Layout Graphs, Random Walks and the t-Wise Independence of SPN Block Ciphers

Tianren Liu[1(✉)], Angelos Pelecanos[2], Stefano Tessaro[3], and Vinod Vaikuntanathan[4]

[1] Peking University, Beijing, China
liutianren@gmail.com
[2] UC Berkeley, Berkeley, USA
[3] University of Washington, Seattle, USA
[4] MIT CSAIL, Cambridge, USA

Abstract. We continue the study of t-wise independence of substitution-permutation networks (SPNs) initiated by the recent work of Liu, Tessaro, and Vaikuntanathan (CRYPTO 2021).

Our key technical result shows that when the S-boxes are *randomly and independently chosen* and kept secret, an r-round SPN with input length $n = b \cdot k$ is $2^{-\Theta(n)}$-close to t-wise independent within $r = O(\min\{k, \log t\})$ rounds for any t almost as large as $2^{b/2}$. Here, b is the input length of the S-box and we assume that the underlying mixing achieves maximum branch number. We also analyze the special case of AES parameters (with random S-boxes), and show it is 2^{-128}-close to pairwise independent in 7 rounds. Central to our result is the analysis of a random walk on what we call the *layout graph*, a combinatorial abstraction that captures equality and inequality constraints among multiple SPN evaluations.

We use our technical result to show concrete security bounds for SPNs with actual block cipher parameters and *small-input S-boxes*. (This is in contrast to the large body of results on ideal-model analyses of SPNs.) For example, for the censored-AES block cipher, namely AES with most of the mixing layers removed, we show that 192 rounds suffice to attain 2^{-128}-closeness to pairwise independence. The prior such result for AES (Liu, Tessaro and Vaikuntanathan, CRYPTO 2021) required more than 9000 rounds.

1 Introduction

The design of block ciphers like the Advanced Encryption Standard (AES) is one of the most central topics in practical cryptography. Our confidence in their security stems from decades of cryptanalysis, spanning a wide range of attacks including linear [38] and differential [4] cryptanalysis, higher-order [33], truncated [31] and impossible [30] differential attacks, interpolation [25] and algebraic attacks [13], integral cryptanalysis [32], biclique attacks [5], and so on. These attacks have so far failed to make a dent in the conjectured security of AES as a (fixed-parameter) pseudorandom permutation. Nonetheless, we remain very

H. Handschuh and A. Lysyanskaya (Eds.): CRYPTO 2023, LNCS 14083, pp. 694–726, 2023.
https://doi.org/10.1007/978-3-031-38548-3_23

far from rigorously justifying that security actually holds. Crucially, the design methodology behind most block ciphers iterates a very weak round function (too weak to achieve any meaningful security notion). It is not clear whether it is even possible to formulate a meaningful non-tautological assumption that implies the security of a block cipher within the classical framework of provable security.

t-wise independent ciphers. Facing the above limitations, this paper continues a line of work justifying the security of block ciphers against *restricted* classes of attacks, with a focus on *substitution permutation networks* (SPNs), an important class of block ciphers that includes AES. In particular, we build on top of recent work by Liu, Tessaro, and Vaikuntanathan (LTV) [37] that studies the t-wise independence of SPNs as a "catch-all" security property that prevents all t-input statistical attacks. (The notion was already studied earlier [7,24] for less standard block cipher constructions.)

We take a *quantitative* angle where, for a given t, we aim to know the smallest $\epsilon = \epsilon(r)$ for which an r-round SPN is $\epsilon(r)$-close to a t-wise independent permutation. The case $t = 2$ already implies, for a small enough ϵ, security against *linear* [38] and *differential* [4] attacks, which have (on their own) been the subject of hundreds of works. Similarly, security against degree-d higher-order differential attacks [33] follows when $t = 2^d$.

The results from [37] suffer however from two major limitations, which we aim to address here: First, they only prove *pairwise* independence of SPNs. Second, for AES-like parameters, their pairwise-independence bound effectively requires *thousands of rounds* to achieve meaningful security matching practical expectations. (Concretely, more than 9000.[1])

Our Contributions, in a Nutshell. In this work, we study the t-wise independence of SPNs when the S-boxes are *randomly chosen*, *independent*, and *secret*, and thus act as the actual secret keys. Unlike a number of recent works in the random S-box model (e.g., [10,18,40]), which assume the S-box inputs to be as large as the security parameter, here we target a scenario with *small-input* S-boxes (e.g., 8 bits, as in AES), which presents a unique challenge. Random S-box SPNs were for instance also studied by Baignères and Vaudenay [3], who quantified the linear and differential probabilities in the limit as the number of rounds goes to infinity. Here, instead, we prove *concrete* bounds for the *stronger* property of t-wise independence. A summary of our results is given in Table 1.

While it is interesting to study random S-boxes in their own right, as they have been used in actual ciphers (e.g., GOST [41] and AES variants [45]), we really want to derive conclusions for block ciphers with fixed S-boxes (as [37] did) from our results. An *optimistic* interpretation of our results is that random, secret, S-boxes yield a good heuristic approximation of the behavior of SPNs with a concrete S-box (e.g., the inversion map $x \mapsto x^{2^b-1}$ as in AES). But we also offer a more *pragmatic* interpretation, based on the fact that a random S-box can be approximated by the sequential composition of an actual S-box (where a key is XORed prior to each call). Our analyses in the random S-box model

[1] LTV prove that $6r$-round AES is $2^{r-1}(0.472)^r$-close to pairwise independent, which becomes smaller than 2^{-128} for $r \geq 1528$.

Table 1. Results for the t-wise independence of SPN* and AES*. Here, b is the length of the input to the S-box (the word length or block size), and k is the width for SPN* (equivalently, the number of parallel S-box invocations). All of the SPN* results assume a linear mixing layer with maximum branch number. The AES* result uses the AES mixing layer, $k = 16$, $b = 8$.

	Rounds	t	**Closeness**	**Theorem**
SPN*	2	$O(1)$	$2^{-\Omega(kb)}$	Theorem 2
	2	$2^{(0.499-1/(4k))b}$	2^{-b}	Theorem 3
	$O(k)$	$2^{(0.499-1/(4k))b}$	$2^{-\Omega(kb)}$	Theorem 3 + [28,39]
	$O(\log t)$	$2^{0.499b}$	$2^{-\Omega(kb)}$	Theorem 4
AES*	7	2	2^{-128}	Theorem 6
censored AES	192	2	2^{-128}	Theorem 7

therefore carry over to a *concrete* block cipher which can be thought of as an SPN with a number of mixing layers removed (what we refer to as a "censored" SPN or SPN*).

We now go back to our contributions in a bit more in detail.

Substitution-permutation Networks. To state our results more concretely, recall that a substitution permutation network (SPN) with *word length* b, *width* k, and r *rounds*, is defined by an invertible *substitution box* (or S-box) $S : \mathbb{F} \to \mathbb{F}$, where $\mathbb{F} = \mathbb{F}_{2^b}$, and an invertible *mixing layer* $M : \mathbb{F}^k \to \mathbb{F}^k$. (One usually focuses on *linear* mixing functions as we do in this paper.) Computation proceeds in r rounds, given input vector $\mathbf{x}^{(\mathsf{in})} = \mathbf{y}^{(0)} \in \mathbb{F}^k$ and round keys $\mathbf{k}^{(0)}, \ldots, \mathbf{k}^{(r)} \in \mathbb{F}^k$. For $i = 1, \ldots, r+1$ we compute

$$\mathbf{x}^{(i)} = \left[S\left(\mathbf{y}^{(i-1)}[1] + \mathbf{k}^{(i-1)}[1]\right), \ldots, S\left(\mathbf{y}^{(i-1)}[k] + \mathbf{k}^{(i-1)}[k]\right) \right] .$$

$$\mathbf{y}^{(i)} = M\mathbf{x}^{(i)}$$

The final output is $\mathbf{y}^{(\mathrm{out})} = \mathbf{x}^{(r+1)}$. See Fig. 1 for an illustration. (Note that in this representation, the final operation is the application of S-boxes, with no further mixing. This differs from some of the literature; however, the difference is inconsequential to our results.) In an actual block cipher, one would compute the round keys from a short key via a suitable key-scheduling algorithm, but here we follow the convention from prior works of using independent keys for the analysis.

Typical choices for the above parameters are those from AES, where $k = 16$ and $b = 8$, and one should think of these when assessing whether a result is meaningful.

t-wise Independence for Random S-Boxes. The bulk of our results will be concerned with the analysis of SPNs in a model where the S-boxes are ideal, i.e.,

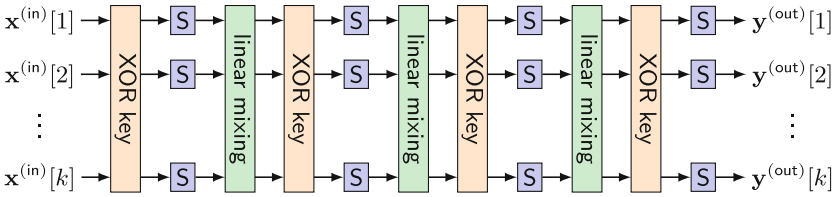


Fig. 1. Illustration of a 3-round SPN.

randomly chosen and secret. In other words, we replace the step

$$\mathbf{x}^{(i)}[j] \leftarrow S(\mathbf{y}^{(i-1)}[j] \oplus \mathbf{k}^{(i-1)}[j])$$

for $i = 1, \ldots, r+1$ and $j = 1, \ldots, k$ with

$$\mathbf{x}^{(i)}[j] \leftarrow S_j^{(i-1)}(\mathbf{y}^{(i-1)}[j])$$

where $S_j^{(i-1)}$ is a uniformly chosen random permutation on $\mathbb{F}$. Here, we can think of the S-box descriptions as part of a longer key, and following the notation from [3], we refer to this variant as SPN*.

Formally, we measure the proximity to t-wise independence by picking t arbitrary distinct input vectors and obtain the t output vectors processed by the r-round SPN* construction. We then give an upper bound on the statistical distance of these output vectors from t uniformly sampled, but distinct, vectors. As observed in [37], such a distance bound also gives explicit concrete bounds for the linear and differential probabilities. (In particular, our result gives concrete bounds for such quantities, as opposed to [3] which only shows eventual convergence to a particular probability as the number of rounds goes to infinity.)

Layouts and Random Walks. At the core of our results is the formalization of the concept of a *layout*, which allows us to reduce the question of t-wise independence to the analysis of a random walk which is entirely defined by the mixing layer M. Concretely, if we are given a t-tuple of vectors $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$, and map them to $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$ by applying the same k random S-boxes to each of the vectors, we observe that the mapping respects equality and inequality constraints. For example, if $\mathbf{y}_i[j] = \mathbf{y}_{i'}[j]$ for $i \neq i'$, then $\mathbf{x}_i[j] = \mathbf{x}_{i'}[j]$. Inequalities are also similarly preserved. A t-wise *layout* I is, formally, a description of equality/inequality constraints among t k-dimensional vectors over $\mathbb{F}$. Crucially, applying random S-boxes to *any* t-tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$ satisfying the layout I results in a t-tuple picked *uniformly at random* from the set of *all* t-tuples that satisfy the same layout I. For the special case of $t = 2$, a layout is equivalent to an *activity pattern* formulated and studied in the AES literature [1].

This means in particular that the evaluation of an r-round SPN* on t inputs corresponds to taking r random steps on the *layout* graph. We start with an arbitrary layout I_0, and step $i = 1, \ldots, r$ consists of:

– Picking a random t-tuple $(\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_t^{(i)})$ that lies in layout I_{i-1};
– Compute $\mathbf{y}_j^{(i)} = M\mathbf{x}_j^{(i)}$ for all $j = 1, \ldots, t$; and
– Set I_i to be the (unique) layout satisfied by $(\mathbf{y}_1^{(i)}, \ldots, \mathbf{y}_t^{(i)})$.

The convergence of this walk to the distribution over layouts induced by a uniformly sampled t-tuple of distinct vectors directly yields t-wise independence of the r-round SPN*. For the case $t = 2$, this random walk was also described in [3] without any explicit convergence guarantees, which we provide here.

We provide a careful analysis of this random walk by first characterizing the transition probability of going from a layout I to a layout J and then derive an upper bound on the distance from the stationary distribution after one single step, provided we start from a nice enough layout, i.e., one that does not induce too many collisions. Then, very roughly, one shows that a nice layout is reached in one round with very high probability. We use this analysis to derive a number of theorems, which all assume that the mixing layer achieves maximum branch number, i.e., for all $\mathbf{x} \in \mathbb{F}^k \setminus \{0\}$, we have $\mathsf{wt}(\mathbf{x}) + \mathsf{wt}(M\mathbf{x}) \geq k + 1$, where $\mathsf{wt}(\cdot)$ denotes Hamming weight, i.e., the number of non-zero components.

Our first two theorems give the smallest ϵ depending on whether t is small or large.

Theorem 2. 2-round SPN* is ε-close to t-wise independent, for $\varepsilon = \frac{t^2 \cdot 2^{k+1}}{(2^b)^{k/(2t)}} + t \cdot \left(\frac{8 \cdot t^3}{2^b}\right)^{k/2}$.

Theorem 3. For any $\alpha \in (0, 1]$, 2-round SPN* is ε-close to t-wise independent, where $\varepsilon = \frac{t^2}{\alpha \cdot 2^b} + t \cdot \left(\frac{(2t)^{2-\alpha}}{(2^b)^{1-\alpha}}\right)^k$.

A standard goal is to make ϵ equal $2^{-\Omega(k \cdot b)}$, as $n = k \cdot b$ is the input length of the SPN, and the first theorem implies that for small constant $t = O(1)$, we achieve distance $2^{-\Omega(n)}$ already after two rounds. In contrast, by picking the suitable α, the second theorem allows t to become almost as large as $2^{b/2}$ (concretely, we require $t < 2^{(0.499-1/(4k))b}$, which is as large as 14 for AES-like parameters), but only gives $\epsilon = 2^{-\Omega(b)}$. However, one can then amplify this using existing amplification results [28,39] to achieve $\epsilon = 2^{-\Omega(bk)}$ after $2k$ rounds.

We also show an alternative theorem that also yields $\epsilon = 2^{-\Omega(bk)}$, but this time using $O(\log t)$ rounds, instead of $O(k)$. This follows from the following.

Theorem 4. Let $t = 2^r$. Then, r-round SPN* is ε-close to 2^r-wise independent for $\varepsilon = \frac{t \cdot 2^{\frac{11}{4}}}{1 - 2^{-\frac{k}{4}}} \cdot \left(\frac{8 \cdot t^2}{2^b}\right)^{k/4}$ if $k > 4$.

The Case of AES. The specific case of AES is interesting because its mixing layer does not achieve the maximal branch number. One could in fact extend some of our techniques above to a more relaxed branch number. However, we give a more precise analysis of a variant of AES with random S-boxes which, unlike the above SPN*, uses the *actual* AES mixing layer (alternating the ShiftRows and MixColumn operations). It also sets $k = 16$ and $b = 8$. We refer to this

variant as AES*. We show that AES* is 2^{-128}-close to pairwise independent already for *seven* rounds. To achieve these results, we combine experimental computations with our random walk framework. We note that this result could have been obtained computationally also using results from [3], in particular their description of the random walk on layouts for the special case of AES* and $t = 2$. (Their description is however not sufficient to yield the results in the other sections of this paper, nor do they actually carry out the computation, or target a security property as strong as pairwise-independence.)

Concrete S-Boxes and Censored SPNs. For the special case of pairwise independence, one can easily transform our results for random S-boxes into results for concrete S-boxes if we are willing to replace the application of a *single* random S-box $S_j^{(i)}$ with the repeated application of the AES S-box (namely, the patched inversion function over $\mathbb{F}$) alternated with the addition of a key value prior to each S-box call. We refer to the resulting cipher as *censored SPN* (or *censored* AES), because it is equivalent to an SPN where a fraction of mixing layers have been removed (i.e., "censored"). We give a censored variant of AES which is 2^{-128}-close to pairwise independent after 192 rounds. We conjecture that 192-round of AES itself is also 2^{-128}-close to pairwise independent, i.e., the censoring mixing layers never increases security.

This should be contrasted with [37], which shows that AES is 2^{-128}-close to pairwise independent after (more than) 9000 rounds.

1.1 Related Work: The "Large" S-Box Model

A number of works [10,18,40] have considered SPNs with random S-boxes when the input length b is large (i.e., it can be thought of as the security parameter), and aims to prove an r-round SPN to be a (strong) pseudorandom permutation. Miles and Viola [40] deal with *secret* S-boxes (as we do here), whereas [10,18] consider a single public S-box (accessible as a random oracle) which is then keyed within the construction. (But clearly, this implies an analysis in a model where the S-box is secret.) These works fit within the bigger scope of a long line of works [2,6,8,9,11,12,15,16,19–23,35,36,44]) analyzing block cipher constructions in ideal models. A recent paper by Dodis, Karthikeyan, and Wichs [17] then suggests conjectures under which these large S-box analyses could imply security in the small S-box regime (for full pseudorandomness).

While the result is not explicitly stated, one can, in fact, apply the toolkit from [10], which in turn relies on the H-coefficient method [43], to show that a 1-round SPN is ϵ-close to t-wise independent for $\epsilon = O(kt^2/2^b)$. For $b = 8$ and $k = 16$, one might hope to achieve $\epsilon = 1/2$ for $t = 2$ (and in turn, this can be boosted using [39]), but the involved constants prevent that. In addition, we observe that this bound has the unnatural feature that it *degrades* as a function of the width parameter k, which is exactly what we show *not to be the case*. Our results adopt completely different techniques, that rely on the analysis of random walks on the layout graph, and indeed also indicate an improvement of the achievable ϵ as k grows, as intuition would suggest.

While (almost) t-wise independent permutations can be constructed in many other ways (see, e.g. [29]), that is not the point of this paper. Our goal is to analyze natural constructions, in this case following the substitution-permutation paradigm, which are *provably* almost t-wise independent and *plausibly* pseudorandom.

1.2 Technical Overview

In this overview, we briefly explain how our technique works in the special case of 2-wise (or pairwise) independence of SPN* (i.e., SPN with random S-boxes). A more detailed analysis of the pairwise setting can be found in Sect. 4. The more involved analysis of the general t-wise setting follows the same framework, and is presented in Sect. 5. Concrete bounds for censored AES are given in Sect. 6.

Differences and Layouts. As we only consider two inputs, we can follow the standard differential cryptanalysis approach of working with *differences*. For any input difference $\mathbf{x}_\Delta^{(\mathsf{in})} = \mathbf{x}_1^{(\mathsf{in})} - \mathbf{x}_2^{(\mathsf{in})}$, we need to show that the corresponding distribution of the output difference $\mathbf{y}_\Delta^{(\mathsf{out})} = \mathbf{y}_1^{(\mathsf{out})} - \mathbf{y}_2^{(\mathsf{out})}$ is close to uniform. We consider a two-round SPN*, so we can define analogously differences $\mathbf{x}_\Delta^{(1)}$, $\mathbf{y}_\Delta^{(1)}$, $\mathbf{x}_\Delta^{(2)}$, $\mathbf{y}_\Delta^{(2)}$, and $\mathbf{y}_\Delta^{(\mathsf{out})}$. See Fig. 2 for an illustration.

Let $I^{(0)}$ denote the layout of $(\mathbf{x}_1^{(\mathsf{in})}, \mathbf{x}_2^{(\mathsf{in})})$. In the pairwise setting, the layout can be defined as a subset $I^{(0)} \subseteq [k]$ including the coordinates where $\mathbf{x}_1^{(\mathsf{in})}, \mathbf{x}_2^{(\mathsf{in})}$ collide, or, equivalently, $I^{(0)}$ consists of all the coordinates where $\mathbf{x}_\Delta^{(\mathsf{in})}$ is zero. In general, we say $I \subseteq [k]$ is the layout of $\mathbf{x} \in \mathbb{F}^k$, or $\mathbf{x}$ is in layout I, if I consists precisely of the zero coordinates of $\mathbf{x}$. That is,

$$\mathbf{x} \text{ in } I \quad \text{means} \quad \forall i \in [k],\ i \in I \iff \mathbf{x}[i] = 0.$$

Due to the randomness of the S-boxes, $\mathbf{x}_\Delta^{(1)}$ is distributed uniformly among all vectors in layout $I^{(0)}$. Similarly, if we let $I^{(1)}$ (resp. $I^{(2)}$) denote the layout of $\mathbf{y}_\Delta^{(1)}$ (resp. $\mathbf{y}_\Delta^{(2)}$), then $\mathbf{x}_\Delta^{(2)}$ (resp. $\mathbf{y}_\Delta^{(\mathsf{out})}$) is distributed uniformly among all vectors in layout $I^{(1)}$ (resp. $I^{(2)}$).

It is easy to show that if $I^{(2)}$ is close to the distribution on layouts induced by a random (non-zero) vector, then the distribution of $\mathbf{y}_\Delta^{(\mathsf{out})}$ is close to uniform. Thus the heart of the analysis is to understand how the distribution of $I^{(r)}$ depends on that of $I^{(r-1)}$. Evidently, this depends on the characteristics of the linear mixing layer. In particular, we show the following lemma.

> **Lemma** 3 (informal). If $I^{(r-1)}$ is *nice* in the sense that $|I^{(r-1)}| \leq k/2$, then $I^{(r)}$ is $2^{-\Omega(kb)}$-close in variation distance to the layout of a random vector.

The Blueprint. We now use the above lemma to prove that 2-round SPN* is close to 2-wise independent using the following blueprint. All the error terms in the analysis have magnitude $2^{-\Omega(kb)}$.

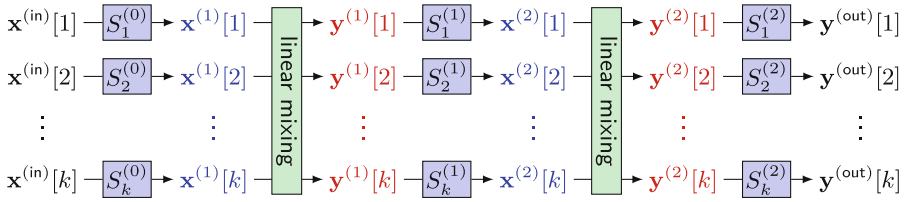


Fig. 2. Illustration of a 2-round SPN* Network. Each S-box is a uniformly random permutation from $\mathbb{F}$ to $\mathbb{F}$. These S-boxes form the key of the SPN* network.

In the first round: If $I^{(0)}$ is nice, then $I^{(1)}$ is statistically close to the layout of a random vector by Lemma 3 above, so $I^{(1)}$ is nice with high probability. If $I^{(0)}$ is not nice, then we claim that $I^{(1)}$ must be nice due to the fact that the linear mixing matrix M has maximal branch number. Recall that this guarantees $\mathsf{wt}(\mathbf{x}) + \mathsf{wt}(M\mathbf{x}) \geq k+1$ for all $\mathbf{0} \neq \mathbf{x} \in \mathbb{F}^k$. Thus, if $I^{(0)}$ is not nice, $I^{(1)}$ must be nice. In either case, $I^{(1)}$ is very likely to be nice.

In the second round: Since $I^{(1)}$ is very likely to be nice, $I^{(2)}$ is close to the layout of a random vector again by Lemma 3, which implies that $\mathbf{y}_{\Delta}^{(\mathsf{out})}$ is close to uniform.

Our analysis of the t-wise setting in Sect. 5 follows the same high-level framework, which requires in particular generalizing the notion of a layout and its niceness.

Proof Sketch of Lemma 3. The rest of this overview provides a proof sketch of the lemma. The transition probability from $I^{(r-1)}$ to $I^{(r)}$ can be written as

$$\Pr\left[I^{(r)} = J \;\middle|\; I^{(r-1)} = I\right] = \Pr_{\mathbf{x}\,\mathsf{in}\,I}[M\mathbf{x}\ \mathsf{in}\ J] = \frac{\#\{\mathbf{x}\ \text{s.t.}\ \mathbf{x}\ \mathsf{in}\ I \wedge M\mathbf{x}\ \mathsf{in}\ J\}}{\#\{\mathbf{x}\ \text{s.t.}\ \mathbf{x}\ \mathsf{in}\ I\}}\,.$$

Define an indicator function $\mathbb{1}_M$ where $\mathbb{1}_M(\mathbf{x},\mathbf{y}) = 1$ if and only if $M\mathbf{x} = \mathbf{y}$. Then

$$\Pr\left[I^{(r)} = J \;\middle|\; I^{(r-1)} = I\right] = \frac{\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\mathbb{1}_M(\mathbf{x},\mathbf{y})}{\sum_{\mathbf{x}\,\mathsf{in}\,I}1}\,. \tag{1}$$

To compute the numerator, it turns out that it is convenient to relax the notion of being in a layout. In particular, we say that $\mathbf{x}$ *satisfies* layout I as follows:

$$\mathbf{x}\ \mathsf{SAT}\ I \quad \text{means} \quad \forall i \in [k],\ i \in I \implies \mathbf{x}[i] = 0.$$

In particular, if $\mathbf{x}$ is in layout I, it satisfies layout I, but not vice versa.

Note that if M has the maximal branch number, then one can show that

$$\sum_{\mathbf{x}\,\mathsf{SAT}\,I}\sum_{\mathbf{y}\,\mathsf{SAT}\,J}\mathbb{1}_M(\mathbf{x},\mathbf{y}) = \begin{cases}(2^b)^{k-|I|-|J|} & \text{if } |I|+|J| \leq k\,,\\ 1 & \text{if } |I|+|J| > k\,.\end{cases} \tag{2}$$

Also, note that

$$\sum_{\mathbf{x}\,\mathsf{SAT}\,I}\sum_{\mathbf{y}\,\mathsf{SAT}\,J}\frac{1}{(2^b)^k}=(2^b)^{k-|I|-|J|} \tag{3}$$

is very close to (2), off by at most 1 for any I and J. In order to express the numerator of (1) in closed form, we first note that (2) and (3) should remain close if the sum operator is replaced by $\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}$. That is

$$\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\left(\mathbb{1}_M(\mathbf{x},\mathbf{y})-\frac{1}{(2^b)^k}\right)=O(2^{2k}).$$

This can be verified by the inclusion-exclusion principle (details in Sect. 4.1).

Plugging it in (1) gives a good bound on the transition probability

$$\Pr\left[I^{(r)}=J\;\middle|\;I^{(r-1)}=I\right]=\frac{\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k}+O(2^k)}{\sum_{\mathbf{x}\,\mathsf{in}\,I}1}=\overbrace{\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k}}^{=\Pr_{\mathbf{y}}[\mathbf{y}\,\mathsf{in}\,J]}+\overbrace{\frac{O(2^{2k})}{(2^b-1)^{k-|I|}}}^{\text{err}}.$$

The error term is of the order of $2^{-\Omega(kb)}$ if I is nice (i.e., $|I|\le k/2$). The transition probability is close to $\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k}$, which is the probability that a random vector lies in J. This can then be turned into a bound on the statistical distance to conclude the proof of the lemma.

2 Preliminaries

For any positive integer n, let $[n]$ denote the set $\{1,2,\ldots,n\}$. We will use boldface letters such as $\mathbf{x}$ to denote vectors and will denote the i^{th} coordinate of such a vector by $\mathbf{x}[i]$. For an integer $b\ge 1$, we let $\mathbb{F}_{2^b}$ denote the finite field of size 2^b. We also denote a finite field by $\mathbb{F}$ when the field size is clear from the context.

2.1 Substitution-Permutation Networks (SPN)

A *Substitution-Permutation Network* (SPN) is parameterized by the number of rounds, denoted by r; the word length, denoted by b; the width parameter, denoted by k; the linear mixing permutation, a full rank matrix $M:(\mathbb{F}_{2^b})^k\to(\mathbb{F}_{2^b})^k$; and an S-box permutation $S:\mathbb{F}_{2^b}\to\mathbb{F}_{2^b}$. All these parameters are public. The network is a keyed permutation over $\mathbb{F}_{2^b}^k$, so every input (output) vector is bk-bit long. The key is a tuple of $r+1$ (meant to be uniformly random) vectors $\mathbf{k}_0,\mathbf{k}_1,\ldots,\mathbf{k}_r\in(\mathbb{F}_{2^b})^k$. The "independent round keys" assumption here is very common and rooted in the model of Markov Ciphers from the seminal works of Lai, Massey, and Murphy [34], Nyberg [42] and follow-ups. We follow the convention that the number of rounds is the same as the number of mixing layers. In Fig. 1, we give an illustration of a 3-round SPN.

SPN with Random Secret S-boxes (SPN).* Much of this work will deal with SPN networks where each S-box is chosen independently at random from the set of all permutations on $\mathbb{F} := \mathbb{F}_{2^b}$, and kept secret. In this case, the set of S-boxes acts as the key, and there is no reason to have a separate addition of round keys. Thus, the key of the network consists of $k(r+1)$ permutations $S_j^{(i)} : \mathbb{F} \to \mathbb{F}$ (for $0 \le i \le r, 1 \le j \le k$).

Given input $\mathbf{x}^{(\mathsf{in})} = \mathbf{y}^{(0)} \in \mathbb{F}^k$ and the key, the output $\mathbf{y}^{(\mathsf{out})} = \mathbf{x}^{(r+1)} \in \mathbb{F}^k$ is determined by alternating the following two steps, as illustrated in Fig. 2. For consistency, we let $\mathbf{y}^{(0)}$ be another name for $\mathbf{x}^{(\mathsf{in})}$ and let $\mathbf{x}^{(r+1)}$ be another name for $\mathbf{y}^{(\mathsf{out})}$.

Substitution Step-i ($0 \le i \le r$) For $1 \le j \le k$, let $\mathbf{x}^{(i+1)}[j] = S_{i,j}(\mathbf{y}^{(i)}[j])$,
Permutation Step-i ($1 \le i \le r$) Let $\mathbf{y}^{(i)} = M\mathbf{x}^{(i)}$.

We call $\mathbf{x}^{(i)}$ and $\mathbf{y}^{(i)}$ the intermediate values of the i-th round. Then the input $\mathbf{x}^{(\mathsf{in})}$, also called $\mathbf{y}^{(0)}$, is in "the 0-th round". This gets fed into the substitution step-0 which produces $\mathbf{x}^{(1)}$. Permutation step-i is inside the i-th round. Substitution step-i is the boundary between the i-th round and the $(i+1)$-th round. The output $\mathbf{y}^{(\mathsf{out})}$, also called $\mathbf{x}^{(r+1)}$, is in "the $(r+1)$-th round".

Branch number. We use the definition of the branch number of a matrix that quantifies how well the linear layer "mixes" its input.

Definition 1. *The branch number of a matrix $M \in (\mathbb{F}_{2^b})^{k \times k}$ is defined to be*

$$\mathsf{br}(M) = \mathsf{min}_{0 \neq \alpha \in (\mathbb{F}_{2^b})^k} (\mathsf{wt}(\alpha) + \mathsf{wt}(M\alpha))$$

where wt *denotes the Hamming weight.*

Having the maximal branch number (namely, $k+1$) is considered a desirable feature for mixing functions [14,27].

Summary of notations. The intermediate states in an SPN (or SPN*) network are denoted by boldface letters $\mathbf{x}$ or $\mathbf{y}$. The notation $\mathbf{x}^{(r)}$ (resp. $\mathbf{y}^{(r)}$) is used to denote the state at round r; and $\mathbf{x}^{(r)}[s]$ denotes the s^{th} coordinate of $\mathbf{x}^{(r)}$. When dealing with multiple inputs, we let the subscript denote which input we are referring to: i.e., $\mathbf{x}_i^{(r)}$ denotes round-r state of the i^{th} input. We let $\mathbf{x}_{1:t}^{(r)} = (\mathbf{x}_i^{(r)})_{i \in [t]} = (\mathbf{x}_1^{(r)}, \ldots, \mathbf{x}_t^{(r)})$ be a shorthand for a tuple of vectors.

3 Layouts

This section introduces *layout*, a key notion of this paper. In the pairwise setting, layout is similar to the notions of an activity pattern [26] or support [3] of an input that have been formulated in the literature in the context of differential and linear cryptanalysis. Our notion considers the generalized setting and deals with t-tuples of inputs for an arbitrary t.

Motivation. Given t inputs $\mathbf{x}_1^{(\mathsf{in})}, \ldots, \mathbf{x}_t^{(\mathsf{in})}$ to an SPN* network, we want to characterize the joint distribution of the outputs $\mathbf{y}_1^{(\mathsf{out})}, \ldots, \mathbf{y}_t^{(\mathsf{out})}$ when all the S-boxes are i.i.d. uniform. The evaluation of the SPN* on these t inputs is essentially a Markov chain. The dependency between the intermediate values can be illustrated by the following Bayesian network.

$$\mathbf{x}_{1:t}^{(\mathsf{in})} \longrightarrow \mathbf{x}_{1:t}^{(1)} \longrightarrow \mathbf{y}_{1:t}^{(1)} \longrightarrow \mathbf{x}_{1:t}^{(2)} \longrightarrow \mathbf{y}_{1:t}^{(2)} \longrightarrow \cdots$$

Here $\mathbf{x}_{1:t}^{(r)}$ denotes the tuple of t vectors $(\mathbf{x}_1^{(r)}, \ldots, \mathbf{x}_t^{(r)})$, and so does $\mathbf{y}_{1:t}^{(r)}$.

The tuple $\mathbf{y}_{1:t}^{(r)}$ depends deterministically on $\mathbf{x}_{1:t}^{(r)}$ via the permutation step. The substitution step is more interesting. The randomness of the substitution step-r consists of k S-boxes $S_1^{(r)}, \ldots, S_k^{(r)}$. Each S-box $S_s^{(r)}$ is applied to the corresponding coordinate for all inputs, namely, $\mathbf{y}_i^{(r)}[s]$ for all $i \in [t]$. The substitution step erases most information, but some are preserved. In particular,

- $\mathbf{y}_i^{(r)}[s] = \mathbf{y}_j^{(r)}[s]$ if and only if $\mathbf{x}_i^{(r+1)}[s] = \mathbf{x}_j^{(r+1)}[s]$.

And it is not hard to verify that this is the only information preserved. In particular, the distribution of $\mathbf{x}_{1:t}^{(r+1)}$ is uniform among all tuples that satisfy

$$\forall i, j \in [t],\ \forall s \in [k],\ \mathbf{x}_i^{(r+1)}[s] = \mathbf{x}_j^{(r+1)}[s] \iff \mathbf{y}_i^{(r)}[s] = \mathbf{y}_j^{(r)}[s].$$

To capture and formalize these constraints, we introduce the notion of a *layout* below. The layout of t vectors $\mathbf{x}_{1:t}$ should specify whether $\mathbf{x}_i[s] = \mathbf{x}_j[s]$, for any $i, j \in [t], s \in [k]$.

Definition 2 (layouts). *A t-wise layout I is defined as $I = (I_{i,j})_{1 \le i < j \le t}$. Each $I_{i,j}$ is a subset of $[k]$. For a tuple of t vectors $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t) \in (\mathbb{F}^k)^t$, we say that the tuple is in a layout I, denoted by $\mathbf{x}_{1:t}$ in I, if*

$$\forall 1 \le i < j \le t,\ \forall s \in [k],\ s \in I_{i,j} \iff \mathbf{x}_i[s] = \mathbf{x}_j[s].$$

We say I is the layout of $\mathbf{x}_{1:t}$, denoted by $\mathrm{layout}(\mathbf{x}_{1:t}) = I$, *if $\mathbf{x}_{1:t}$ is in layout I.*

We also define a weaker notion: say $\mathbf{x}_{1:t}$ satisfies *a layout I, denoted by $\mathbf{x}_{1:t}$ SAT I, if and only if*

$$\forall 1 \le i < j \le t,\ \forall s \in [k],\ s \in I_{i,j} \implies \mathbf{x}_i[s] = \mathbf{x}_j[s].$$

Given another layout $J = (J_{i,j})_{1 \le i < j \le t}$, we say J is stricter or equal to I, denoted by $J \supseteq I$ or $I \subseteq J$, if

$$\forall 1 \le i < j \le t,\ J_{i,j} \supseteq I_{i,j}.$$

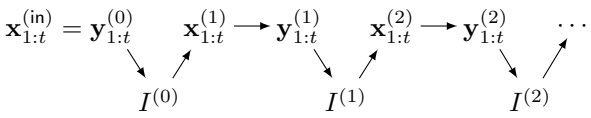
Example 1. Consider the 3-wise layout $I = (I_{1,2}, I_{1,3}, I_{2,3}) = (\{1\}, \{2\}, \{3\})$. Then, the tuple of vectors $\mathbf{x}_1 = [a, b, c']$, $\mathbf{x}_2 = [a, b', c]$, and $\mathbf{x}_3 = [a', b, c]$ lay in the layout I a long as $a \neq a'$, $b \neq b'$, $c \neq c'$.

Note that not all layouts are "valid". For example,

$$I = (I_{1,2}, I_{1,3}, I_{2,3}) = (\{1\}, \varnothing, \{1\}).$$

is not the layout of any 3-tuple. Because $1 \in I_{1,2}$ means the first two vectors agree on coordinate 1, and $1 \in I_{2,3}$ means the last two vectors agree on coordinate 1, by transitivity, these imply $1 \in I_{1,3}$. We say a layout I is *valid* if for all $s \in [k]$ and for all $i < i' < i''$, if any two of $I_{i,i'}, I_{i,i''}, I_{i',i''}$ contain s, so does the third one.

Random Walks on Layouts. Using the notion of layouts, the distribution of $\mathbf{x}_{1:t}^{(r+1)}$ conditioned on $\mathbf{y}_{1:t}^{(r)}$ can be described more concisely: the substitution step simply samples a random $\mathbf{x}_{1:t}^{(r+1)}$ who is in the same layout as $\mathbf{y}_{1:t}^{(r)}$. In other words, the substitution step is equivalent to a two-step process: first extract the layout of $\mathbf{y}_{1:t}^{(r)}$, then sample a random tuple from the layout. If letting $I^{(r)}$ denote the layout of $\mathbf{y}_{1:t}^{(r)}$ (and also $\mathbf{x}_{1:t}^{(r+1)}$, since they are in the same layout), the Bayesian network of the SPN* evaluation can also be written in the following way:



This Bayesian network view through the lens of layouts suggests that the right problem to study is the transition probability from $I^{(r)}$ to $I^{(r+1)}$ (induced by the linear mixing layer). This transition probability could be easier to characterize since the space of all layouts is much smaller than the space of all t-tuples.

All theorems in this paper follow this framework. They essentially prove the following statement: Starting from any layout $I^{(0)}$, after some r rounds, the distribution of $I^{(r)}$ is close to t-wise independent. To complete the framework, we need to answer two questions: 1) What is the definition of a layout being close to t-wise independent; and 2) How does a layout being close to t-wise independent imply that a random tuple in the layout is close to t-wise independent?

Definition 3 (closeness to t-wise independence). *Let $\mathbf{z}_1, \ldots, \mathbf{z}_t$ be sampled uniformly at random from $\mathbb{F}^k$ with (resp. without) replacement. Then we say the tuple $(\mathbf{z}_1, \ldots, \mathbf{z}_t)$ is t-wise independent with (resp. without) replacement.*

Let $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$ be sampled from a distribution. We say $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$ is ε-close to t-wise independent with (resp. without) replacement if

$$\Delta_{TV}\Big((\mathbf{x}_1, \ldots, \mathbf{x}_t)(\mathbf{z}_1, \ldots, \mathbf{z}_t)\Big) \leq \varepsilon.$$

Let layout I be sampled from a distribution. We say I is ε-close to t-wise independent with (resp. without) replacement if

$$\Delta_{TV}\Big(I, \mathrm{layout}(\mathbf{z}_1, \ldots, \mathbf{z}_t)\Big) \leq \varepsilon.$$

We say a keyed permutation (e.g., a SPN) is ε-close to t-wise independent with (resp. without) replacement if for any t distinct input $\mathbf{x}_{1:t}^{(in)}$, the joint distribution of the t corresponding output $\mathbf{y}_{1:t}^{(out)}$ is ε-close to t-wise independent with (resp. without) replacement, assuming the key is sampled properly.*

The following lemma and its corollary show how the distribution of t-tuples is related to the distribution of their layouts, and justify why this 'layout' analysis suffices for our purposes of proving t-wise independence. Their proofs are deferred to the full version of the paper.

Lemma 1. *Assume I and $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$ jointly come from a distribution where $\mathbf{x}_{1:t}$ is a random tuple in I when conditioning on I, and similarly for J and $\mathbf{z}_{1:t}$. Then*

$$\Delta_{TV}(I, J) = \Delta_{TV}(\mathbf{x}_{1:t}, \mathbf{z}_{1:t}).$$

Corollary 1. *Suppose I is sampled from a distribution and $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$ is sampled uniformly within layout I. Then $\mathbf{x}_{1:t}$ is ε-close to t-wise independent if and only if I is ε-close to t-wise independent.*

4 Warm-Up: 2-Wise Independence of 2-Round SPN*

In this section, we present the core idea of our new technique and demonstrate its power by showing that a 2-round SPN* is $2^{-\Theta(kb)}$-close to 2-wise independent. That is, we show that for any two distinct inputs $(\mathbf{x}_1^{(\mathsf{in})}, \mathbf{x}_2^{(\mathsf{in})})$ (which is the same as $(\mathbf{y}_1^{(0)}, \mathbf{y}_2^{(0)})$) the joint distribution of their corresponding outputs $(\mathbf{y}_1^{(\mathsf{out})}, \mathbf{y}_2^{(\mathsf{out})})$ (which is the same as $(\mathbf{x}_1^{(3)}, \mathbf{x}_2^{(3)})$) is close to 2-wise independent.

Theorem 1. *2-round SPN* is ε-close to 2-wise independent, where*

$$\varepsilon \leq \frac{3^k}{(2^{b-1})^{k/2}},$$

if its linear mixing function has maximal branch number (see Definition 1).

The theorem will be proved in Sect. 4.2. At a high level, the proof is the combination of the following two statements.

- **After the first round, the layout is nice w.h.p.** That is, starting from any pair of inputs, the intermediate layout is "nice" with overwhelmingly high probability. A layout is nice if the number of collisions (i.e., coordinates where the two vectors agree) is relatively small.
- **If the layout is nice before the second round, the output is close to 2-wise independent.** That is, conditioning on the intermediate layout being any nice layout, the pair of outputs will be close to 2-wise independent

Let $I^{(r)}$ denote the layout of $(\mathbf{y}_1^{(r)}, \mathbf{y}_2^{(r)})$ and $(\mathbf{x}_1^{(r+1)}, \mathbf{x}_2^{(r+1)})$. Since the section only discusses the 2-wise setting, the representation of a layout can be simplified. A layout is represented by a subset $I \subseteq [k]$, such that $i \in I$ means the two vectors agree on the i-th position.

As pointed out by the standard differential cryptanalysis, it would be helpful to consider the difference between each pair of vectors

$$\mathbf{x}_{\Delta}^{(r)} := \mathbf{x}_1^{(r)} - \mathbf{x}_2^{(r)}, \qquad \mathbf{y}_{\Delta}^{(r)} := \mathbf{y}_1^{(r)} - \mathbf{y}_2^{(r)}.$$

Note that for each $s \in [k]$

$$\mathbf{y}_{\Delta}^{(r)}[s] = 0 \iff (\mathbf{y}_1^{(r)}[s] = \mathbf{y}_2^{(r)}[s]) \iff s \in I^{(r)}.$$

This suggests that $\mathbf{y}_{\Delta}^{(r)}[s]$ is also "in" $I^{(r)}$. This can be formalized by introducing the following simplified definition for the pairwise setting.

Definition 4. *A (pairwise) layout I is a subset of $[k]$. For any vector $\mathbf{x}_{\Delta}$ and layout I, define*

$$\mathbf{x}_{\Delta} \ \mathsf{SAT}\ I \iff (\forall s \in [k],\ s \in I \implies \mathbf{x}[s] = 0),$$
$$\mathbf{x}_{\Delta} \ \mathsf{in}\ I \iff (\forall s \in [k],\ s \in I \iff \mathbf{x}[s] = 0).$$

And we say I is the layout of $\mathbf{x}_{\Delta}$, denoted by $\mathrm{layout}(\mathbf{x}_{\Delta}) = I$, *if* $\mathbf{x}_{\Delta}$ in I.

Then for any vector difference $\mathbf{x}_{\Delta} = \mathbf{x}_1 - \mathbf{x}_2$, we have

$$\mathbf{x}_{\Delta} \ \mathsf{SAT}\ I \iff (\mathbf{x}_1, \mathbf{x}_2) \ \mathsf{SAT}\ I, \qquad \mathbf{x}_{\Delta} \ \mathsf{in}\ I \iff (\mathbf{x}_1, \mathbf{x}_2) \ \mathsf{in}\ I,$$

and $\mathrm{layout}(\mathbf{x}_{\Delta}) = \mathrm{layout}(\mathbf{x}_1, \mathbf{x}_2)$.

Therefore it suffices to *only* consider the difference vectors, since the whole analysis can ignore the original pair of vectors.

- Permutation step: $\mathbf{y}_{\Delta}^{(r)} = M\mathbf{x}_{\Delta}^{(r)}$.
- Substitution step: $\mathbf{x}_{\Delta}^{(r+1)}$ is a random tuple whose layout is the same as $\mathbf{y}_{\Delta}^{(r)}$.
- Output: The pair of output vectors is ε-close to 2-wise independent if and only if $I^{(2)} = \mathrm{layout}(\mathbf{y}_{\Delta}^{(2)})$ is ε-close to 2-wise independent (Corollary 1).

4.1 The Layout Transition Probability

This section computes the transition probability from layout $I^{(r)}$ to $I^{(r+1)}$. Their dependency can be captured by the following Bayesian network.

$$I^{(r-1)} \longrightarrow \mathbf{x}_{\Delta}^{(r)} \xrightarrow{\ \mathsf{M}\ } \mathbf{y}_{\Delta}^{(r)} \longrightarrow I^{(r)}$$

Let trans-prob(I, J) denote the probability $I^{(r)} = J$ conditioning on $I^{(r-1)} = I$. Formally, trans-prob(I, J) is the probability layout$(M\mathbf{x}) = J$ when the (difference) vector $\mathbf{x}$ is sampled uniformly from layout I. By definition

$$\text{trans-prob}(I, J) = \Pr_{\mathbf{x}\ \mathsf{in}\ I}\Big[M\mathbf{x}\ \mathsf{in}\ J\Big] = \frac{\#\big\{\mathbf{x} : \mathbf{x}\ \mathsf{in}\ I \text{ and } M\mathbf{x}\ \mathsf{in}\ J\big\}}{\#\big\{\mathbf{x} : \mathbf{x}\ \mathsf{in}\ I\big\}}.$$

To simplify this expression, we introduce some new notations.

For the denominator, we define free$(I) = k - |I|$, which stands for the number of "free" coordinates. Then $\#\big\{\mathbf{x} : \mathbf{x}\ \mathsf{in}\ I\big\} = (2^b - 1)^{\text{free}(I)}$.

Denote the numerator by trans-count(I, J). Define indicator function $\mathbb{1}_M$ as

$$\mathbb{1}_M(\mathbf{x}, \mathbf{y}) := \begin{cases} 1 & \text{if } M\mathbf{x} = \mathbf{y}, \\ 0 & \text{otherwise.} \end{cases}$$

Then the numerator can be written as

$$\text{trans-count}(I, J) = \#\big\{\mathbf{x} : \mathbf{x}\ \mathsf{in}\ I \text{ and } M\mathbf{x}\ \mathsf{in}\ J\big\} = \sum_{\mathbf{x}\ \mathsf{in} I} \sum_{\mathbf{y}\ \mathsf{in} J} \mathbb{1}_M(\mathbf{x}, \mathbf{y}).$$

The core idea is to also consider another sum operator $\sum_{\mathbf{x}\ \mathsf{SAT} I}$. For any function f, we have

$$\sum_{\mathbf{x}\ \mathsf{SAT} I} f(\mathbf{x}) = \sum_{I' \supseteq I} \sum_{\mathbf{x}\ \mathsf{in} I'} f(\mathbf{x}).$$

Then by the inclusion-exclusion principle,

$$\sum_{\mathbf{x}\ \mathsf{in} I} f(\mathbf{x}) = \sum_{I' \supseteq I} (-1)^{|I' \setminus I|} \sum_{\mathbf{x}\ \mathsf{SAT} I'} f(\mathbf{x}).$$

Consider the following sum

$$\sum_{\mathbf{x}\ \mathsf{SAT} I} \sum_{\mathbf{y}\ \mathsf{SAT} J} \mathbb{1}_M(\mathbf{x}, \mathbf{y}) = \#\big\{\mathbf{x} : \mathbf{x}\ \mathsf{SAT}\ I \text{ and } M\mathbf{x}\ \mathsf{SAT}\ J\big\} \tag{4}$$

that looks similar to trans-count(I, J). The only difference is whether to enumerate vectors *in* I, J or *satisfying* I, J. The value of (4) is easier to compute. It is the number of solutions of a linear system, which must be a power of $|\mathbb{F}| = 2^b$. In particular, if the matrix M has the maximal branch number, we have

$$\sum_{\mathbf{x}\ \mathsf{SAT} I} \sum_{\mathbf{y}\ \mathsf{SAT} J} \mathbb{1}_M(\mathbf{x}, \mathbf{y}) = \begin{cases} (2^b)^{\text{free}(I) + \text{free}(J) - k} & \text{if } \text{free}(I) + \text{free}(J) \geq k, \\ 1 & \text{if } \text{free}(I) + \text{free}(J) < k. \end{cases} \tag{5}$$

Then by the inclusion-exclusion principle,

$$\begin{aligned} \text{trans-count}(I, J) &= \sum_{\mathbf{x}\ \mathsf{in} I} \sum_{\mathbf{y}\ \mathsf{in} J} \mathbb{1}_M(\mathbf{x}, \mathbf{y}) \\ &= \sum_{I' \supseteq I} \sum_{J' \supseteq J} (-1)^{|I' \setminus I| + |J' \setminus J|} \sum_{\mathbf{x}\ \mathsf{SAT} I'} \sum_{\mathbf{y}\ \mathsf{SAT} J'} \mathbb{1}_M(\mathbf{x}, \mathbf{y}) \\ &= \sum_{I' \supseteq I} \sum_{J' \supseteq J} (-1)^{|I' \setminus I| + |J' \setminus J|} \big(2^b\big)^{\max(\text{free}(I') + \text{free}(J') - k, 0)}. \end{aligned} \tag{6}$$

Now we are ready to present our results about the layout transition probability. They are essentially polishing (6).

Lemma 2. *If M has the maximal branch number, the layout transition probability* $\text{trans-prob}(I, J) := \Pr_{\mathbf{x}\,\mathsf{in}\,I}\left[M\mathbf{x}\ \mathsf{in}\ J\right]$ *is bounded by*

$$\left|\text{trans-prob}(I,J) - \frac{(2^b-1)^{\text{free}(J)}}{(2^b)^k}\right| \le \frac{2^{\text{free}(I)+\text{free}(J)}}{(2^b-1)^{\text{free}(I)}}.$$

Proof. Consider function $u(\mathbf{x},\mathbf{y}) = \frac{1}{(2^b)^k}$. If we view $u(\mathbf{x},\mathbf{y})$ as the conditional probability of $\mathbf{y}$ given $\mathbf{x}$, then it captures the process that $\mathbf{y}$ is sampled uniformly at random and is independent of $\mathbf{x}$. Notice that

$$\sum_{\mathbf{x}\,\mathsf{SAT}\,I}\sum_{\mathbf{y}\,\mathsf{SAT}\,J} u(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{x}\,\mathsf{SAT}\,I}\sum_{\mathbf{y}\,\mathsf{SAT}\,J}\frac{1}{(2^b)^k} = (2^b)^{\text{free}(I)+\text{free}(J)-k}$$

is very similar to (5). The difference is no more than 1 for any I, J. Therefore, in some sense, u is a very good approximation of $\mathbb{1}_M$. With this intuition in mind, we expect

$$\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\mathbb{1}_M(\mathbf{x},\mathbf{y}) \quad - \quad \sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J} u(\mathbf{x},\mathbf{y}) \tag{7}$$

to be very small. The difference between them is bounded by

$$\begin{aligned}
&\left|\text{trans-count}(I,J) - \sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k}\right| = \left|\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\left(\mathbb{1}_M(\mathbf{x},\mathbf{y}) - \frac{1}{(2^b)^k}\right)\right| \\
&= \left|\sum_{I'\supseteq I}\sum_{J'\supseteq J}(-1)^{|I'\setminus I|+|J'\setminus J|}\sum_{\mathbf{x}\,\mathsf{SAT}\,I'}\sum_{\mathbf{y}\,\mathsf{SAT}\,J'}\left(\mathbb{1}_M(\mathbf{x},\mathbf{y}) - \frac{1}{(2^b)^k}\right)\right| \\
&\le \sum_{I'\supseteq I}\sum_{J'\supseteq J} 1 = 2^{\text{free}(I)+\text{free}(J)}.
\end{aligned}$$

So we can approximate the transition probability by

$$\text{trans-prob}(I,J) = \frac{\sum_{\mathbf{x}\,\mathsf{in}\,I}\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k} + \text{term (7)}}{\sum_{\mathbf{x}\,\mathsf{in}\,I} 1} = \underbrace{\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k}}_{\text{approximation}} + \underbrace{\frac{\text{term (7)}}{\sum_{\mathbf{x}\,\mathsf{in}\,I} 1}}_{\text{error}}.$$

The approximation term is particularly nice, as it can be interpreted as the probability that a random vector lies in layout J. It equals to

$$\sum_{\mathbf{y}\,\mathsf{in}\,J}\frac{1}{(2^b)^k} = \Pr_{\mathbf{y}\in\mathbb{F}^k}\left[\mathbf{y}\ \mathsf{in}\ J\right] = \frac{(2^b-1)^{\text{free}(J)}}{(2^b)^k}.$$

The absolute value of the error term is at most $2^{\text{free}(I)+\text{free}(J)}/(2^b-1)^{\text{free}(I)}$. □

Lemma 3. *Let $M : \mathbb{F}^k \to \mathbb{F}^k$ be a matrix with maximal branch number. For any layout I. Let J denote the layout of I after one round of SPN. That is,* $\text{trans-prob}(I, J)$ *is the probability mass function of J. Then J is ε-close to 2-wise independent, where $\varepsilon \le 3^k/2(2^{b-1})^{\text{free}(I)}$.*

4.2 The Niceness of a Layout

Implied by Lemma 3, if the starting input difference is in a layout I with large $\text{free}(I)$, then after one round it will be very close to 2-wise independent. However, consider the extreme case when $\text{free}(I) = 1$, that is, the input difference $\mathbf{x}_\Delta$ is zero on all but one coordinate. Then after one round of SPN with maximal branch number mixing, the difference must be non-zero on every coordinate, which is about $(k/2^b)$ away from 2-wise independence.

So for proving 2-wise independent, a layout I with larger $\text{free}(I)$ is "easier" to analyze. We formalize this by defining the *niceness* of a layout. We say a layout I is α-nice if $|I| = k - \text{free}(I) \le \alpha k$.

To prove Theorem 1, we show that after one round, the layout is likely to be nice, then after one more round, it will be close to 2-wise independent.

Lemma 4. *Assume the mixing function has maximal branch number. For any 2-wise layout I, let J be sampled according to* $\text{trans-prob}(I, J)$. *Then for any $\alpha \in [0, 1]$,*

$$\Pr[J \text{ is } \alpha\text{-nice}] \ge 1 - \frac{e \cdot 2^k}{(2^b - 1)^{\alpha k}}.$$

Proof. The proof starts with an upper bound on the transition probability $\text{trans-prob}(I, J)$ that does not depend on I.

$$\begin{aligned}\text{trans-prob}(I, J) &= \frac{\sum_{\mathbf{x}\,\mathsf{in}\,I} \sum_{\mathbf{y}\,\mathsf{in}\,J} \mathbb{1}_M(\mathbf{x}, \mathbf{y})}{(2^b - 1)^{\text{free}(I)}} \\ &\le \frac{\sum_{\mathbf{x}\,\mathsf{SAT}\,I} \sum_{\mathbf{y}\,\mathsf{SAT}\,J} \mathbb{1}_M(\mathbf{x}, \mathbf{y})}{(2^b - 1)^{\text{free}(I)}} = \frac{(2^b)^{\max(\text{free}(I)+\text{free}(J)-k,0)}}{(2^b - 1)^{\text{free}(I)}}.\end{aligned}$$

Focus on the case that $\text{free}(I)+\text{free}(J) > k$, since otherwise $\text{trans-prob}(I, J) = 0$.

$$\begin{aligned}\text{trans-prob}(I, J) &\le \frac{(2^b)^{\text{free}(I)+\text{free}(J)-k}}{(2^b - 1)^{\text{free}(I)}} \\ &\le \left(\frac{2^b}{2^b - 1}\right)^k \cdot \frac{1}{(2^b - 1)^{k-\text{free}(J)}} \le \frac{e}{(2^b - 1)^{k-\text{free}(J)}}.\end{aligned}$$

The last inequality holds because the mixing function has maximal branch number inherently implies $k \le 2^b$.

We finish the proof by applying the union bound over all layouts J that are not α-nice. The number of not-α-nice layouts is no more than 2^k. □

Proof (Theorem 1). Let $I^{(0)}, I^{(1)}, I^{(2)}$ denote the layout of the inputs, the layout of the middle vectors, the layout of the outputs respectively.

By Lemma 4,

$$\Pr\left[I^{(1)} \text{ is } \alpha\text{-nice}\right] \ge 1 - \frac{e \cdot 2^k}{(2^b - 1)^{\alpha k}}.$$

Conditioning on $I^{(1)}$ being an α-nice layout, $I^{(2)}$ is $(3^k/2(2^{b-1})^{(1-\alpha)k})$-close to 2-wise independent, as shown by Lemma 3. Adding up all the errors, $I^{(2)}$ is ε-close to 2-wise independent, where

$$\varepsilon \leq \frac{3^k}{2 \cdot (2^{b-1})^{(1-\alpha)k}} + \frac{e \cdot 2^k}{(2^b - 1)^{\alpha k}}.$$

Set $\alpha = 1/2$ to minimize the statistical distance bound.

□

5 The General Case of t-Wise Independence

In this section, we generalize our analysis of 2-wise independence in Sect. 4 to the t-wise setting. The high-level framework is mostly the same:

- Introducing the proper notion of *nice* layouts.
- Starting from any t distinct inputs $(\mathbf{x}_1^{(\mathsf{in})}, \ldots, \mathbf{x}_t^{(\mathsf{in})})$, after one round (or a few rounds), the tuple will fall into some nice layout with high probability.
- Core lemma: For any nice layout I, if t inputs $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$ are uniformly sampled from layout I, then after the linear mixing, the layout of $(\mathbf{y}_1, \ldots, \mathbf{y}_1) := (M\mathbf{x}_1, \ldots, M\mathbf{x}_t)$ is close to t-wise independent.

We define *nice* layouts as follows: For any t-wise layout $I = \{I_{i,j}\}_{1 \leq i < j \leq t}$, we say I is α-*nice* if and only if for all $1 < j \leq t$,

$$\left| \bigcup_{i<j} I_{i,j} \right| < \alpha k.$$

Here $\alpha \in [0, 1]$ is a parameter quantifying the niceness of the layout. An equivalent definition is as follows: For any t-tuple $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$, say $\mathbf{x}_j$ collides with $\mathbf{x}_{1:j-1} = (\mathbf{x}_1, \ldots, \mathbf{x}_{j-1})$ on coordinate s if and only if there exists $i < j$ such that $\mathbf{x}_i[s] = \mathbf{x}_j[s]$. Then $\mathbf{x}_{1:t}$ is in an α-nice layout if and only if for every $1 < j \leq t$, $\mathbf{x}_j$ collides with $\mathbf{x}_{1:j-1}$ on at most αk coordinates.

If a t-tuple is sampled from a nice layout, it will be close to t-wise independent after one more round, as shown by our core lemma (Lemma 5). At a high level, the proof inductively uses the technique of its pairwise analog in Sect. 4.

Thanks to this core lemma, in order to show a r-round SPN* is close to t-wise independent, it suffices to show that after the first $r - 1$ rounds, the tuple falls into some nice layout with high probability. We present three different results of this flavor. They differ in the following three criteria

- How large t can be (the core lemma supports t up to $2^{0.499b}$);
- How small the statistical error is (we are aiming for $2^{-\Theta(bk)}$ error); and
- How many rounds are required (ideally 2 rounds).

Each of our results optimizes two of the criteria, and compromises on the third criterion. Section 5.2 can only handle small t. Section 5.3 supports t up to $2^{0.499b}$ but the statistical error is slightly larger. Section 5.4 supports large t and keeps the statistical error $2^{-\Theta(bk)}$, but it requires $O(\log t)$ rounds.

5.1 Core Lemma and Conditional Transition Probability

Lemma 5. *For $\alpha \in [0,1]$ and any α-nice t-wise layout I, if tuple $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$ is sampled uniformly from layout I and let $(\mathbf{y}_1, \ldots, \mathbf{y}_t) = (M\mathbf{x}_1, \ldots, M\mathbf{x}_t)$, then the layout of $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$ is ε-close to t-wise independence with replacement, where*

$$\varepsilon \le t \cdot \left(\frac{2t}{2^b}\right)^{(1-\alpha)k} (2t)^k = t \cdot \left(\frac{(2t)^{2-\alpha}}{(2^b)^{1-\alpha}}\right)^k$$

and we assume the mixing function M has maximal branch number.

This section proves Lemma 5, which is the core of our analysis. The lemma says, if the tuple is in a nice layout at the beginning of a round (must be uniform within this layout due to the S-boxes), then the tuple will become very close to t-wise independent after this round.

The lemma is proved by induction. Assume the lemma holds for smaller t. Say $I = \{I_{a,b}\}_{1 \le a < b \le t}$ is a nice layout, $\mathbf{x}_{1:t}$ is sampled uniformly from layout I and $\mathbf{y}_{1:t} = M\mathbf{x}_{1:t}$, as in the lemma statement. By the definition of niceness, $\mathbf{x}_{1:t-1}$ is sampled uniformly from a nice $(t-1)$-wise layout $I' = \{I_{a,b}\}_{1 \le a < b \le t-1}$. By the induction hypothesis, layout$(\mathbf{y}_{1:t-1})$ is close to $(t-1)$-wise independent. To complete the induction, we need to show that the "conditional layout" of $\mathbf{y}_t$ is close to uniform. First, we need to formalize "conditional layout".

We want to analyze the distribution of $(\mathbf{x}_t, \mathbf{y}_t)$ conditioning on the value of $\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}$. Let's start with a simpler question: What is the conditional distribution of $\mathbf{x}_t$? Since the tuple is sampled from layout I, any constraint in I saying $\mathbf{x}_a[i] = \mathbf{x}_t[i]$ (i.e., if $i \in I_{a,t}$) affects the conditional distribution of $\mathbf{x}_t$. In more detail, the constraints on $\mathbf{x}_t$ can be formalized as[2]

$$I_{\mathrm{c}}(i) = \begin{cases} \mathbf{x}_a[i] & \text{if } i \in I_{a,t} \text{ for some } a < t, \\ \perp & \text{otherwise.} \end{cases} \tag{8}$$

For each $i \in [k]$, if $I_{\mathrm{c}}(i) \neq \perp$ then $\mathbf{x}_t[i]$ must equal to $I_{\mathrm{c}}(i)$, otherwise $\mathbf{x}_t[i]$ is uniform in $\mathbb{F} \setminus \{\mathbf{x}_1[i], \ldots, \mathbf{x}_{t-1}[i]\}$.

Inspired by the above discussion, we formally define *conditional layouts*. When conditioning on $\mathbf{x}_{1:t-1}$ and $\mathbf{y}_{1:t-1} = M\mathbf{x}_{1:t-1}$. For any $i, j \in [k]$, define

$$S_i = \{\mathbf{x}_a[i] \mid a < t\}, \qquad T_j = \{\mathbf{y}_a[j] \mid a < t\}.$$

A *conditional layout* for $\mathbf{x}_t$ is specified by a function $I_{\mathrm{c}} : [k] \to \mathbb{F} \cup \{\perp\}$ such that $I_{\mathrm{c}}(i) \in S_i \cup \{\perp\}$ for every $i \in [k]$. Define $\mathbf{x}_t$ is in I_{c} (denoted by $\mathbf{x}_t$ in I_{c}) and $\mathbf{x}_t$ satisfies I_{c} (denoted by $\mathbf{x}_t$ SAT I_{c}) as

$$\mathbf{x}_t \ \mathsf{in}\ I_{\mathrm{c}} \iff \forall i \in [k], \begin{pmatrix} I_{\mathrm{c}}(i) \neq \perp \implies \mathbf{x}_t[i] = I_{\mathrm{c}}(i), \\ I_{\mathrm{c}}(i) = \perp \implies \mathbf{x}_t[i] \notin S_i \end{pmatrix},$$

$$\mathbf{x}_t \ \mathsf{SAT}\ I_{\mathrm{c}} \iff \forall i \in [k], \Big(I_{\mathrm{c}}(i) \neq \perp \implies \mathbf{x}_t[i] = I_{\mathrm{c}}(i)\Big).$$

[2] Even if there exists distinct a, a' such that $i \in I_{a,t} \cap I_{a',t}$, I_{c} is still well-defined. Because in such case, we must have $i \in I_{a,a'}$ (otherwise I is not a valid layout), then $\mathbf{x}_a[i] = \mathbf{x}_{a'}[i]$.

We say I_c is the layout of $\mathbf{x}_t$, denoted by $\text{layout}_c(\mathbf{x}^{(t)}) = I_c$, if $\mathbf{x}_t \in I_c$. Define

$$\text{free}(I_c) := |I_c^{-1}(\perp)| = \#\{i \in [k] \text{ s.t. } I_c(i) = \perp\}$$

as the number of coordinates that I_c outputs $\perp$. Note that, if I_c is derived from an α-nice layout I as in (8), then

$$\text{free}(I_c) = k - \left|\bigcup_{a<t} I_{a,t}\right| \geq (1-\alpha)k.$$

Define I'_c is stricter or equal to I_c, denoted by $I'_c \supseteq I_c$, as

$$I'_c \supseteq I_c \iff \forall i \in [k], \Big(I_c(i) \neq \perp \implies I'_c(i) = I_c(i)\Big).$$

Symmetrically, a conditional layout for $\mathbf{y}_t$ is specified by a function $J_c : [k] \to \mathbb{F} \cup \{\perp\}$ such that $J_c(j) \in T_j \cup \{\perp\}$ for every $j \in [k]$. We adopt the same notations and terminology from the conditional layout of $\mathbf{x}_t$.

Let $\mathbf{y}^*$ be sampled uniformly at random from $\mathbb{F}^k$. Then

$$\Pr\Big[\text{layout}_c(\mathbf{y}^*) = J_c\Big] = \sum_{\mathbf{y} \text{ in} J_c} \frac{1}{2^{bk}} = \frac{\prod_{j \in [k] \text{ s.t. } J_c(j)=\perp} (2^b - |T_j|)}{2^{bk}}. \tag{9}$$

We hope $\text{layout}_c(\mathbf{y}_t)$ is close to $\text{layout}_c(\mathbf{y}^*)$ by distribution. So we analyze the transition probability from I_c to J_c. That is, if $\mathbf{x}$ is sampled from layout I_c, what is the distribution of the layout of $\mathbf{y} = M\mathbf{x}$. We found that, if $\text{free}(I_c)$ is large enough, the layout of $\mathbf{y}$ is close to the layout of random $\mathbf{y}^*$ by distribution.

Lemma 6. *Assume the linear mixing M has maximal branch number. Conditioning on any sets $S_1, \ldots, S_k, T_1, \ldots, T_k$, each of size at most $t-1$. For any conditional layout I_c, if $\mathbf{x}$ is sampled uniformly at random from layout I_c and let $\mathbf{y} := M\mathbf{x}$, then the statistical distance between* $\text{layout}_c(\mathbf{y})$ *and the conditional layout of a random vector is no greater than*

$$\Big(\frac{2t-1}{2^b}\Big)^{\text{free}(I_c)} (2t-1)^k.$$

We start by bounding the transition probability. For any conditional layouts I_c, J_c, the transition probability from I_c to J_c, denoted by $\text{trans-prob}(I_c, J_c)$, is the probability $M\mathbf{x}$ in J_c when $\mathbf{x}$ is sampled from layout I_c. By definition,

$$\text{trans-prob}(I_c, J_c) = \frac{\text{trans-count}(I_c, J_c)}{\text{size of layout } I_c} = \frac{\sum_{\mathbf{x} \text{ in} I_c} \sum_{\mathbf{y} \text{ in} J_c} \mathbb{1}_M(\mathbf{x}, \mathbf{y})}{\sum_{\mathbf{x} \text{ in} I_c} 1} \tag{10}$$

where $\mathbb{1}_M$ is defined as

$$\mathbb{1}_M(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } M\mathbf{x} = \mathbf{y}, \\ 0 & \text{otherwise.} \end{cases}$$

We show that if $\text{free}(I_c)$ is sufficiently large, then the transition probability $\text{trans-prob}(I_c, J_c)$ is close to the probability that random $\mathbf{y}^*$ lies in layout J_c.

Lemma 7. *Assume the linear mixing has maximal branch number. Conditioning on any sets $S_1, \ldots, S_k, T_1, \ldots, T_k$, each of size at most $t-1$. For any (conditional) layouts I_c, J_c, the transition probability from I_c to J_c is bounded by*

$$\left|\text{trans-prob}(I_c, J_c) - \sum_{\mathbf{y}\ \mathsf{in} J_c} \frac{1}{2^{bk}}\right| \le \left(\frac{2t-1}{2^b}\right)^{\text{free}(I_c)} t^{\text{free}(J_c)}.$$

Proof. In the definition of transition probability (Eq. (10)), the sum is over $\mathbf{x}$ in I_c, which is hard to analyze. But we know how $\sum_{\mathbf{x}\ \mathsf{SAT} I_c}$ and $\sum_{\mathbf{x}\ \mathsf{in} I_c}$ are closely connected. On the easy direction, we have

$$\sum_{\mathbf{x}\ \mathsf{SAT} I_c} \equiv \sum_{I'_c \supseteq I_c} \sum_{\mathbf{x}\ \mathsf{in} I'_c}, \qquad \sum_{\mathbf{y}\ \mathsf{SAT} J_c} \equiv \sum_{J'_c \supseteq J_c} \sum_{\mathbf{y}\ \mathsf{in} J'_c}.$$

Then by the inclusion-exclusion principle

$$\sum_{\mathbf{x}\ \mathsf{in} I_c} \equiv \sum_{I'_c \supseteq I_c} (-1)^{\Delta(I'_c, I_c)} \sum_{\mathbf{x}\ \mathsf{SAT} I'_c}, \qquad \sum_{\mathbf{y}\ \mathsf{in} J_c} \equiv \sum_{J'_c \supseteq J_c} (-1)^{\Delta(J'_c, J_c)} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c},$$

where Δ denotes the Hamming distance. Since $I'_c \supseteq I_c$, the Hamming distance can also be written as $\Delta(I'_c, I_c) = \text{free}(I_c) - \text{free}(I'_c)$.

We can apply the inclusion-exclusion principle to the numerator of (10),

$$\begin{aligned}\text{trans-count}(I_c, J_c) &= \sum_{\mathbf{x}\ \mathsf{in} I_c} \sum_{\mathbf{y}\ \mathsf{in} J_c} \mathbb{1}_M(\mathbf{x}, \mathbf{y}) \\ &= \sum_{I'_c \supseteq I_c} \sum_{J'_c \supseteq J_c} (-1)^{\Delta(I'_c, I_c) + \Delta(J'_c, J_c)} \sum_{\mathbf{x}\ \mathsf{SAT} I'_c} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c} \mathbb{1}_M(\mathbf{x}, \mathbf{y}).\end{aligned}$$

As we have observed in previous sections, $\sum_{\mathbf{x}\ \mathsf{SAT} I'_c} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c} \mathbb{1}_M(\mathbf{x}, \mathbf{y})$ is easy to bound. Since the linear mixing has maximal branch number,

$$\sum_{\mathbf{x}\ \mathsf{SAT} I'_c} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c} \mathbb{1}_M(\mathbf{x}, \mathbf{y}) = \begin{cases} (2^b)^{\text{free}(I'_c) + \text{free}(J'_c) - k} & \text{if } \text{free}(I'_c) + \text{free}(J'_c) \ge k \\ 0 \text{ or } 1 & \text{otherwise.} \end{cases}$$

It can be approximated by

$$\sum_{\mathbf{x}\ \mathsf{SAT} I'_c} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c} \frac{1}{2^{bk}} = (2^b)^{\text{free}(I'_c) + \text{free}(J'_c) - k},$$

such that the absolute value of the error is no more than 1 for any I'_c, J'_c.

As $\sum_{\mathbf{x}\ \mathsf{SAT} I'_c} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c} \frac{1}{2^{bk}}$ is a good approximation of $\sum_{\mathbf{x}\ \mathsf{SAT} I'_c} \sum_{\mathbf{y}\ \mathsf{SAT} J'_c} \mathbb{1}_M(\mathbf{x}, \mathbf{y})$ and the inclusion-exclusion principle has small coefficients, $\sum_{\mathbf{x}\ \mathsf{in} I'_c} \sum_{\mathbf{y}\ \mathsf{in} J'_c}$

$\frac{1}{2^{bk}}$ should also be a fairly good approximation of trans-count(I_c, J_c).

$$\begin{aligned}
&\left|\text{trans-count}(I_c, J_c) - \sum_{\mathbf{x}\,\text{in}\,I_c}\sum_{\mathbf{y}\,\text{in}\,J_c}\frac{1}{2^{bk}}\right| = \left|\sum_{\mathbf{x}\,\text{in}\,I_c}\sum_{\mathbf{y}\,\text{in}\,J_c}\left(\mathbb{1}_M(\mathbf{x},\mathbf{y}) - \frac{1}{2^{bk}}\right)\right| \\
&= \left|\sum_{I'_c \supseteq I_c}\sum_{J'_c \supseteq J_c}(-1)^{\Delta(I'_c,I_c)+\Delta(J'_c,J_c)}\sum_{\mathbf{x}\,\text{SAT}\,I'_c}\sum_{\mathbf{y}\,\text{SAT}\,J'_c}\left(\mathbb{1}_M(\mathbf{x},\mathbf{y}) - \frac{1}{2^{bk}}\right)\right| \qquad (11) \\
&\le \sum_{I'_c \supseteq I_c}\sum_{J'_c \supseteq J_c} 1 \le t^{\text{free}(I_c)+\text{free}(J_c)}.
\end{aligned}$$

This can be translated into a bound on the transition probability,

$$\left|\text{trans-prob}(I_c, J_c) - \frac{\sum_{\mathbf{x}\,\text{in}\,I_c}\sum_{\mathbf{y}\,\text{in}\,J_c}\frac{1}{2^{bk}}}{\sum_{\mathbf{x}\,\text{in}\,I_c} 1}\right| \le \frac{t^{\text{free}(I_c)+\text{free}(J_c)}}{\sum_{\mathbf{x}\,\text{in}\,I_c} 1}.$$

In the fraction on the left-hand side, the $\sum_{\mathbf{x}\,\text{in}\,I_c} 1$ in the numerator and in the denominator can cancel out. So

$$\begin{aligned}
\left|\text{trans-prob}(I_c, J_c) - \sum_{\mathbf{y}\,\text{in}\,J_c}\frac{1}{2^{bk}}\right| &\le \frac{t^{\text{free}(I_c)+\text{free}(J_c)}}{\sum_{\mathbf{x}\,\text{in}\,I_c} 1} \\
&\le \frac{t^{\text{free}(I_c)+\text{free}(J_c)}}{(2^b-(t-1))^{\text{free}(I_c)}} \le \left(\frac{2t-1}{2^b}\right)^{\text{free}(I_c)} t^{\text{free}(J_c)}.
\end{aligned}$$

The last inequality assumes $t \le 2^{b-1}$, we can assume this without loss of generality, because the lemma is trivialized otherwise. □

Now we can prove Lemma 6, by adding up the error term over all layouts J_c.

Proof (Lemma 6). The statistical distance between the conditional layout of $\mathbf{y}$ and the conditional layout of a random $\mathbf{y}^* \in \mathbb{F}^k$ is bounded by

$$\sum_{J_c}\left(\frac{2t-1}{2^b}\right)^{\text{free}(I_c)} t^{\text{free}(J_c)} \le \left(\frac{2t-1}{2^b}\right)^{\text{free}(I_c)}(2t-1)^k.$$

The inequality holds because

$$\sum_{J_c} t^{\text{free}(J_c)} = \sum_i \sum_{\substack{J_c \text{ s.t.} \\ \text{free}(J_c)=i}} t^i \le \sum_i \binom{k}{i}(t-1)^{k-i}t^i = (2t-1)^k.$$

□

We are now ready to complete our inductive proof of the core lemma (Lemma 5).

Proof (Lemma 5). Let $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$ be sampled uniformly from an α-nice layout I. We need to show that the layout of $\mathbf{y}_{1:t} := (M\mathbf{x}_1, \ldots, M\mathbf{x}_t)$ is statistically close to the layout of t random vectors.

Consider $\mathbf{x}_{1:t}^{(\mathsf{next})} = (\mathbf{x}_1^{(\mathsf{next})}, \ldots, \mathbf{x}_t^{(\mathsf{next})})$, which is obtained by applying k independent random S-boxes on $\mathbf{y}_{1:t}$. By Corollary 1, it is equivalent to study the statistical distance between $\mathbf{x}_{1:t}^{(\mathsf{next})}$ and t random vectors. Denote this statistical distance by $\varepsilon(t)$. Clearly $\varepsilon(1) = 0$.

For $t > 1$, assume the lemma holds for smaller t. By our definition of niceness, $\mathbf{x}_{1:t-1}$ is sampled from an α-nice layout I'. By the induction hypothesis, $\mathbf{x}_{1:t-1}^{(\mathsf{next})}$ is $\varepsilon(t-1)$-close to uniform by distribution. Implied by Lemma 6, the distribution of $\mathbf{x}_t^{(\mathsf{next})}$ conditioning on the values of $\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}, \mathbf{x}_{1:t-1}^{(\mathsf{next})}$ is very close to uniform. The (conditional) statistical distance is at most $(\frac{2t-1}{2^b})^{\mathrm{free}(I_c)}(2t-1)^k$ where I_c is determined by (8). Since I is α-nice, $\mathrm{free}(I_c) \geq (1-\alpha)k$. Therefore, the statistical distance between $\mathbf{x}_{1:t}^{(\mathsf{next})}$ and t random vectors is bounded by

$$\varepsilon(t) \leq \varepsilon(t-1) + \left(\frac{2t-1}{2^b}\right)^{(1-\alpha)k}(2t-1)^k.$$

By induction on t,

$$\varepsilon(t) \leq \sum_{t'=2}^{t}\left(\frac{2t'-1}{2^b}\right)^{(1-\alpha)k}(2t'-1)^k \leq t \cdot \left(\frac{2t}{2^b}\right)^{(1-\alpha)k}(2t)^k.$$

□

5.2 2-Round SPN* is $2^{-\Theta(bk)}$-Close to $O(1)$-Wise Independence

In this section, we use the core lemma (Lemma 5) to prove that a 2-round SPN* is $2^{-\Theta(bk)}$-close to t-wise independent, for constant t.

Theorem 2. *The 2-round SPN* is ε-close to t-wise independent, where*

$$\varepsilon = \frac{t^2 \cdot 2^{k+1}}{(2^b)^{k/(2t)}} + t \cdot \left(\frac{8 \cdot t^3}{2^b}\right)^{k/2},$$

if the linear mixing has maximal branch number.

When t is a constant, the distance satisfies $\varepsilon = 2^{-\Theta(bk)}$.

The proof follows the high-level framework introduced at the beginning of Sect. 5. Lemma 8 shows that for constant t, the first-round tuple $\mathbf{y}_{1:t}^{(1)}$ will be in an α-nice layout with high probability. Thus, the core lemma (Lemma 5) implies that the layout of the second-round tuple $\mathbf{y}_{1:t}^{(2)}$ is exponentially close to t-wise independent.

Lemma 8. *For any $\alpha \in [0,1]$ and any t-wise layout I, if tuple $\mathbf{x}_{1:t}$ is sampled uniformly from layout I, and let $\mathbf{y}_{1:t} = M\mathbf{x}_{1:t}$, then $J = \mathrm{layout}(\mathbf{y}_{1:t})$ is α-nice with probability*

$$\Pr[J \text{ is } \alpha\text{-nice}] \geq 1 - \frac{2^{k+1} \cdot t^2}{(2^b)^{\alpha k/t}}.$$

Proof. We will upper bound the probability that J is α-nice by requiring that each pair of vectors collide in at most $\alpha k/(t-1)$ coordinates. Then every vector collides with other vectors on at most αk coordinates, which implies that the layout of the tuple is α-nice.

The number of collisions between each pair of vectors can be bounded by Lemma 4, which does not depend on the starting layout. The probability $|J_{i,j}| > \alpha k/t$ is no more than $e \cdot 2^k/(2^b-1)^{\alpha k/t}$.

$$\Pr\Big[J \text{ is not } \alpha\text{-nice}\Big] \le \Pr\left[\bigwedge_{1\le i<j\le t} |J_{i,j}| > \frac{\alpha k}{t}\right] \le \frac{t^2 \cdot 2^{k+1}}{(2^b-1)^{\alpha k/t}}$$

The last inequality is obtained by applying the union bound inequality over all $\binom{t}{2} \le \frac{t^2}{2}$ pairs of vectors. □

We are now ready to present the proof of the main theorem of this section.

Proof (Theorem 2). Lemma 8 shows that

$$\varepsilon_1 := \Pr[J \text{ is not } \alpha\text{-nice}] \le \frac{t^2 \cdot 2^{k+1}}{(2^b-1)^{\alpha k/t}}.$$

Conditioning on J being α-nice, consider the (conditional) distribution of $\mathbf{y}^{(2)}_{1:t}$. The core lemma (Lemma 5) shows that the conditional distribution is ε_2-close to t-wise independent, for

$$\varepsilon_2 \le t \cdot \left(\frac{(2t)^{2-\alpha}}{(2^b)^{1-\alpha}}\right)^k.$$

In conclusion, the output tuple $\mathbf{x}^{(3)}_{1:t}$, alias $\mathbf{y}^{(\mathsf{out})}_{1:t}$, is $(\varepsilon_1+\varepsilon_2)$-close to t-wise independent. If we set $\alpha = \frac{1}{2}$, the statistical distance is bounded by

$$\varepsilon_1 + \varepsilon_2 \le \frac{t^2 \cdot 2^{k+1}}{(2^b)^{k/(2t)}} + t \cdot \left(\frac{8 \cdot t^3}{2^b}\right)^{k/2}.$$

□

5.3 2-Round SPN* is $2^{-\Theta(b)}$-Close to t-Wise independent

This section shows a similar result for larger t. In particular, we prove that 2-round SPN* with a maximal-branch-number mixing is $2^{-\Theta(b)}$-close to t-wise independent, for t almost up to $2^{0.499b}$.

By applying the amplification result of Maurer, Pietrzak, and Renner [39], we can reduce the error to $2^{-\Theta(bk)}$ by having $O(k)$ rounds.

Theorem 3. *For any $\alpha \in (0,1]$, the 2-round SPN* is ε-close to t-wise independent, where*

$$\varepsilon = \frac{t^2}{\alpha \cdot 2^b} + t \cdot \left(\frac{(2t)^{2-\alpha}}{(2^b)^{1-\alpha}}\right)^k,$$

if the mixing function has the maximal branch number.

If $t < 2^{(0.499-1/(4k))b}$, the distance is $\varepsilon = 2^{-\Theta(b)}$ by choosing the optimal α.

Corollary 2. *Assuming $t < 2^{(0.499-1/(4k))b}$, $\Theta(k)$-round SPN* with maximal-branch-number linear mixing is $2^{-\Theta(bk)}$-close to t-wise independent.*

The proof of this theorem is in the full version of the paper.

5.4 ($\log t$)-Rounds SPN* is $2^{-\Theta(bk)}$-Close to t-Wise Independent

In this section, we discuss how to achieve $2^{-\Theta(bk)}$-closeness to t-wise independent, for t up to $2^{0.499b}$, at the cost of a slightly larger number of rounds.

This result is proved by induction. The base case is closeness to 2-wise independent in 2 rounds. Assume that we have already shown ε-closeness to t-wise independent in r rounds. As the inductive step, we will prove the closeness to $(2t-1)$-wise independent in $r+1$ rounds.

As for notations, let $\mathbf{x}_{1:2t-1}^{(\mathsf{in})}$ denote $2t-1$ distinct inputs, let $\mathbf{y}_{1:2t-1}^{(\mathsf{out})}$ denote their corresponding outputs, and let $\mathbf{x}_{1:2t-1}^{(\mathsf{last})}, \mathbf{y}_{1:2t-1}^{(\mathsf{last})}$ denote the intermediate values in the last round (as illustrated in Fig. 3).

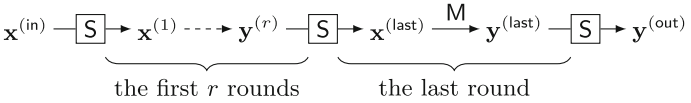


Fig. 3. Illustration of a $(r+1)$-round SPN

Due to the core lemma (Lemma 5), it suffices to show that: With overwhelming probability, $(\mathbf{x}_{1:2t-1}^{(\mathsf{last})})$ lies in a α-nice layout for some $\alpha \in (0,1)$ of our choice.

By the induction hypothesis, we know that the distribution of $\mathbf{x}_{1:t}^{(\mathsf{last})}$ is $\varepsilon(t)$-close to t-wise independent. If they are actually t-wise independent, then the probability $\mathbf{x}_t^{(\mathsf{last})}$ collides with $\mathbf{x}_{1:t-1}^{(\mathsf{last})}$ in more than $\alpha k/2$ coordinates is exponentially small due to Chernoff bound. The same argument also bounds the probability that $\mathbf{x}_t^{(\mathsf{last})}$ collides with $\mathbf{x}_{t+1:2t-1}^{(\mathsf{last})}$ in more than $\alpha k/2$ coordinates. Then the probability $\mathbf{x}_t^{(\mathsf{last})}$ collides with the other $2t-2$ vectors in at most αk coordinates is bounded by the union bound. Due to the symmetry and the union bound, $\mathbf{x}_{1:2t-1}^{(\mathsf{last})}$ is α-nice with good probability. Then we can finish the induction step by Lemma 5.

Such analysis can show $\varepsilon(t)$-closeness to t-wise independent in $O(\log t)$ rounds, where $\varepsilon(t)$ is inductively bounded by

$$\varepsilon(2t-1) \le O(t) \cdot \Big(\varepsilon(t) + \underset{\text{from Chernoff bound}}{\text{a small term}}\Big) + \underset{\text{from Lemma 5}}{\text{another small term.}}$$

The $O(t)$ multiplicative factor before $\varepsilon(t)$ turns out to be problematic. It results in a multiplicative blow-up of order $t^{O(\log t)}$. When $t = 2^{\Theta(b)}$, this blow-up is about $2^{O(b^2)}$, which is unacceptable especially if $b = \Omega(k)$. In the actual proof

of our result (Theorem 4), we conduct a more sophisticated analysis, though the high-level inductive idea is the same.

Theorem 4. *If $k > 4$, r-round SPN* is ε-close to 2^r-wise independent for*

$$\varepsilon = \frac{2^{r+\frac{3}{4}}}{1-2^{-\frac{k}{4}}} \cdot \left(\frac{2^{2r+3}}{2^b}\right)^{k/4} = \frac{t \cdot 2^{\frac{11}{4}}}{1-2^{-\frac{k}{4}}} \cdot \left(\frac{8 \cdot t^2}{2^b}\right)^{k/4}.$$

As usual, let $\mathbf{x}_{1:t}^{(r)}, \mathbf{y}_{1:t}^{(r)}$ denote the intermediate values in the r-th round. We also introduce a new notation $\mathbf{x}_{i,\times 2^\rho}^{(r)}$

$$\mathbf{x}_{i,\times 2^\rho}^{(r)} := \mathbf{x}_{i2^\rho+1:i2^\rho+2^\rho}^{(r)} = (\mathbf{x}_{i2^\rho+1}^{(r)}, \ldots, \mathbf{x}_{i2^\rho+2^\rho}^{(r)})$$

to denote 2^ρ consecutive vectors. Similarly we define $\mathbf{y}_{i,\times 2^\rho}^{(r)}$.

In the ρ-th round, for $0 \le i < j < 2^{r-\rho}$, define $A_{i,j}^{(\rho)}$ as the event that

$$\Big(\underbrace{\mathbf{x}_{i2^{\rho-1}+1}^{(\rho)}, \ldots, \mathbf{x}_{i2^{\rho-1}+2^{\rho-1}}^{(\rho)}}_{\mathbf{x}_{i,\times 2^{\rho-1}}^{(\rho)}}, \underbrace{\mathbf{x}_{j2^{\rho-1}+1}^{(\rho)}, \ldots, \mathbf{x}_{j2^{\rho-1}+2^{\rho-1}}^{(\rho)}}_{\mathbf{x}_{j,\times 2^{\rho-1}}^{(\rho)}}\Big) \tag{12}$$

is in an α_ρ-nice layout. For $0 \le i < j < 2^{r-\rho+1}$, define $B_{i,j}^{(\rho)}$ as the event that

$$\Big(\underbrace{\mathbf{x}_{i2^{\rho-2}+1}^{(\rho)}, \ldots, \mathbf{x}_{i2^{\rho-2}+2^{\rho-2}}^{(\rho)}}_{\mathbf{x}_{i,\times 2^{\rho-2}}^{(\rho)}}, \underbrace{\mathbf{x}_{j2^{\rho-2}+1}^{(\rho)}, \ldots, \mathbf{x}_{j2^{\rho-2}+2^{\rho-2}}^{(\rho)}}_{\mathbf{x}_{j,\times 2^{\rho-2}}^{(\rho)}}\Big) \tag{13}$$

is in a $\frac{1}{3}\alpha_\rho$-nice layout. The value of α_ρ will be fixed later.

The proof of Theorem 4 is inductive. The induction hypothesis is that with overwhelming probability $\bigwedge_{0 \le i < j < 2^{r-\rho}} A_{i,j}^{(\rho)}$ holds. Then by Lemma 5, the joint distribution of $\mathbf{x}_{i,\times 2^{\rho-1}}^{(\rho)}, \mathbf{x}_{j,\times 2^{\rho-1}}^{(\rho)}$ is close to 2^ρ-wise uniform, for each $0 \le i < j < 2^{r-\rho}$. Then by the following Lemma 9, they are very likely to be $\frac{1}{3}\alpha_{\rho+1}$-nice, that is, $B_{i,j}^{(\rho+1)}$ is likely to hold. To complete the induction step, we bridge the remaining gap by proving the following statement for $\rho > 2$,

$$\bigwedge_{0 \le i < j < 2^{r-\rho+1}} B_{i,j}^{(\rho)} \implies \bigwedge_{0 \le i < j < 2^{r-\rho}} A_{i,j}^{(\rho)}. \tag{14}$$

Lemma 9. *Assume $\mathbf{x}_{1:t}$ are uniformly sampled from $(\mathbb{F}^k)^t$, for any $\alpha > \frac{t-1}{2^b}$,*

$$\Pr\big[\text{layout}(\mathbf{x}_1, \ldots, \mathbf{x}_t) \text{ is } \alpha\text{-nice}\big] \ge 1 - \frac{t \cdot 2^k}{1+\alpha k} \cdot \left(\frac{t}{2^b}\right)^{\alpha k}.$$

The proofs of statement (14) and of Lemma 9 are deferred to the full version of the paper.

Now we are nearly ready to prove Theorem 4. We introduce a few additional notations. For $\rho \geq 1$, define

$$A_\rho := \bigwedge_{0 \leq i < j < 2^{r-\rho}} A_{i,j}^{(\rho)}, \qquad \delta_\rho := 1 - \Pr[A_\rho].$$

Define $\varepsilon_{\rho,i,j}$ as the statistical distance between the uniform distribution and the distribution of $\mathbf{x}_{i,\times 2^{\rho-1}}^{(\rho+1)}, \mathbf{x}_{j,\times 2^{\rho-1}}^{(\rho+1)}$ (the vectors in the definition of $B_{i,j}^{(\rho+1)}$) *conditioning on event* A_ρ. Lemma 5 shows that

$$\varepsilon_{\rho,i,j} \leq 2^\rho \cdot \left(\frac{2^{\rho+1}}{2^b}\right)^{(1-\alpha_\rho)k} (2^{\rho+1})^k.$$

for all $\rho \geq 2$. Define $\varepsilon_\rho = \sum_{0 \leq i < j < 2^{r-\rho}} \varepsilon_{\rho,i,j}$.

Note that $\varepsilon_r = \varepsilon_{r,1,2}$ is the statistical distance between the 2^r output vectors and uniform, conditioning on A_r. So r-round SPN* is $(\delta_r + \varepsilon_r)$-close to 2^r-wise independent.

Proof (Theorem 4). For each $2 < \rho \leq r$, conditional on $A_{\rho-1}$, the (conditional) distribution of $\mathbf{x}_{i,\times 2^{\rho-2}}^{(\rho)}, \mathbf{x}_{j,\times 2^{\rho-2}}^{(\rho)}$ is $\varepsilon_{\rho-1,i,j}$-close to uniform. Then by Lemma 9

$$\Pr\left[\neg B_{i,j}^{(\rho)} \;\middle|\; A_{\rho-1}\right] \leq \varepsilon_{\rho-1,i,j} + 2^{\rho-1} \cdot 2^k \cdot \left(\frac{2^{\rho-1}}{2^b}\right)^{\frac{1}{3}\alpha_\rho k}.$$

By the union bound,

$$\Pr\left[\neg \bigwedge_{0 \leq i < j < 2^{r-\rho+1}} B_{i,j}^{(\rho)} \;\middle|\; A_{\rho-1}\right] \leq \varepsilon_{\rho-1} + \frac{(2^{r-\rho+1})^2}{2} \cdot 2^{\rho-1} \cdot 2^k \cdot \left(\frac{2^{\rho-1}}{2^b}\right)^{\frac{1}{3}\alpha_\rho k}.$$

By (14), the left-hand side is lower bounded by $\Pr\left[\neg A_\rho \;\middle|\; A_{\rho-1}\right]$. And we know

$$\Pr\left[\neg A_\rho \;\middle|\; A_{\rho-1}\right] \geq \Pr\left[\neg A_\rho \wedge A_{\rho-1}\right] \geq \delta_\rho - \delta_{\rho-1}.$$

So

$$\delta_\rho \leq \delta_{\rho-1} + \varepsilon_{\rho-1} + \frac{(2^{r-\rho+1})^2}{2} 2^{\rho-1} \cdot 2^k \cdot \left(\frac{2^{\rho-1}}{2^b}\right)^{\frac{1}{3}\alpha_\rho k}.$$

For the base case $\rho = 2$, Lemma 4 directly bounds the probability of $B_{i,j}^{(2)}$ by $\frac{e \cdot 2^k}{(2^b-1)^{\frac{1}{3}\alpha_2 k}}$. Then by the union bound

$$\delta_2 \leq \Pr\left[\neg \bigwedge_{i,j} B_{i,j}^{(2)}\right] \leq \frac{(2^r)^2}{2} \frac{e \cdot 2^k}{(2^b-1)^{\frac{1}{3}\alpha_2 k}}.$$

As the final goal is to bound $\delta_r + \varepsilon_r$, we are interested in how $\delta_\rho + \varepsilon_\rho$ depends on $\delta_{\rho-1} + \varepsilon_{\rho-1}$,

$$\begin{aligned}
&(\delta_\rho + \varepsilon_\rho) - (\delta_{\rho-1} + \varepsilon_{\rho-1}) \\
&\leq \frac{(2^{r-\rho})^2}{2} 2^\rho \cdot \left(\frac{2^{\rho+1}}{2^b}\right)^{(1-\alpha_\rho)k} (2^{\rho+1})^k + \frac{(2^{r-\rho+1})^2}{2} 2^{\rho-1} \cdot 2^k \cdot \left(\frac{2^{\rho-1}}{2^b}\right)^{\frac{1}{3}\alpha_\rho k} \\
&= 2^{2r-\rho-1} \left(\left(\frac{2^{\rho+1}}{2^b}\right)^{(1-\alpha_\rho)k} (2^{\rho+1})^k + 2^{k+1} \cdot \left(\frac{2^{\rho-1}}{2^b}\right)^{\frac{1}{3}\alpha_\rho k} \right). \qquad (15)
\end{aligned}$$

Table 2. Statistical (TV) distance from pairwise independence of the r-round AES* given two inputs that differ in exactly one coordinate. This corresponds to starting from a layout I with Hamming weight 1, e.g. $I = \{1, \ldots, k-1\}$.

Number of rounds r	$\log_2$(TV distance from 2-wise ind.)
3	−23.4275
4	−48.9916
5	−117.1745
6	−126.3073
7	−141.2575

The value of α_ρ should be chosen so that (15) is minimized. Note that

$$\text{Right-hand side of (15)} \approx \left(\frac{2^\rho}{2^b}\right)^{-\alpha_\rho k}\left(\frac{2^{2\rho}}{2^b}\right)^k + \left(\frac{2^\rho}{2^b}\right)^{\frac{1}{3}\alpha_\rho k}$$

so (15) is minimized when $\alpha \approx \frac{3}{4}\frac{b-2\rho}{b-\rho}$, and the minimum value is about $\left(\frac{2^{2\rho}}{2^b}\right)^{k/4}$. If we tune the value of α_ρ, we get

$$(\delta_\rho + \varepsilon_\rho) - (\delta_{\rho-1} + \varepsilon_{\rho-1}) \le 2^{2r-\rho} \cdot 2^{\frac{1}{2}k\rho - \frac{1}{4}kb + \frac{3}{4}k + \frac{3}{4}} = 2^{2r-\rho+\frac{3}{4}} \cdot \left(\frac{2^{2\rho+3}}{2^b}\right)^{k/4}.$$

We defer the analysis of the base case to the full version.

$$\delta_r + \varepsilon_r \le \delta_2 + \varepsilon_2 + \sum_{\rho=3}^{r} 2^{2r-\rho+\frac{3}{4}} \cdot \left(\frac{2^{2\rho+3}}{2^b}\right)^{k/4} \le \frac{2^{r+\frac{3}{4}}}{1-2^{-\frac{k}{4}}} \cdot \left(\frac{2^{2r+3}}{2^b}\right)^{k/4}.$$

□

6 Pairwise Independence of AES* and Censored AES

In this section, we obtain concrete bounds on the pairwise independence of (1) an SPN cipher with random, independent S-boxes and the *actual* AES mixing (we refer to this as AES*) as well as (2) a "censored" version of the actual AES block cipher (with the *actual* AES S-box, but some mixing layers removed). We will use partially computational methods for our theorems. The source code for our computations is available at https://github.com/AnPelec/t-wise-ind-SPN.

6.1 Pairwise Independence of AES*

We can represent the evaluation of AES* as a Markov chain over $2^{16}-1$ layouts. Our goal is to describe this random walk exactly, and then use numerical calculations to infer an upper bound on the statistical distance of an output pair after a certain number of rounds. To compute the transition probabilities, we start with an exact version of Lemma 2. A similar lemma was already proved in [3], by relating the number of transitions to the number of codewords of specific weight in an MDS code.

Lemma 10. *If M has the maximal branch number, the layout transition probability* $\text{trans-prob}(I, J) := \Pr_{\mathbf{x} \text{ in } I}\left[M\mathbf{x} \text{ in } J\right]$ *equals*

$$\text{trans-prob}(I, J) = \sum_{i=0}^{\text{free}(I)+\text{free}(J)-k-1} (-1)^i \frac{\binom{k-1+i}{k-1}}{(2^b-1)^{k-\text{free}(J)+i}}. \tag{16}$$

Lemma 10 assumes however a full-branch mixing layer, which is not the case for AES mixing. Another issue is that the number of layouts is still quite high and poses a non-trivial computational challenge. Thankfully, we can overcome this obstacle by representing the AES mixing layer in terms of permutations and full-branch mixings, an observation first made by [3]. More details can be found in the full version of this paper.

As our starting point, we numerically compute the total variation distance from uniform after r rounds starting with a pair of inputs that differ in exactly one 8-bit word. The results are summarized in Table 2, and are obtained by computing the corresponding r-th power of the transition matrix of the random walk. (This requires leveraging a number of symmetries to be computationally feasible.)

We then derive conjectures on the maximum distance over all possible input layouts and verify that our conjectures hold by computing the statistical distance for all input layouts. As a result of this, we obtain the following theorems.

Theorem 5. *The* 3*-round AES* is* $2^{-23.42}$*-close to pairwise independent.*

Theorem 6. *The* 7*-round AES* is* 2^{-128}*-close to pairwise independent.*

6.2 Censored AES

To translate our results from the random S-box setting to the AES S-box, we replace a random S-box by consecutive applications of the AES one, namely the patched inverse function over $\mathbb{F}_{2^8}$ where the input is XOR with a fresh key byte. Note that the resulting SPN which we refer to as "censored" AES is simply AES with several mixing layers removed.

We numerically compute the closeness to pairwise independence of the sequential composition of AES S-boxes over $\mathbb{F}_{2^8}$, where a fresh key byte is XORed into the input prior to each call. These distances can be found in the full version of this paper. Note that analytical bounds were obtained in [37], however here we obtain tighter numerical bounds for our parameter settings. We defer the implementation details to the full version of this paper.

Overall, we prove the following theorem. It considers what we (informally) refer to as "192-round censored AES." One should think of this as a 191-round SPN (thus with 192 layers of S-boxes), with independent keys, using the true AES S-box (patched inverse) and the AES mixing layer, but with a subset of mixing layers removed. Which mixing layers remain can be inferred from the proof below.

Theorem 7. 192*-round censored AES is* 2^{-128}*-close to pairwise independent.*

Proof. First off, Theorem 5 implies that 3-round AES* (that is, 4 layers of random S-boxes) is $\varepsilon_{ideal} = 2^{-23.42}$-close to pairwise independent. We then replace each random S-box with the sequential composition of c consecutive AES S-boxes (and xoring an independent uniform key byte to each call) and show that the resulting construction (which consists of $4c$ layers of S-boxes) is ϵ-close to pairwise independent, for some suitable ϵ. This value of ϵ will be then amplified, via further sequential composition. By the amplification theorem of [28,39], the resulting $4cr$-round censored AES is in particular $(2^{r-1}\epsilon^r)$-close to pairwise independent. The exact constants c and r are chosen to optimize the final number of rounds required to reach 2^{-128}-closeness.

First of all, we pick $c = 8$. Indeed, according to our findings in the full version, the 8-fold sequential composition of the S-box (with independent key bytes XORed to each S-box input) is $\varepsilon_{sim} \leq 2^{-29.39}$-close to pairwise independent, and hence to the behavior of a random S-box. Recall that the random S-box in AES* is applied to $k = 16$ blocks in parallel, hence by the triangle inequality we deduce that we can simulate 4 random S-box layers with an error of at most $16 \cdot 4 \cdot \varepsilon_{sim} \leq 2^{-23.39}$.

Therefore, we conclude that this partial 32-round censored AES is ϵ-close to pairwise independent for

$$\epsilon \leq \varepsilon_{ideal} + 16 \cdot 4 \cdot \varepsilon_{sim} \leq 2^{-23.42} + 2^{-23.39} < 2^{-22.39} .$$

Then, amplification for $r = 6$ repetitions gives that the 192-round censored AES is

$$2^5 \cdot (2^{-22.39})^6 = 2^{5-22.39 \cdot 6} < 2^{-128}$$

close to pairwise independent. □

If one believes that the mixing layers are useful for AES to achieve pseudorandomness, then it is natural to expect that removing a large fraction of them should only hurt the convergence to pairwise independence. This leads us to conjecture that 192-round AES is 2^{-128}-close to pairwise independent. We view proving this conjecture formally to be an outstanding open problem.

Acknowledgements. Pelecanos was supported by DARPA under Agreement No. HR00112020023. Tessaro was supported in part by NSF grants CNS-2026774, CNS-2154174, a JP Morgan Faculty Award, a CISCO Faculty Award, and a gift from Microsoft. Vaikuntanathan was supported by DARPA under Agreement No. HR00112020023, NSF CNS-2154149, and a Thornton Family Faculty Research Innovation Fellowship.

References

1. Advanced Encryption Standard (AES). National Institute of Standards and Technology, NIST FIPS PUB 197, U.S. Department of Commerce (Nov 2001)
2. Andreeva, E., Bogdanov, A., Dodis, Y., Mennink, B., Steinberger, J.P.: On the indifferentiability of key-alternating ciphers. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013. LNCS, vol. 8042, pp. 531–550. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40041-4_29

3. Baignères, T., Vaudenay, S.: Proving the security of AES substitution-permutation network. In: Preneel, B., Tavares, S. (eds.) SAC 2005. LNCS, vol. 3897, pp. 65–81. Springer, Heidelberg (2006). https://doi.org/10.1007/11693383_5
4. Biham, E., Shamir, A.: Differential cryptanalysis of des-like cryptosystems. J. Cryptol. **4**(1), 3–72 (1991)
5. Bogdanov, A., Khovratovich, D., Rechberger, C.: Biclique cryptanalysis of the full AES. In: Lee, D.H., Wang, X. (eds.) ASIACRYPT 2011. LNCS, vol. 7073, pp. 344–371. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25385-0_19
6. Bogdanov, A., Knudsen, L.R., Leander, G., Standaert, F.-X., Steinberger, J., Tischhauser, E.: Key-alternating ciphers in a provable setting: encryption using a small number of public permutations. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 45–62. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29011-4_5
7. Brodsky, A., Hoory, S.: Simple permutations mix even better. Random Struct. Algorithms **32**(3), 274–289 (2008). https://doi.org/10.1002/rsa.20194
8. Chen, S., Lampe, R., Lee, J., Seurin, Y., Steinberger, J.: Minimizing the two-round even-mansour cipher. In: Garay, J.A., Gennaro, R. (eds.) CRYPTO 2014. LNCS, vol. 8616, pp. 39–56. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44371-2_3
9. Chen, S., Steinberger, J.: Tight security bounds for key-alternating ciphers. In: Nguyen, P.Q., Oswald, E. (eds.) EUROCRYPT 2014. LNCS, vol. 8441, pp. 327–350. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-55220-5_19
10. Cogliati, B., et al.: Provable security of (tweakable) block ciphers based on substitution-permutation networks. In: Shacham, H., Boldyreva, A. (eds.) CRYPTO 2018. LNCS, vol. 10991, pp. 722–753. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96884-1_24
11. Cogliati, B., Seurin, Y.: On the Provable Security of the Iterated Even-Mansour Cipher Against Related-Key and Chosen-Key Attacks. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015. LNCS, vol. 9056, pp. 584–613. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46800-5_23
12. Coron, J.S., Holenstein, T., Künzler, R., Patarin, J., Seurin, Y., Tessaro, S.: How to build an ideal cipher: The indifferentiability of the Feistel construction. J. Cryptol. **29**(1), 61–114 (2016). https://doi.org/10.1007/s00145-014-9189-6
13. Courtois, N.T., Pieprzyk, J.: Cryptanalysis of block ciphers with overdefined systems of equations. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 267–287. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36178-2_17
14. Daemen, J.: Cipher and hash function design strategies based on linear and differential cryptanalysis. Ph.D. Thesis, KU Leuven (1995)
15. Dai, Y., Seurin, Y., Steinberger, J., Thiruvengadam, A.: Indifferentiability of iterated even-mansour ciphers with non-idealized key-schedules: five rounds are necessary and sufficient. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017. LNCS, vol. 10403, pp. 524–555. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63697-9_18
16. Dai, Y., Steinberger, J.: Indifferentiability of 8-round feistel networks. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016. LNCS, vol. 9814, pp. 95–120. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-53018-4_4
17. Dodis, Y., Karthikeyan, H., Wichs, D.: Small-box cryptography. In: Braverman, M. (ed.) 13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA. LIPIcs, vol. 215, pp. 56:1–56:25. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2022)

18. Dodis, Y., Katz, J., Steinberger, J., Thiruvengadam, A., Zhang, Z.: Provable security of substitution-permutation networks. Cryptology ePrint Archive, Report 2017/016 (2017). https://eprint.iacr.org/2017/016
19. Dodis, Y., Stam, M., Steinberger, J., Liu, T.: Indifferentiability of confusion-diffusion networks. In: Fischlin, M., Coron, J.-S. (eds.) EUROCRYPT 2016. LNCS, vol. 9666, pp. 679–704. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49896-5_24
20. Even, S., Mansour, Y.: A construction of a cipher from a single pseudorandom permutation. J. Cryptol. **10**(3), 151–162 (1997). https://doi.org/10.1007/s001459900025
21. Farshim, P., Procter, G.: The related-key security of iterated even–mansour ciphers. In: Leander, G. (ed.) FSE 2015. LNCS, vol. 9054, pp. 342–363. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48116-5_17
22. Guo, C., Lin, D.: On the indifferentiability of key-alternating feistel ciphers with no key derivation. In: Dodis, Y., Nielsen, J.B. (eds.) TCC 2015. LNCS, vol. 9014, pp. 110–133. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46494-6_6
23. Hoang, V.T., Tessaro, S.: Key-alternating ciphers and key-length extension: exact bounds and multi-user security. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016. LNCS, vol. 9814, pp. 3–32. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-53018-4_1
24. Hoory, S., Magen, A., Myers, S.A., Rackoff, C.: Simple permutations mix well. Theor. Comput. Sci. **348**(2–3), 251–261 (2005)
25. Jakobsen, T., Knudsen, L.R.: The interpolation attack on block ciphers. In: FSE. Lecture Notes in Computer Science, vol. 1267, pp. 28–40. Springer (1997). https://doi.org/10.1007/bfb0052332
26. Joan, D., Vincent, R.: The design of rijndael: Aes-the advanced encryption standard. Information Security and Cryptography (2002)
27. Kang, J.S., Hong, S., Lee, S., Yi, O., Park, C., Lim, J.: Practical and provable security against differential and linear cryptanalysis for substitution-permutation networks. Etri J. **23** (02 2002). https://doi.org/10.4218/etrij.01.0101.0402
28. Kaplan, E., Naor, M., Reingold, O.: Derandomized constructions of k-wise (almost) independent permutations. In: APPROX-RANDOM. Lecture Notes in Computer Science, vol. 3624, pp. 354–365. Springer (2005). https://doi.org/10.1007/s00453-008-9267-y
29. Kaplan, E., Naor, M., Reingold, O.: Derandomized constructions of k-wise (almost) independent permutations. Algorithmica **55**(1), 113–133 (2009). https://doi.org/10.1007/s00453-008-9267-y
30. Knudsen, L.: Deal - a 128-bit block cipher. In: NIST AES Proposal (1998)
31. Knudsen, L.R.: Truncated and higher order differentials. In: Preneel, B. (ed.) FSE 1994. LNCS, vol. 1008, pp. 196–211. Springer, Heidelberg (1995). https://doi.org/10.1007/3-540-60590-8_16
32. Knudsen, L., Wagner, D.: Integral cryptanalysis. In: Daemen, J., Rijmen, V. (eds.) FSE 2002. LNCS, vol. 2365, pp. 112–127. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45661-9_9
33. Lai, X.: Higher Order Derivatives and Differential Cryptanalysis, pp. 227–233. Springer, US, Boston, MA (1994). https://doi.org/10.1007/978-1-4615-2694-0_23
34. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 17–38. Springer, Heidelberg (1991). https://doi.org/10.1007/3-540-46416-6_2

35. Lampe, R., Patarin, J., Seurin, Y.: An asymptotically tight security analysis of the iterated even-mansour cipher. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012. LNCS, vol. 7658, pp. 278–295. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34961-4_18
36. Lampe, R., Seurin, Y.: Security Analysis of Key-Alternating Feistel Ciphers. In: Cid, C., Rechberger, C. (eds.) FSE 2014. LNCS, vol. 8540, pp. 243–264. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46706-0_13
37. Liu, T., Tessaro, S., Vaikuntanathan, V.: The t-wise independence of substitution-permutation networks. In: Malkin, T., Peikert, C. (eds.) CRYPTO 2021. LNCS, vol. 12828, pp. 454–483. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-84259-8_16
38. Matsui, M., Yamagishi, A.: A new method for known plaintext attack of FEAL cipher. In: Rueppel, R.A. (ed.) EUROCRYPT 1992. LNCS, vol. 658, pp. 81–91. Springer, Heidelberg (1993). https://doi.org/10.1007/3-540-47555-9_7
39. Maurer, U., Pietrzak, K., Renner, R.: Indistinguishability Amplification. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 130–149. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74143-5_8
40. Miles, E., Viola, E.: Substitution-permutation networks, pseudorandom functions, and natural proofs. In: Safavi-Naini, R., Canetti, R. (eds.) CRYPTO 2012. LNCS, vol. 7417, pp. 68–85. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32009-5_5
41. National Soviet Bureau of Standards: Information processing system - cryptographic protection - cryptographic algorithm gost 28147–89 (1989)
42. Nyberg, K.: Differentially uniform mappings for cryptography. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 55–64. Springer, Heidelberg (1994). https://doi.org/10.1007/3-540-48285-7_6
43. Patarin, J.: A proof of security in $O(2^n)$ for the Benes scheme. In: Vaudenay, S. (ed.) AFRICACRYPT 08. LNCS, vol. 5023, pp. 209–220. Springer, Heidelberg (Jun (2008)
44. Tessaro, S.: Optimally secure block ciphers from ideal primitives. In: Iwata, T., Cheon, J.H. (eds.) ASIACRYPT 2015. LNCS, vol. 9453, pp. 437–462. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48800-3_18
45. Tiessen, T., Knudsen, L.R., Kölbl, S., Lauridsen, M.M.: Security of the AES with a secret S-box. In: Leander, G. (ed.) FSE 2015. LNCS, vol. 9054, pp. 175–189. Springer, Heidelberg (Mar 2015). https://doi.org/10.1007/978-3-662-48116-5_9