i < i' \leq N_j$, we have $y_\ell - s_{j,i} \neq y_{t+1} - s_{j,i'}$,
- for any $j \in [2(t+1)]$, the size of $\{y_\ell - s_i | \ell \in [t+1], i < N_j\}$ is no more than $\frac{N}{2}$.

The existence of such sequences is guaranteed by the following algorithm

1: Initialize $N_j = 0$ for $j \in [2(t+1)]$
2: Initialize set $\Omega_j = \varnothing$ for $j \in [2(t+1)]$
3: **for all** $s \in \mathfrak{D}$ **do**
4: Find some $j \in [2(t+1)]$ satisfying $y_{t+1} - s \notin \Omega_j$ and $|\Omega_j| \le \frac{N}{2}$
5: Set $N_j \leftarrow N_j + 1$, $s_{j,N_j} \leftarrow s$
6: Set $\Omega_j \leftarrow \Omega_j \cup \left\{ y_\ell - s \;\middle|\; \ell \in [t+1] \right\}$

The existence of j in line 4 is guaranteed by two facts:

1. $\sum_j |\Omega_j| \le N(t+1)$. It holds because at any time during the algorithm, $|\Omega_j| \le N_j(t+1)$ and $\sum_j N_j \le N$. The fact implies that $|\Omega_j| \le N/2$ for at least $t+1$ distinct j's.

2. $y_{t+1} - s$ is contained by at most t of Ω_j's. Let $\delta := y_{t+1} - s$. δ is added to one of Ω_j's only when the for loop hits one of $y_1 - \delta, \ldots, y_{t+1} - \delta$. In the current iteration, for loop hits $s = y_{t+1} - \delta$. Before that, δ was added to one of Ω_j's for at most t times.

Once $\mathfrak{D}$ is divided into $2(t+1)$ sequences, we can prove, for each $j \in [2(t+1)]$, the concentration of $\sum_{i=1}^{N_j} X_{s_{j,i}}$ using tail bound of martingale. We need to show that the conditional distribution of $X_{s_{j,i}}$ given $X_{s_{j,1}} + \cdots + X_{s_{j,i-1}}$ has good properties. In particular, we would like to show that for each $j \in [2(t+1)], i \in [N_j]$, the conditional expectation

$$\mathop{\mathbb{E}}_{p}\left[X_{s_{j,i}} \;\middle|\; X_{s_{j,1}} + \cdots + X_{s_{j,i-1}} = z \right]$$

is close to $\frac{1}{N^{\underline{t+1}}}$ for any possible z. Define $\Omega_{j,i}$ as

$$\Omega_{j,i} := \left\{ y_\ell - s_{j,i'} \;\middle|\; \ell \in [t+1], i' \le i \right\}.$$

Note that $X_{s_{j,1}} + \cdots + X_{s_{j,i-1}}$ is determined by the value of p^{-1} on set $\Omega_{j,i-1}$, thus it suffices to bound the expectation of $X_{s_{j,i}}$ conditioning on the values of $p^{-1}(\Omega_{j,i-1})$.

For any $Z \in \mathfrak{D}^{|\Omega_{j,i-1}|}$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{p}\left[X_{s_{j,i}} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z \right] \\
&= \mathop{\Pr}_{p,f}\left[f(x_1) = p^{-1}(y_1 - s_{j,i}), \ldots, f(x_{t+1}) = p^{-1}(y_{t+1} - s_{j,i}) \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z \right] \\
&= \mathop{\Pr}_{p,f}\left[f(x_1) = p^{-1}(y_1 - s_{j,i}), \ldots, f(x_t) = p^{-1}(y_t - s_{j,i}) \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z \right] \\
&\quad \cdot \mathop{\Pr}_{p,f}\left[f(x_{t+1}) = p^{-1}(y_{t+1} - s_{j,i}) \;\middle|\; f(x_1) = p^{-1}(y_1 - s_{j,i}), \ldots, f(x_t) = p^{-1}(y_t - s_{j,i}), p^{-1}(\Omega_{j,i-1}) = Z \right] \\
&\le \frac{1+\varepsilon}{N^{\underline{t}}} \cdot \frac{1}{N - |\Omega_{j,i-1}| - t} \\
&\le \frac{2(1+\varepsilon)}{N^{\underline{t+1}}}.
\end{aligned}$$

As $X_{s_{j,i}} \in [0, \frac{1+\varepsilon}{N^{\underline{t}}}]$, we have

$$\mathop{\mathrm{Var}}_{p}\left[X_{s_{j,i}} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z \right] \le \mathop{\mathbb{E}}_{p}\left[X_{s_{j,i}}^2 \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z \right] \le \frac{2(1+\varepsilon)^2}{(N-t)(N^{\underline{t}})^2}$$

Thus by Corollary 2.21, (Formally, the filters are defined as $\mathfrak{F}_i := p^{-1}(\Omega_{j,i})$.)

$$\Pr\left[\sum_{i=1}^{N_j} X_{s_{j,i}} - N_j \cdot \frac{2(1+\varepsilon)}{N^{\underline{t+1}}} > \tau\right] \le \exp\left(-\frac{\frac{1}{2}\tau^2}{N_j \frac{2(1+\varepsilon)^2}{(N-t)(N^{\underline{t}})^2} + \frac{1}{3} \cdot \frac{(1+\varepsilon)\tau}{N^{\underline{t}}}}\right).$$

Let $\tau = \frac{1+\varepsilon}{N^{\underline{t}}}\tau'$, then

$$\Pr\left[\sum_{i=1}^{N_j} X_{s_{j,i}} - N_j \cdot \frac{2(1+\varepsilon)}{N^{\underline{t+1}}} > \frac{1+\varepsilon}{N^{\underline{t}}}\tau'\right] \le \exp\left(-\frac{\frac{1}{2}\tau'^2}{\frac{2N_j}{N-t} + \frac{1}{3} \cdot \tau'}\right) \le \exp\left(-\frac{\frac{1}{2}\tau'^2}{1 + \frac{1}{3} \cdot \tau'}\right).$$

Choose $\tau' = 3(t+1)\ln N + \ln(2(t+1)) = O(t \log N)$, then for each $j \le 2(t+1)$

$$\Pr\left[\sum_{i=1}^{N_j} X_{s_{j,i}} - N_j \cdot \frac{2(1+\varepsilon)}{N^{\underline{t+1}}} > \frac{1+\varepsilon}{N^{\underline{t}}}(3(t+1)\ln N + \ln(2(t+1)))\right] \le \frac{1}{2(t+1)N^{3(t+1)}}.$$

By union bound over all $j \le 2(t+1)$,

$$\Pr\left[\sum_{s\in\mathfrak{D}} X_s - N \cdot \frac{2(1+\varepsilon)}{N^{\underline{t+1}}} > 2(t+1) \cdot \frac{1+\varepsilon}{N^{\underline{t}}}(3(t+1)\ln N + \ln(2(t+1)))\right] \le \frac{1}{N^{3(t+1)}}.$$

That is, with probability $1 - \frac{1}{N^{3(t+1)}}$ (over the random choice of p),

$$\frac{1}{N}\sum_{s\in\mathfrak{D}} X_s \le \frac{2(1+\varepsilon)}{N^{\underline{t+1}}} + \frac{2(1+\varepsilon)(t+1)(3(t+1)\ln N + \ln(2(t+1)))}{N \cdot N^{\underline{t}}} = \frac{O((1+\varepsilon)(t+1)^2 \log N)}{N^{\underline{t+1}}}.$$

The proof is completed by a union bound over all possible $x_1, \ldots, x_{t+1}, y_1, \ldots, y_{t+1}$. □

4.4 Distance Amplification

We now show how to go from a permutation that is close to t-wise independent *and* somewhat $(t+1)$-wise independent permutation into a close to $(t+1)$-wise independent permutation by adding one round of the key-alternating cipher.

Lemma 4.7. *Let $\mathcal{F}$ be a permutation distribution that is pointwise ε-close to t-wise independence and is pointwise ε'-close to $(t+1)$-wise independence. At least $1-1/N^{t+1}$ of the possible permutations p satisfy the property that $\mathcal{S} \circ p \circ \mathcal{F}$ is pointwise $\left(\varepsilon + 4\varepsilon'(t+1)\sqrt[3]{\ln N/N}\right)$-close to $(t+1)$-wise independence.*

Proof. For any distinct $x_1, \ldots, x_{t+1} \in \mathfrak{D}$ and for any distinct $y_1, \ldots, y_{t+1} \in \mathfrak{D}$, define a random variable X_s for each $s \in \mathfrak{D}$ as

$$X_s := \Pr_{f \leftarrow \mathcal{F}}\left[(\sigma_s \circ p \circ f)(x_1) = y_1, \ldots, (\sigma_s \circ p \circ f)(x_{t+1}) = y_{t+1}\right].$$

The definition is the same as in the proof of Lemma 4.6.

As $\mathcal{F}$ is pointwise ε-close to t-wise independence and pointwise ε'-close to $(t+1)$-wise independence, X_s lays in $[0, \frac{1+\varepsilon}{N^{\underline{t}}}] \cap [\frac{1-\varepsilon'}{N^{\underline{t+1}}}, \frac{1+\varepsilon'}{N^{\underline{t+1}}}]$.

Similar to the proof of Lemma 4.6, sort all elements in $\mathfrak{D}$ into $\frac{t+1}{\mu}$ sequences $\{s_{j,i}\}_{j\in[\frac{t+1}{\mu}], i\in[N_j]}$, where $\mu \le 1$ is a parameter we will choose later, such that

- $\sum_{j\in[\frac{t+1}{\mu})]} N_j = N$
- define $\Omega_{j,i} := \{y_\ell - s_{j,i'} \mid \ell \in [t+1], i' < i\}$
- for any $j \in [\frac{t+1}{\mu}], i \in [N_j]$, we have $y_{t+1} - s_{j,i} \notin \Omega_{j,i-1}$
- for any $j \in [\frac{t+1}{\mu}], i \in [N_j]$, we have $|\Omega_{j,i}| \le \mu N$

The existence of such $\frac{t+1}{\mu}$ sequences is guaranteed by the following algorithm

Initialize $N_j = 0$ for $j \in [\frac{t+1}{\mu}]$
Initialize set $\Omega_{j,0} = \varnothing$ for $j \in [\frac{t+1}{\mu}]$
for all $s \in \mathfrak{D}$ **do**
    Find the smallest j satisfying $y_{t+1} - s \notin \Omega_{j,N_j}$ and $|\Omega_{j,N_j}| \le \mu N$
    Set $N_j \leftarrow N_j + 1$, $s_{j,N_j} \leftarrow s$
    Set $\Omega_{j,N_j} \leftarrow \Omega_{j,N_j-1} \cup \{y_\ell - s \mid \ell \in [t+1]\}$

The correctness of the algorithm can be analyzed the same way as in Lemma 4.6.

Let W_ℓ denote the random variable $p^{-1}(y_\ell - s_{j,i})$ for $\ell \in [t+1]$. For any $Z \in \mathfrak{D}^{|\Omega_{j,i-1}|}$, we want to bound

$$\begin{aligned}
&\mathop{\mathbb{E}}_{p}\left[X_{s_{j,i}} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z\right] \\
&= \mathbb{E}\left[\mathop{\mathbb{E}}_{p}\left[X_{s_{j,i}} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z, W_1, \ldots, W_t\right] \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z\right] \\
&= \mathbb{E}\left[\mathop{\Pr}_{p,f}\left[f(x_1) = W_1, \ldots, f(x_{t+1}) = W_{t+1} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z, W_1, \ldots, W_t\right] \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z\right]
\end{aligned} \tag{6}$$

To do so, it suffices to bound the following for any $w_1, \ldots, w_k \in \mathfrak{D}$,

$$\begin{aligned}
&\mathop{\Pr}_{p,f}\left[f(x_1) = W_1, \ldots, f(x_{t+1}) = W_{t+1} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z, W_1 = w_1, \ldots, W_t = w_k\right] \\
&= \mathop{\Pr}_{p,f}\left[f(x_1) = w_1, \ldots, f(x_k) = w_k, f(x_{t+1}) = W_{t+1} \;\middle|\; p^{-1}(\Omega_{j,i-1}) = Z, W_1 = w_1, \ldots, W_t = w_k\right] \\
&= \mathop{\Pr}_{\substack{f\leftarrow\mathcal{F} \\ W_{t+1}\leftarrow\mathfrak{D}\setminus(Z\cup\{w_1,\ldots,w_k\})}}\left[f(x_1) = w_1, \ldots, f(x_t) = w_t, f(x_{t+1}) = W_{t+1}\right] \\
&= \mathop{\mathbb{E}}_{W_{t+1}\leftarrow\mathfrak{D}\setminus(Z\cup\{w_1,\ldots,w_k\})}\left[\mathop{\Pr}_{f\leftarrow\mathcal{F}}\left[f(x_1) = w_1, \ldots, f(x_t) = w_t, f(x_{t+1}) = W_{t+1}\right]\right]
\end{aligned} \tag{7}$$

It is hard to proceed the above analysis because of Z. For a moment, let us ignore the existence of Z, and consider another random variable

$$\mathop{\Pr}_{f\leftarrow\mathcal{F}}\left[f(x_1) = w_1, \ldots, f(x_t) = w_t, f(x_{t+1}) = W_{t+1}\right], \tag{8}$$

where the randomness comes from $W_{t+1} \leftarrow \mathfrak{D} \setminus \{w_1, \ldots, w_k\}$. To estimate the support of (8), as $\mathcal{F}$ is pointwise ε'-close to $(t+1)$-wise independence,

$$\mathop{\Pr}_{f\leftarrow\mathcal{F}}\left[f(x_1) = w_1, \ldots, f(x_t) = w_t, f(x_{t+1}) = W_{t+1}\right] \in \left[\frac{1-\varepsilon'}{N^{\underline{t+1}}}, \frac{1+\varepsilon'}{N^{\underline{t+1}}}\right]. \tag{9}$$

To bound the expectation of (8), as $\mathcal{F}$ is pointwise ε-close to t-wise independence,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{W_{t+1}\leftarrow\mathfrak{D}\setminus\{w_1,\ldots,w_k\}}\Big[\Pr_{f\leftarrow\mathcal{F}}\Big[f(x_1)=w_1,\ldots,f(x_t)=w_t,f(x_{t+1})=W_{t+1}\Big]\Big]\\
=&\Pr_{\substack{f\leftarrow\mathcal{F}\\ W_{t+1}\leftarrow\mathfrak{D}\setminus\{w_1,\ldots,w_k\}}}\Big[f(x_1)=w_1,\ldots,f(x_t)=w_t,f(x_{t+1})=W_{t+1}\Big]\\
=&\Pr_{f\leftarrow\mathcal{F}}\Big[f(x_1)=w_1,\ldots,f(x_t)=w_t\Big]\\
&\cdot\Pr_{\substack{f\leftarrow\mathcal{F}\\ W_{t+1}\leftarrow\mathfrak{D}\setminus\{w_1,\ldots,w_k\}}}\Big[f(x_{t+1})=W_{t+1}\;\Big|\;f(x_1)=w_1,\ldots,f(x_t)=w_t\Big]\\
=&\Pr_{f\leftarrow\mathcal{F}}\Big[f(x_1)=w_1,\ldots,f(x_t)=w_t\Big]\cdot\frac{1}{N-t}\\
\in&\left[\frac{1-\varepsilon}{N^{\underline{t+1}}},\frac{1+\varepsilon}{N^{\underline{t+1}}}\right].
\end{aligned}\tag{10}$$

Combining (9) and (10) gives an estimation of (7). We now know the random variable (8) lays in $\left[\frac{1-\varepsilon'}{N^{\underline{t+1}}},\frac{1+\varepsilon'}{N^{\underline{t+1}}}\right]$, and its expectation is in $\left[\frac{1-\varepsilon}{N^{\underline{t+1}}},\frac{1+\varepsilon}{N^{\underline{t+1}}}\right]$, then what is its expectation conditioning on an event[10] of probability no less than $1-\mu$? By the law of total expectation, it's easy to show that the conditional expectation, which is (7), lays in the range $\left[(1-\frac{\varepsilon+\mu\varepsilon'}{1-\mu})/N^{\underline{t+1}},(1+\frac{\varepsilon+\mu\varepsilon'}{1-\mu})/N^{\underline{t+1}}\right]$. Then combine with (6), we have

$$\mathop{\mathbb{E}}_{p}\Big[X_{s_{j,i}}\;\Big|\;p^{-1}(\Omega_{j,i-1})=Z\Big]\in\left[\frac{1-\frac{\varepsilon+\mu\varepsilon'}{1-\mu}}{N^{\underline{t+1}}},\frac{1+\frac{\varepsilon+\mu\varepsilon'}{1-\mu}}{N^{\underline{t+1}}}\right].$$

By the Azuma bound,

$$\Pr\left[\sum_{i\in[N_j]}X_{s_{j,i}}-N_j\cdot\frac{1+\frac{\varepsilon+\mu\varepsilon'}{1-\mu}}{N^{\underline{t+1}}}\geq\tau\right]\leq\exp\Big(-\frac{2\tau^2}{N_j\cdot(\frac{2\varepsilon'}{N^{\underline{t+1}}})^2}\Big)\leq\exp\Big(-\frac{1}{2\mu N}\Big(\frac{\tau}{\varepsilon'/N^{\underline{t+1}}}\Big)^2\Big).$$

Symmetrically,

$$\Pr\left[\sum_{i\in[N_j]}X_{s_{j,i}}-N_j\cdot\frac{1-\frac{\varepsilon+\mu\varepsilon'}{1-\mu}}{N^{\underline{t+1}}}\leq-\tau\right]\leq\exp\Big(-\frac{1}{2\mu N}\Big(\frac{\tau}{\varepsilon'/N^{\underline{t+1}}}\Big)^2\Big).$$

As we are going to use union bound, we have to ensure that the tail probability is no more than $\frac{1}{2\cdot\frac{t+1}{\mu}\cdot N^{3(t+1)}}$. (Among the denominator: 2 is for union bound of the Azuma bound in two directions; $\frac{t+1}{\mu}$ is for the union bound over all $\frac{t+1}{\mu}$ sequences; $N^{2(t+1)}$ is for the union bound over all $(x_1,\ldots,x_{t+1},y_1,\ldots,y_{t+1})$; the remaining N^{t+1} is the bound we need.) Thus we choose $\tau=\frac{\varepsilon'\cdot\sqrt{\ln 2+\ln(\frac{t+1}{\mu})+3(t+1)\ln N}\cdot\sqrt{2\mu N}}{N^{\underline{t+1}}}\leq\frac{\varepsilon'\cdot\sqrt{4(t+1)\ln N}\cdot\sqrt{2\mu N}}{N^{\underline{t+1}}}$. By union bound, at least $1-\frac{1}{N^{t+1}}$ of the

[10] Conditional on the event that $W_{t+1}\notin Z$.

possible permutations p satisfy the properties that

$$\sum_{s\in\mathfrak{D}} X_s \le N\cdot\frac{1+\frac{\varepsilon+\mu\varepsilon'}{1-\mu}}{N^{\underline{t+1}}}+\frac{t+1}{\mu}\frac{\varepsilon'\cdot\sqrt{4(t+1)\ln N}\cdot\sqrt{2\mu N}}{N^{\underline{t+1}}},$$
$$\sum_{s\in\mathfrak{D}} X_s \ge N\cdot\frac{1-\frac{\varepsilon+\mu\varepsilon'}{1-\mu}}{N^{\underline{t+1}}}-\frac{t+1}{\mu}\frac{\varepsilon'\cdot\sqrt{4(t+1)\ln N}\cdot\sqrt{2\mu N}}{N^{\underline{t+1}}},$$

for any $x_1,\ldots,x_{t+1},y_1,\ldots,y_{t+1}\in\mathfrak{D}$. Then $\mathcal{S}\circ p\circ\mathcal{F}$ is pointwise $\left(\frac{\varepsilon+\mu\varepsilon'}{1-\mu}+\frac{\varepsilon'\sqrt{8(t+1)^3\ln N}}{\sqrt{\mu N}}\right)$-close to $(t+1)$-wise independence.

The distance is

$$\frac{\varepsilon+\mu\varepsilon'}{1-\mu}+\frac{\varepsilon'\sqrt{8(t+1)^3\ln N}}{\sqrt{\mu N}}\le\varepsilon+\varepsilon'\Big(\mu+\frac{\sqrt{8(t+1)^3\ln N/N}}{\sqrt{\mu}}\Big).$$

To minimize the distance, we choose $\mu=\sqrt[3]{4(t+1)^3\ln N/N}$. Then distance becomes

$$\varepsilon+\varepsilon'(1+\sqrt{2})(t+1)\sqrt[3]{4\ln N/N}\le\varepsilon+4\varepsilon'(t+1)\sqrt[3]{\ln N/N}. \qquad \square$$

Acknowledgments. The third author thanks Charlie Rackoff for inspiring discussions a decade ago on the topic of what can be proved about the security of block ciphers. He also thanks Orr Dunkelman and Thomas Peyrin for answering his questions about the role of key schedule in the security of AES. We thank the Simons Institute for hosting us at the "Lattices: Algorithms, Complexity and Cryptography" program where the seeds of this work were planted. TL was supported by NSF grants CNS-1528178, CNS-1929901, CNS-1936825 (CAREER), CNS-2026774, a JP Morgan AI research Award, and a Simons Foundation Collaboration Grant on Algorithmic Fairness. ST was supported in part by NSF grants CNS-1930117 (CAREER), CNS-1926324, CNS-2026774, a Sloan Research Fellowship, and a JP Morgan Faculty Award. VV was supported by DARPA under Agreement No. HR00112020023, a grant from the MIT-IBM Watson AI, a grant from Analog Devices, a Microsoft Trustworthy AI grant, and a DARPA Young Faculty Award.

References

[ABD+13] Elena Andreeva, Andrey Bogdanov, Yevgeniy Dodis, Bart Mennink, and John P. Steinberger. On the indifferentiability of key-alternating ciphers. In *CRYPTO 2013*, volume 8042 of *Lecture Notes in Computer Science*, pages 531–550. Springer, 2013. 7

[AGR+16] Martin R. Albrecht, Lorenzo Grassi, Christian Rechberger, Arnab Roy, and Tyge Tiessen. Mimc: Efficient encryption and cryptographic hashing with minimal multiplicative complexity. In *ASIACRYPT (1)*, volume 10031 of *Lecture Notes in Computer Science*, pages 191–219, 2016. 2

[AL13] Noga Alon and Shachar Lovett. Almost k-wise vs. k-wise independent permutations, and uniformity for general group actions. *Theory Comput.*, 9:559–577, 2013. 2

[BH08] Alex Brodsky and Shlomo Hoory. Simple permutations mix even better. *Random Struct. Algorithms*, 32(3):274–289, 2008. 2

[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. 7

[BKR11] Andrey Bogdanov, Dmitry Khovratovich, and Christian Rechberger. Biclique cryptanalysis of the full AES. In *ASIACRYPT*, volume 7073 of *Lecture Notes in Computer Science*, pages 344–371. Springer, 2011. 1

[BN15] Céline Blondeau and Kaisa Nyberg. Perfect nonlinear functions and cryptography. *Finite Fields Their Appl.*, 32:120–147, 2015. 6

[BS91] Eli Biham and Adi Shamir. Differential cryptanalysis of des-like cryptosystems. *J. Cryptol.*, 4(1):3–72, 1991. 1, 6

[BV05] Thomas Baignères and Serge Vaudenay. Proving the security of AES substitution-permutation network. In *Selected Areas in Cryptography*, volume 3897 of *Lecture Notes in Computer Science*, pages 65–81. Springer, 2005. 6

[Car53] L. Carlitz. Permutations in a finite field. *Proceedings of American Mathematical Society*, page 538, 1953. 44

[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 *CRYPTO 2018*, volume 10991 of *Lecture Notes in Computer Science*, pages 722–753. Springer, 2018. 7

[CG75] Don Coppersmith and Edna Grossman. Generators for certain alternating groups with applications to crytography. *SIAM Journal on Applied Mathematics*, 29(4):624–627, 1975. 5

[CL06] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006. 14

[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. 7

[CLL+14] Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the two-round even-mansour cipher. In *CRYPTO (1)*, volume 8616 of *Lecture Notes in Computer Science*, pages 39–56. Springer, 2014. 7

[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. 7

[CVS09] Andrea Caranti, Francesca Dalla Volta, and Massimiliano Sala. An application of the o'nan-scott theorem to the group generated by the round functions of an aes-like cipher. *Des. Codes Cryptogr.*, 52(3):293–301, 2009. 5

[Dae95] Joan Daemen. Cipher and hash function design strategies based on linear and differential cryptanalysis, 1995. Ph.D. Thesis, KU Leuven. 22

[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. 7

[DR06] Joan Daemen and Vincent Rijmen. Understanding two-round differentials in AES. In *SCN 2006*, volume 4116 of *Lecture Notes in Computer Science*, pages 78–94. Springer, 2006. 6

[DSSL16] Yevgeniy Dodis, Martijn Stam, John P. Steinberger, and Tianren Liu. Indifferentiability of confusion-diffusion networks. In *EUROCRYPT 2016*, volume 9666 of *Lecture Notes in Computer Science*, pages 679–704. Springer, 2016. 7

[GL15a] Chun Guo and Dongdai Lin. On the indifferentiability of key-alternating feistel ciphers with no key derivation. In *TCC (1)*, volume 9014 of *Lecture Notes in Computer Science*, pages 110–133. Springer, 2015. 7

[GL15b] Chun Guo and Dongdai Lin. A synthetic indifferentiability analysis of interleaved double-key even-mansour ciphers. In *ASIACRYPT (2)*, volume 9453 of *Lecture Notes in Computer Science*, pages 389–410. Springer, 2015. 7

[HLL+00] Seokhie Hong, Sangjin Lee, Jongin Lim, Jaechul Sung, Dong Hyeon Cheon, and Inho Cho. Provable security against differential and linear cryptanalysis for the SPN structure. In *FSE*, volume 1978 of *Lecture Notes in Computer Science*, pages 273–283. Springer, 2000. 6

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

[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. 7

[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. 1, 44

[JRS88] Burton S. Kaliski Jr., Ronald L. Rivest, and Alan T. Sherman. Is the data encryption standard a group? (results of cycling experiments on DES). *J. Cryptol.*, 1(1):3–36, 1988. 5

[Kel04] Liam Keliher. Refined analysis of bounds related to linear and differential cryptanalysis for the AES. In *AES 2004*, volume 3373 of *Lecture Notes in Computer Science*, pages 42–57. Springer, 2004. 1, 6

[KHL+02] Ju-Sung Kang, Seokhie Hong, Sangjin Lee, Okyeon Yi, Choonsik Park, and Jongin Lim. Practical and provable security against differential and linear cryptanalysis for substitution-permutation networks. *Etri Journal*, 23, 02 2002. 6, 22

[KMT01a] Liam Keliher, Henk Meijer, and Stafford E. Tavares. Improving the upper bound on the maximum average linear hull probability for rijndael. In *Selected Areas in Cryptography*, volume 2259 of *Lecture Notes in Computer Science*, pages 112–128. Springer, 2001. 1, 6

[KMT01b] Liam Keliher, Henk Meijer, and Stafford E. Tavares. New method for upper bounding the maximum average linear hull probability for spns. In *EUROCRYPT*, volume 2045 of *Lecture Notes in Computer Science*, pages 420–436. Springer, 2001. 1, 6

[KNR05] Eyal Kaplan, Moni Naor, and Omer Reingold. Derandomized constructions of k-wise (almost) independent permutations. In *APPROX-RANDOM*, volume 3624 of *Lecture Notes in Computer Science*, pages 354–365. Springer, 2005. 2

[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. 1, 6, 8

[Knu98] Lars Knudsen. Deal - a 128-bit block cipher. In *NIST AES Proposal*, 1998. 1, 6

[KS07] Liam Keliher and Jiayuan Sui. Exact maximum expected differential and linear probability for two-round advanced encryption standard. *IET Inf. Secur.*, 1(2):53–57, 2007. 1, 6

[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. 1

[Lai94] Xuejia Lai. *Higher Order Derivatives and Differential Cryptanalysis*, pages 227–233. Springer US, Boston, MA, 1994. 1, 6, 8

[LMM91] Xuejia Lai, James L. Massey, and Sean Murphy. Markov ciphers and differential cryptanalysis. In *EUROCRYPT*, volume 547 of *Lecture Notes in Computer Science*, pages 17–38. Springer, 1991. 5, 6

[LN86] Rudolf Lidl and Harald Niederreiter. *Introduction to Finite Fields and Their Applications*. Cambridge University Press, USA, 1986. 11

[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. 7

[LTV21] Tianren Liu, Stefano Tessaro, and Vinod Vaikuntanathan. The t-wise independence of substitution-permutation networks. *IACR Cryptol. ePrint Arch.*, 2021:507, 2021. 19, 32

[MPR07] Ueli M. Maurer, Krzysztof Pietrzak, and Renato Renner. Indistinguishability amplification. In *CRYPTO 2007*, volume 4622 of *Lecture Notes in Computer Science*, pages 130–149. Springer, 2007. 3, 8

[MPW94] Sean Murphy, Kenneth G. Paterson, and Peter R. Wild. A weak cipher that generates the symmetric group. *J. Cryptol.*, 7(1):61–65, 1994. 6

[MV15] Eric Miles and Emanuele Viola. Substitution-permutation networks, pseudorandom functions, and natural proofs. *J. ACM*, 62(6):46:1–46:29, 2015. 6, 22

[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. 1, 6

[NK95] Kaisa Nyberg and Lars R. Knudsen. Provable security against a differential attack. *J. Cryptol.*, 8(1):27–37, 1995. 1, 6

[Nyb93] Kaisa Nyberg. Differentially uniform mappings for cryptography. In *EUROCRYPT '93*, volume 765 of *Lecture Notes in Computer Science*, pages 55–64. Springer, 1993. 3, 5, 6, 12, 19

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, USA, 2014. 13

[PSC+02] Sangwoo Park, Soo Hak Sung, Seongtaek Chee, E-Joong Yoon, and Jongin Lim. On the security of rijndael-like structures against differential and linear cryptanalysis. In *ASIACRYPT*, volume 2501 of *Lecture Notes in Computer Science*, pages 176–191. Springer, 2002. 1, 6

[PSLL03] Sangwoo Park, Soo Hak Sung, Sangjin Lee, and Jongin Lim. Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES. In *FSE*, volume 2887 of *Lecture Notes in Computer Science*, pages 247–260. Springer, 2003. 1, 6

[Ste12] John P. Steinberger. Improved security bounds for key-alternating ciphers via hellinger distance. *IACR Cryptol. ePrint Arch.*, 2012:481, 2012. 7

[Vau03] Serge Vaudenay. Decorrelation: A theory for block cipher security. *J. Cryptol.*, 16(4):249–286, 2003. 2, 6

[WYCD20] Yusai Wu, Liqing Yu, Zhenfu Cao, and Xiaolei Dong. Tight security analysis of 3-round key-alternating cipher with a single permutation. In *ASIACRYPT (1)*, volume 12491 of *Lecture Notes in Computer Science*, pages 662–693. Springer, 2020. 7

A Pairwise Independence – Two-round Convergence

While this follows directly from our result on SPN, we give here a direct (and tighter) proof that *two* rounds of a key-alternating cipher with block size n with the inverse permutation $P(x) = x^{2^n-2}$ are enough to obtain convergence to pairwise-independence. As a bonus, we obtain a nearly tight bound on the differential probabilities of this S-box thanks to the direct analysis.

Theorem A.1. *Assume that n is even. The two-round key alternating cipher with invese permutation is $(\frac{9}{2\cdot 2^n} + \frac{7}{2^{2n}})$-close to pairwise-independent. Thus, for $n \geq 4$, the statistical distance from uniform is at most $\frac{5}{2^n}$.*

We note that experimental validation suggests that the real bound is closer $(c/2^n)$ for some $c \leq 2$. The proof below is somewhat loose, and perhaps unavoidably so.

Proof. (of Theorem A.1.) For two any fixed inputs with difference δ, we consider the probability distribution of the output differences. Because of the final key xored to the output, it follows by a standard argument that the distance from uniform of the two outputs is at most the distance from uniform of the output difference.

The probability that an input difference of $\delta \neq 0$ results in an output difference of δ' is $\sum_{\gamma\neq 0} p_{\delta,\gamma} p_{\gamma,\delta'}$, where $p_{\alpha,\beta}$ is the differential probability for the inverse permutation mapping difference α to difference β.

Case 1: $\delta = \delta'$. If $\gamma = \delta^{-1} = (\delta')^{-1}$, then

$$p_{\delta,\gamma} p_{\gamma,\delta'} = (4/2^n)^2 = 16/2^{2n}$$

In all other cases $(\gamma \neq 0, \delta^{-1})$ of which there are $2^{n-1} - 2$, we have

$$p_{\delta,\gamma} p_{\gamma,\delta'} = (2/2^n)^2 = 4/2^{2n}$$

In total,

$$\sum_{\gamma\neq 0} p_{\delta,\gamma} p_{\gamma,\delta'} = \frac{16}{2^{2n}} + (2^n - 2)\frac{4}{2^{2n}} = \frac{4}{2^n} + \frac{8}{2^{2n}}$$

Case 2: $\delta \neq \delta'$. The set of all β $(= \gamma^{-1})$ such that $\mathrm{Tr}(\delta^{-1}\beta) = \mathrm{Tr}((\delta')^{-1}\beta)$ is a subspace of dimension $n-2$, by Lemma 2.13. However, since the elements of the subspace are assigned slightly different probabilities, we will need to consider three further subcases.

Subcase 2.1: $\mathrm{Tr}(\delta(\delta')^{-1}) = \mathrm{Tr}(\delta'\delta^{-1}) = 0$. In this case,

$$\sum_{\gamma\neq 0} p_{\delta,\gamma} p_{\gamma,\delta'} = \left(\frac{2^n}{4} - 3\right)\frac{4}{2^{2n}} + 2\cdot\frac{4\cdot 2}{2^{2n}} = \frac{1}{2^n} + \frac{4}{2^{2n}}$$

Subcase 2.2: $\mathsf{Tr}(\delta(\delta')^{-1}) = \mathsf{Tr}(\delta'\delta^{-1}) = 1$. In this case,

$$\sum_{\gamma \neq 0} p_{\delta,\gamma} p_{\gamma,\delta'} = \left(\frac{2^n}{4} - 1\right)\frac{4}{2^{2n}} = \frac{1}{2^n} - \frac{4}{2^{2n}}$$

This is the case since $p_{\delta,(\delta')^{-1}} = p_{\delta^{-1},\delta'} = 0$.

Subcase 2.3: $\mathsf{Tr}(\delta(\delta')^{-1}) \neq \mathsf{Tr}(\delta'\delta^{-1})$. In this case,

$$\sum_{\gamma \neq 0} p_{\delta,\gamma} p_{\gamma,\delta'} = \left(\frac{2^n}{4} - 2\right)\frac{4}{2^{2n}} + \frac{4 \cdot 2}{2^{2n}} = \frac{1}{2^n}$$

By Proposition 2.4, to compute the total variation distance from uniform, it suffices to consider δ' that fall into cases 1 and 2.1 since those are exactly the δ' whose the probability is more than $1/(2^n - 1)$, the uniform probability density. We also use the fact that the number of δ' that fall into case 2.1 is at most $2^n/2$.

The statistical distance from the uniform distribution over $\mathbb{F}_{2^n} \setminus \{0\}$ is therefore

$$\begin{aligned} d_{TV} &\leq \left(\frac{4}{2^n} + \frac{8}{2^{2n}} - \frac{1}{2^n - 1}\right) + \frac{2^n}{2} \cdot \left(\frac{1}{2^n} + \frac{4}{2^{2n}} - \frac{1}{2^n - 1}\right) \\ &\leq \left(\frac{3}{2^n} + \frac{7}{2^{2n}}\right) + \frac{3}{2 \cdot 2^n} \\ &\leq \frac{9}{2 \cdot 2^n} + \frac{7}{2^{2n}} \end{aligned}$$

For large enough n (in particular, $n \geq 4$), this is at most $\frac{5}{2^n}$. □

Finally, Lemma 2.6 gives us the following direct corollary.

Corollary A.2. *Assume that $n \geq 4$ is even. Then, the $2t$-round key-alternating cipher with permutation $P(x) = x^{2^n-2}$ is $(\frac{10}{2^n})^t$-close to pairwise-independent.*

B Attack against 4-wise Independence

We show that for a modest number of rounds $r \ll 2^n$, the function family key alternating cipher with permutation $P(x) = x^{2^n-2}$ is not 4-wise independent. The attack is essentially the same as the interpolation attack of Jakobsen and Knudsen [JK97].

The main observation here is that the output of the function can be written as a rational function of x where the numerator and denominator are linear functions in x. This result should be seen in contrast to a result of Carlitz [Car53] who shows that with a large number, in particular $2^{\Omega(n)}$, rounds, the key alternating cipher with permutation P *would actually be* indistinguishable from a *truly random permutation*. Indeed, for such large number of rounds, it is impossible to write the output as a rational function due to the inevitable emergence of the zero input. For convenience, in the following, we note by $F^{(r)}_{P,K_1,\ldots,K_r}$ the r-round key-alternating cipher *without* the final key XORed to the output. (It is clear that the attack can be extended with this final key.)

Lemma B.1. *For every* r*, with probability* $1 - \frac{r}{2^n}$ *over a random choice of* $K_1, \ldots, K_r \leftarrow \mathbb{F}_{2^n}$*, there are* $L_1, L_2, L_3 \in \mathbb{F}_2^n$ *such that*

$$F^{(r)}_{P,K_1,\ldots,K_r}(x) = (x + L_1) \cdot (L_2 x + L_3)^{-1}$$

Proof. We prove this by induction. For $r = 0$, $L_1 = 0$, $L_2 = 0$ and $L_3 = 1$.

Assume the statement is true for i. Then:

$$\begin{aligned} F^{(i+1)}_{P,K_1,\ldots,K_r}(x) &= (F^{(i)}_{P,K_1,\ldots,K_{i-1}}(x) + K_i)^{-1} \\ &= \left(\frac{x + L_1}{L_2 x + L_3} + K_i\right)^{-1} \\ &= \frac{L_2 x + L_3}{(K_i L_2 + 1)x + (K_i L_3 + L_1)} \end{aligned}$$

which is of the same form as well. The only way this fails is if one of the numerators in the expression turns out to be 0. The probability of this happening for any one of the r rounds is at most $r/2^n$. □

Given Lemma B.1, the attack is simple. Every evaluation of F gives us a linear equation in L_1, L_2 and L_3. Three such linearly independent equations suffice to recover them and we can use the fourth to check that these are indeed correct.

A Note on AES and 4-wise Independence. The fact that there is way to break the 4-wise independence of a function so closely related to AES might give us some pause. However, when one extends the function family by performing the inverse function in parallel on small blocks and mixing them in every round, just like AES does, the attack as above seems to dissipate. This is because the number of unknowns L_i grows exponentially with k, the number of parallel invocations of the S-box.

B.1 Warm-up: Pointwise t-wise Independence of Generalized KAC

This section shows that the generalized KAC defined below is t-wise independent. This is intended as a warmup to the main results of Section 4 and can be skipped on a first read.

A t-round generalized KAC is the composition

$$\mathcal{F}_1 \circ \mathcal{F}_2 \circ \ldots \circ \mathcal{F}_t,$$

where $\mathcal{F}_1, \ldots, \mathcal{F}_t$ are permutations distributions with (presumably) low entropy. Then KAC is the special case by letting $\mathcal{F}_1 = \mathcal{S} \circ p_1$, $\ldots$, $\mathcal{F}_{t-1} = \mathcal{S} \circ p_{t-1}$, $\mathcal{F}_t = \mathcal{S}$.

The simplified question is to find permutation distributions $\mathcal{F}_1, \ldots, \mathcal{F}_t$ with low max-entropy, such that the their composition (i.e., t-round gKAC) is close to t- or $(t - O(1))$-wise independent.

The plan is to prove by induction. Assume $\mathcal{F}$ is (close to) t-wise independent. Sample $\mathcal{H}$ as the uniform distribution over a random subset of permutations. Then we want to prove $\mathcal{H} \circ \mathcal{F}$ is close to $(t + 1)$-wise independent.

Start with a concrete example $\mathcal{F} = \mathcal{L}$, which is 2-wise independent.

Lemma B.2. *Sample $\mathcal{H}$ as a random collection of M permutations. Let $\mathcal{H}$ also denote the uniform distribution over $\mathcal{H}$. Then $\mathcal{H} \circ \mathcal{L}$ is pointwise ϵ-close to 3-wise independent with probability $1 - \exp(6\log(N) - \Theta(\epsilon^2 M/N))$.*

Proof. For any distinct $x_1, x_2, x_3 \in \mathfrak{D}$ and distinct $y_1, y_2, y_3 \in \mathfrak{D}$, we need to show that the number of $(h, \ell) \in \mathcal{H} \times \mathcal{L}$ satisfying

$$\Big(h(\ell(x_1)), h(\ell(x_2)), h(\ell(x_3))\Big) = (y_1, y_2, y_3) \tag{11}$$

is close to $\frac{MN^{\underline{2}}}{N^{\underline{3}}} = \frac{M}{N-2}$. Equivalently, Eq. (11) can be written as

$$\Big(h^{-1}(y_1), h^{-1}(y_2), h^{-1}(y_3)\Big) = \Big(\ell(x_1), \ell(x_2), \ell(x_3)\Big).$$

By the nature of $\mathcal{L}$, the triples $\big(\ell(x_1), \ell(x_2), \ell(x_3)\big)$ are all distinct.[11] Thus it suffices show that the number of $h \in \mathcal{H}$ such that

$$\Big(h^{-1}(y_1), h^{-1}(y_2), h^{-1}(y_3)\Big) \in \Big\{\big(\ell(x_1), \ell(x_2), \ell(x_3)\big) \;\Big|\; \ell \in \mathcal{L}\Big\} \tag{12}$$

is close to $\frac{M}{N-2}$. The set in the right-hand side of (12) is of size $N^{\underline{2}}$. A random permutation h satisfies (12) with probability $\frac{N^{\underline{2}}}{N^{\underline{3}}} = \frac{1}{N-2}$.

$\mathcal{H}$ consists of M randomly sampled permutations, let $\#h$ temporarily denote the number of $h \in \mathcal{H}$ that satisfies (12). By Bernstein's bound,

$$\Pr\left[\left|\#h - \frac{M}{N-2}\right| > \tau\right] \le 2 \cdot \exp\Big(-\frac{\frac{1}{2}\tau^2}{\mathrm{Var}[\#h] + \frac{1}{3}\tau}\Big).$$

As $\#h$ is sum of independent Bernoulli, $\mathrm{Var}[\#h] \le \mathbb{E}[\#h] = \frac{M}{N-2}$, thus

$$\Pr\left[\left|\#h - \frac{M}{N-2}\right| > \epsilon\frac{M}{N-2}\right] \le 2 \cdot \exp\Big(-\frac{\frac{1}{2}\epsilon^2}{1 + \frac{1}{3}\epsilon}\frac{M}{N-2}\Big).$$

The proof is finished by a union bound over all $x_1, x_2, x_3, y_1, y_2, y_3$. □

Lemma B.3. *Assume $\mathcal{F}$ is a permutation distribution that is pointwise $(\lambda - 1)$-close to t-wise independent. Sample $\mathcal{H}$ as a collection of M permutations. Let $\mathcal{H}$ also denote the uniform distribution over $\mathcal{H}$. Then $\mathcal{H} \circ \mathcal{F}$ is pointwise ϵ-close to $(t+1)$-wise independent with probability $1 - \exp(2(t+1)\log(N) - \Theta(\epsilon^2 M/\lambda N))$.*

Proof. Consider any distinct $x_1, \ldots, x_{t+1} \in \mathfrak{D}$ and distinct $y_1, \ldots, y_{t+1} \in \mathfrak{D}$, we need to show that

$$\Pr_{h \leftarrow \mathcal{H}, f \leftarrow \mathcal{F}}\Big[h(f(x_1)) = y_1, \ldots, h(f(x_{t+1})) = y_{t+1}\Big] \tag{13}$$

is close to $\frac{1}{N^{\underline{t+1}}}$. Note that (13) equals to

$$\frac{1}{M}\sum_{h \in \mathcal{H}} \Pr_{f \leftarrow \mathcal{F}}\Big[f(x_1) = h^{-1}(y_1), \ldots, f(x_{t+1}) = h^{-1}(y_{t+1})\Big].$$

[11] A stronger statement holds: for any two distinct $\ell, \ell' \in \mathcal{L}$, we have $(\ell(x_1), \ell(x_2)) \neq (\ell'(x_1), \ell'(x_2))$.

Define p_h as

$$p_h := \Pr_{f \leftarrow \mathcal{F}}\Big[f(x_1) = h^{-1}(y_1), \ldots, f(x_{t+1}) = h^{-1}(y_{t+1})\Big].$$

Obviously $p_h \geq 0$. As $\mathcal{F}$ is pointwise $(\lambda - 1)$-close to t-wise independent, $p_h \leq \frac{\lambda}{N^{\underline{t}}}$.[12] For a random permutation h, we have $\mathbb{E}_h[p_h] = \frac{1}{N^{\underline{t+1}}}$ and $\mathrm{Var}_h[p_h] \leq \frac{\lambda}{N^{\underline{t}} \cdot N^{\underline{t+1}}}$.

$\mathcal{H}$ consists of M randomly sampled permutations, by Bernstein's bound,

$$\Pr\Big[\Big|\sum_{h \in \mathcal{H}} p_h - \frac{M}{N^{\underline{t+1}}}\Big| > \tau\Big] \leq 2 \cdot \exp\Big(-\frac{\frac{1}{2}\tau^2}{\frac{\lambda M}{N^{\underline{t}} \cdot N^{\underline{t+1}}} + \frac{1}{3}\frac{\lambda \tau}{N^{\underline{t}}}}\Big).$$

The case we care about is when $\tau = \frac{\epsilon M}{N^{\underline{t+1}}}$, in such case

$$\Pr\Big[\Big|\frac{1}{M}\sum_{h \in \mathcal{H}} p_h - \frac{1}{N^{\underline{t+1}}}\Big| > \frac{\epsilon}{N^{\underline{t+1}}}\Big] \leq 2 \cdot \exp\Big(-\frac{\frac{1}{2}(\frac{\epsilon M}{N^{\underline{t+1}}})^2}{\frac{\lambda M}{N^{\underline{t}} \cdot N^{\underline{t+1}}} + \frac{1}{3}\frac{\lambda}{N^{\underline{t}}}\frac{\epsilon M}{N^{\underline{t+1}}}}\Big) = 2 \cdot \exp\Big(-\frac{\frac{1}{2}\epsilon^2}{\lambda(1 + \frac{1}{3}\epsilon)} \cdot \frac{M}{N - t}\Big).$$

Recall that $\frac{1}{M}\sum_{h \in \mathcal{H}} p_h$ equals to (13).

The proof is finished by a union bound over all $x_1, \ldots, x_{t+1}, y_1, \ldots, y_{t+1}$. □

C Alternative Proof of the Extraction Lemma

Our extraction lemma (Lemma 3.1) assumes a specific arithmetic structure on the S-boxes, namely, the output (difference) of the S-box is a random vector orthogonal to the input (difference). This section, shows that the arithmetic structure required in the extraction lemma is not necessary. We will present an alternative proof of the extraction lemma. The core of the new proof is to bound Rényi entropy and collision probability, both of which are essentially the ℓ_2-norm. This also explains why the Fourier analysis is so effective and gives such a clean and succinct proof of the extraction lemma.

Lemma C.1. *Let k, b be positive integers and $n = bk$. Let Ω be a finite set of size 2^b. For each $v \in \Omega$, there is a set $S_v \subseteq \Omega$ satisfying,*

$$|S_v| \geq 2^{b-1} \text{ for all } v \in \Omega,$$
$$\frac{|S_v \cap S_u|}{|S_v| \cdot |S_u|} \leq \frac{1}{2^b} \text{ for all distinct } v, u \in \Omega.$$

Let $\mathcal{D}$ be a distribution over Ω^k and consider the following probabilistic process called $\mathsf{Samp}_{\mathcal{D}}$.

1. *Sample $(v_1, \ldots, v_k) \leftarrow \mathcal{D}$.*
2. *For each $i \in [k]$, sample $x_i \leftarrow S_{v_i}$ independently at random, and output $(x_1, \ldots, x_k)$.*

For any $T \subseteq [k]$, let v_T denote the concatenation of $(v_i)_{i \in T}$, let $\mathcal{D}_T$ denote the distribution of v_T, let $H_2[\mathcal{D}_T]$ denote its Rényi entropy. Then, the statistical distance between the joint distribution of $(x_1, \ldots, x_k)$ and the uniform distribution over Ω^k is at most

$$\frac{1}{2}\sqrt{\sum_{T \subseteq [k], T \neq \varnothing} 2^{-H_2[\mathcal{D}_T]}}.$$

[12] A stronger statement holds: for any $z_1, \ldots, z_t$, we have $\Pr_{f \leftarrow \mathcal{F}}[f(x_1) = z_1, \ldots, f(x_t) = z_t] \leq \frac{\lambda}{N^{\underline{t}}}$.

Lemma C.1 implies the extraction lemma (Lemma 3.1). The two lemmas have the same conclusion, while the hypothesis of the former is weaker. Let set Ω be $\mathbb{F}_2^b$. Let each $S_v := \{0, v\}^\perp$ be the subspace orthogonal to v. Then the hypothesis of Lemma C.1 is satisfied:

- $|S_v| = 2^{b-1}$ for all $v \neq 0$; and $|S_0| = 2^b$.
- $\frac{|S_u \cap S_v|}{|S_u| \cdot |S_v|} = \frac{2^{b-2}}{2^{b-1} \cdot 2^{b-1}} = \frac{1}{2^b}$ for non-zero distinct u, v; and $\frac{|S_u \cap S_0|}{|S_u| \cdot |S_0|} = \frac{|S_u|}{|S_u| \cdot 2^b} = \frac{1}{2^b}$ for non-zero u.

Proof of Lemma C.1. Assume $(x'_1, \ldots, x'_k)$ is independently sampled from $\mathsf{Samp}_{\mathcal{D}}$ and $(v'_1, \ldots, v'_k)$ is the intermediate variable. Then the connection between Rényi entropy and collision probability implies that $\Pr\left[v_T = v'_T\right] = 2^{-H_2[\mathcal{D}_T]}$ for each $T \subseteq [k]$ and $\Pr\left[(x_1, \ldots, x_k) = (x'_1, \ldots, x'_k)\right] = 2^{-H_2[x_1, \ldots, x_k]}$.

We start by expending the collision probability using the law of total probability,

$$\begin{aligned} &\Pr\left[(x_1, \ldots, x_k) = (x'_1, \ldots, x'_k)\right] \\ &= \sum_{T \subseteq [k]} \Pr\left[(x_1, \ldots, x_k) = (x'_1, \ldots, x'_k) \;\middle|\; \forall i, v_i = v'_i \iff i \in T\right] \Pr\left[\forall i, v_i = v'_i \iff i \in T\right]. \end{aligned} \tag{14}$$

For each $T \subseteq [k]$,

$$\begin{aligned} &\Pr\left[(x_1, \ldots, x_k) = (x'_1, \ldots, x'_k) \;\middle|\; \forall i, v_i = v'_i \iff i \in T\right] \\ &= \prod_{i \in T} \Pr\left[x_i = x'_i \mid v_i = v'_i\right] \cdot \prod_{i \notin T} \Pr\left[x_i = x'_i \mid v_i \neq v'_i\right] \\ &\leq \prod_{i \in T} \frac{1}{2^{b-1}} \cdot \prod_{i \notin T} \frac{1}{2^b} \\ &= \frac{2^{|T|}}{2^{bk}}. \end{aligned}$$

Thus Equation (14) becomes

$$\Pr\left[(x_1, \ldots, x_k) = (x'_1, \ldots, x'_k)\right] \leq \sum_{T \subseteq [k]} \frac{2^{|T|}}{2^{bk}} \Pr\left[\forall i, v_i = v'_i \iff i \in T\right].$$

Then hinted by the inclusion-exclusion principle,

$$
\begin{aligned}
&\Pr\Big[(x_1,\ldots,x_k)=(x_1',\ldots,x_k')\Big] \\
\le &\sum_{T\subseteq[k]} \frac{2^{|T|}}{2^{bk}} \Pr\Big[\forall i, v_i = v_i' \iff i\in T\Big] \\
= &\frac{1}{2^{bk}} \sum_{T\subseteq[k]} \sum_{T'\subseteq T} \Pr\Big[\forall i, v_i = v_i' \iff i\in T\Big] \\
= &\frac{1}{2^{bk}} \sum_{T'\subseteq[k]} \sum_{T'\subseteq T\subseteq[k]} \Pr\Big[\forall i, v_i = v_i' \iff i\in T\Big] \\
= &\frac{1}{2^{bk}} \sum_{T'\subseteq[k]} \Pr\Big[\forall i\in T', v_i = v_i'\Big] \\
= &\frac{1}{2^{bk}} \sum_{T'\subseteq[k]} 2^{-H_2[\mathcal{D}_{T'}]}.
\end{aligned}
\tag{15}
$$

Similar to the proof of Lemma 3.1, we define f as the probability mass function of $(x_1,\ldots,x_n)$ and let u denote the uniform distribution over Ω^k. Equation (15) essentially shows an upper bound of $\|f\|_2^2$. The ℓ_2-norm is tightly related to the ℓ_2 distance from uniform.

$$
\|f-u\|_2^2 = \|f\|_2^2 - \frac{1}{|\Omega|^k} \le \frac{1}{2^{bk}} \sum_{T\subseteq[k]} 2^{-H_2[\mathcal{D}_T]} - \frac{1}{2^{bk}} = \frac{1}{2^{bk}} \sum_{T\subseteq[k], T\neq\varnothing} 2^{-H_2[\mathcal{D}_T]}.
$$

The proof concludes with Cauchy-Schwartz.

$$
d_{TV}(f,u) = \frac{1}{2}\|f-u\|_1 \le \frac{1}{2}\sqrt{2^{kb}\|f-u\|_2^2} = \frac{1}{2}\sqrt{\sum_{T\subseteq[k], T\neq\varnothing} 2^{-H_2[\mathcal{D}_T]}}. \qquad \square
$$