Prouff and Rivain's Formal Security Proof of Masking, Revisited

Loïc Masure, François-Xavier Standaert

Prouff & Rivain's Formal Security Proof of Masking, Revisited

Tight Bounds in the Noisy Leakage Model

Loïc Masure and François-Xavier Standaert

ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
loic.masure@lirmm.fr

Abstract. Masking is a counter-measure that can be incorporated to software and hardware implementations of block ciphers to provably secure them against side-channel attacks. The security of masking can be proven in different types of threat models. In this paper, we are interested in directly proving the security in the most realistic threat model, the so-called *noisy leakage* adversary, that captures well how real-world side-channel adversaries operate. Direct proofs in this leakage model have been established by Prouff & Rivain at Eurocrypt 2013, Dziembowski *et al.* at Eurocrypt 2015, and Prest *et al.* at Crypto 2019. These proofs are complementary to each other, in the sense that the weaknesses of one proof are fixed in at least one of the others, and conversely. These weaknesses concerned in particular the strong requirements on the noise level and the security parameter to get meaningful security bounds, and some requirements on the type of adversary covered by the proof — *i.e.*, chosen or random plaintexts. This suggested that the drawbacks of each security bound could actually be proof artifacts. In this paper, we solve these issues, by revisiting Prouff & Rivain's approach.

1 Introduction

1.1 Context

Side-Chanel Analysis (SCA) represents an important threat for cryptographic implementations on embedded devices such as smart-cards, Micro-Controller Units (MCUs), *etc.* [37,38]. In such attacks, the adversary has a physical access to the target device.More precisely, the adversary is assumed to measure some physical metrics of the device called *leakages* — *e.g.* the power consumption of the device or the Electro-Magnetic (EM) emanations around the target — during one or several encryptions. It is then possible to use this side information — beside leveraging plaintexts and ciphertexts — to guess the values of *sensitive* variables, *i.e.* the values of intermediate calculations depending on some chunks of secret. This way, an SCA adversary may independently recover the secret in a divide-and-conquer approach, making the typical complexity of such attacks often negligible compared to a regular cryptanalysis. That is why the SCA threat

should carefully be taken into account in the design of cryptographic implementations.

Thankfully, this does not prevent the deployment and the use of embedded cryptography, as this threat can be mitigated by incorporating counter-measures in the implementation. At a very high level, most of the counter-measures such as *masking* or *shuffling* turn a deterministic cryptographic primitive into a non-deterministic implementation by injecting some randomness during the execution of the primitive, either at a physical level or at an algorithmic level. In this paper, we focus on the main counter-measure considered so far in SCA, namely masking [29,18], a.k.a. "Multi-Party Computation (MPC) on silicon" [34]. In a nutshell, any sensitive variable is submitted to a $(d+1)$-linear secret-sharing, where d is the security parameter that the designer may control in order to achieve the desired security level. The implementation is then modified in a way such that all the subsequent calculations involving a sensitive variable are now replaced by some *gadgets* operating on the shares separately, as in multi-party computation. As a result, any SCA adversary must have access to the *noisy* observation of every share of secret to be able to recover any piece of information about a sensitive variable. If any noisy observation induces some uncertainty on the actual value of the corresponding share, it results in an *amplified* uncertainty on the actual value of the target sensitive variable — an intuition that dates back to the seminal works of Chari *et al.* at Crypto 99 [18]. As a consequence, the complexity of any SCA attack increases exponentially fast — up to some limits — with the security parameter d, at the price of quadratic (or super-linear) runtime and memory overheads in the implementation only [34].

1.2 Provable Security of Masking

The latter intuition has been formalized over the past few years by masking security proofs. Generally speaking, a masking security proof takes as inputs an abstract representation of the implementation, the number of shares $d+1$ (where d act as he security parameter) and a measure of the noisiness of the leakage, usually characterized from the device embedding the implementation. The masking security proof then returns an upper bound on a metric depicting the security level of the implementation.

There exists different strategies to establish a masking security proof. In this paper, we focus on masking security bounds directly stated in the most realistic threat model. This approach has been first considered by Chari *et al.* [18], before being formalized by Prouff and Rivain [46]. Concretely, a *noisy* observation of an intermediate calculation may be turned into a conditional Probability Mass Function (p.m.f.) over all the hypothetical values that the operands may take: the closer the p.m.f. to the uniform distribution, the noisier the leakage.

The idea of security proofs in the noisy leakage model is to assume that any noisy leakage accessed by the adversary is δ-close to the uniform distribution, for some real-valued parameter δ stated in a metric that can be measured by the

practitioner.[1] Then, the goal is to prove that the p.m.f. of the secret key, given an access to the full leakage, is in turn ϵ-close to the p.m.f. that an adversary without access to side-channel would get, for some real-valued parameter ϵ depending on δ, the security parameter d, and some other specifications of the implementation.

This direct approach has gained the reputation of being "not convenient" [8,10] to work with, up to the point that most masking security proofs are now established in much simpler yet unrealistic threat models [34,6,8,9,17], relying on a non-tight reduction from the noisy leakage model to such simpler threat models [25]. As a result, only three previous works tackled masking security proofs through this direct way so far. These works, from Prouff and Rivain at Eurocrypt 2013 [46], Dziembowski *et al.* at Eurocrypt 2015 [27], and Prest *et al.* at Crypto 2019 [45], considered implementations of block ciphers protected with an Ishai-Sahai-Wagner (I.S.W.) masking scheme [34,48], assuming leak-free refreshings. The latter assumption is a drawback, as it is unrealistic — otherwise studying leaky computations would not be relevant — and some real-world refreshings could critically decrease the security level [21]. Interestingly, these three proofs are quite complementary to each other, in the sense that the weaknesses of one proof are fixed in at least one of the others, and conversely. We give hereafter a brief overview of these pros and cons — also synthesized in Table 1:

1. **Strong noise requirements [46].** Prouff and Rivain's bound required the baseline noise parameter δ to scale *polynomially* with the field size, which is prohibitive for concrete block ciphers, *e.g.*, the Advanced Encryption Standard (AES) whose field size is 256. On the opposite, Dziembowski *et al.*'s bound have a nearly tight noise requirement that does not depend on the field size.
2. **Lack of incentive for noisier leakage [27].** In Dziembowski *et al.*'s security bound assuming that the noise requirement is verified, the bound no longer depends on the actual baseline noise level δ. This suggests that to reach the desired security level ϵ, the designer would have no incentive in choosing a noisier device on which implementing the block cipher, which sounds unrealistic. In the extreme case where the device is so noisy enough that $\delta \leq \epsilon$, masking would not be necessary, whereas Dziembowski *et al.*'s bound would still require a prohibitive number of shares to be meaningful. On the opposite, the bounds of Prouff and Rivain and Prest *et al.* still carry some incentive towards noisier baseline leakage.
3. **Too conservative and hard to estimate metric [45].** Contrary to the other proofs, the baseline noise in Prest *et al.*'s security bound is assumed to be measured in a *worst-case* metric, the so-called Relative Error (RE). This contrasts with all the other works considering *average-case* metrics, such as the MI [46] or the SD [27], and does not fit either with SCA security metrics such as Guessing Entropy (GE) or Success Rate (SR) [51] that are averaged metrics as well. Using worst-case metrics has two main drawbacks. First, a baseline noise characterization made with a worst-case metric necessarily

[1] *e.g.*, the Statistical Distance (SD), the Euclidean Norm (EN), or the Mutual Information (MI). Notice that in our context, "noisier" means a *lower* δ.

results in more conservative requirements than with average-case metrics. Second, worst-case metrics are by definition harder to estimate on concrete devices by evaluators, and hereupon the RE may not be efficiently tractable — especially for high-dimensional leakage — nor even be formally defined in some cases. As an example, Prest *et al.* even needed to use tedious tail-cut arguments on the exemplary leakage distributions of their case study [45, Remark 2].

4. **Leakage from the gates vs. from the wires.** Beyond using different metrics from one past work to another, the formal modelization of an implementation may also differ. On the one hand, the works of Prouff and Rivain originally took inspiration from the "only computations leak" paradigm, in which a cryptographic computation is split into a sequence of elementary operations that each leaks information on the accessed part of the device state [47]. On the other hand, Duc *et al.* and all their subsequent works have assumed the leakage observed by the adversary to be occured by the *wires* of the device storing some intermediate values of the computation. While the latter one can be encompassed by the former one [11, Lemma 1], the past literature suggested that these different views might result in different security levels [25, Sec. 5.5]. Whether such differences were actually proof artifacts was not widely discussed in the literature, so it remains an open problem.
5. **Random message attacks [46].** Last but not least, Prouff and Rivain's security bounds are given for random message attacks, whereas Dziembowski *et al.* and Prest *et al.* state security bounds for chosen plaintext attacks. Even if most of state-of-the-art SCA adversaries consider random plaintext attacks, this contrasts with the common practice in cryptography, where the adversary is assumed to (adaptively) choose the message or the ciphertext.

Table 1: Comparison between all proofs in the Noisy Leakage model: Prouff & Rivain [46], Dziembowski *et al.* [27], Prest *et al.* [45].

Feature	[46]	[27]	[45]	Our work
Strong noise requirement	Yes	No	No	No
Leak-free refreshing	Yes	Yes	Yes (Sec. 6)	Yes
Incentive to small δ	✓	✗	✓	✓
Average-case metric	✓	✓	✗	✓
Adaptive attacks	✗	✓	✓	✓

1.3 Recent Improvements on Security Bounds for Encodings Only

In light of the previous drawbacks listed so far, Duc *et al.* conjectured at EUROCRYPT 2015 that the weaknesses (1-3) were actually proof artifacts [26]. More

precisely, it would be possible to prove a masking security bound in terms of MI with tight noise requirement, and tight amplification rates, while covering the leakage of the full block cipher. In a recent line of works, Ito *et al.* [35], Masure *et al.* [42], and Béguinot *et al.* [15] have been able to prove a reduced version of Duc *et al.*'s conjectured security bound, for the leakage of one encoding *only*. While these works represent a first milestone, they were limited in that they did not cover the leakage coming from the *computations*, and Duc *et al.*'s conjecture remained to be proven for the leakage of a full block cipher.

1.4 Our Contribution

In this paper, we prove new masking security bounds stated in the noisy leakage model, in the same setting as the one of the previous works discussed so far — namely Rivain-Prouff's masking scheme, with leak-free refreshings [46]. To this end, we revisit Prouff and Rivain's approach, by showing that some drawbacks of their results can be circumvented.

- **A tight bound with respect to the noise parameter δ.** We leverage the recent results of Ito *et al.* [35], Masure *et al.* [42] and Béguinot *et al.* [15], to bound the amount of informative leakage of computations coming from a full block cipher, masked with an I.S.W.-like masking scheme. In our result, we consider the two cases where the leakage comes from the wires of comes from the gates, which results into different security bounds with a non-trivial gap. Nevertheless, our noise requirement is tight in both cases [33], while carrying a much higher incentive to noisier leakage than in the previous works.
- **A security bound with low dependency on the field size.** With the previous contribution alone, our final security bound would still carry a constant factor scaling *quadratically* with the size of the field over which the block cipher operates, regardless of the number of shares. While this is much better than Prouff & Rivain's bound and competitive with Dziembowski *et al.*'s bound, this still sounds unnatural, as it does not perfectly fit Duc *et al.*'s conjecture [26], and might be fatal for block ciphers operating over large fields. To tackle this problem, we show how a careful scrutiny of the implementation, under mild assumptions on the Sbox, can allow us to make this constant factor *quasi-linear* with the field size.
- **Security Bound with Average Metric.** In our masking security proof, any metric, be it the baseline noise δ or the final security bound ϵ, is expressed in MI. This contrasts with Prouff & Rivain's work where the parameters δ and ϵ are not expressed in the same metric. Since MI is an averaged metric, it is quite easy to estimate by evaluators when characterizing the behavior of the target device in worst-case evaluations [4].
- **Attacks with Chosen Messages.** Eventually, we argue how our security bounds stated for random plaintext attacks can be extended to the case of chosen plaintext attacks, using a similar argument as the one stated by Dziembowski *et al.* in their follow-up work at Tcc 2016 [28].

Overall, our work is the first to state a masking proof with meaningful security bounds, *i.e.*, for which the desired security level can be reached with a reasonable amount of masking shares, and requiring a reasonable amount of noise from the device. Therefore, our masking security bound can be practically used by an SCA evaluator to upper bound current state-of-the-art SCA adversaries. This suggests that masking proofs directly stated in the noisy leakage model can be seen as complementary to the more generic proofs in other threat models. The only shortcoming of our proof, in line of the previous works, concerns the use of leak-free refreshings. We hope future works may allow to relax this assumption, and thereby provide a comparable setting with masking security proofs in the indirect approach taking advantage of reductions between models.

Erratum. A previous version of this paper, published on eprint and included in the proceedings of Crypto 2023, contained two flaws: one on Theorem 5 and another one in a section formerly situated between section 3 and section 4. We would like to testify our gratefulness to Julien Béguinot (Télécom Paris) for spotting the flaw in the proof of section 3 [16] and Jürgen Pulkus (G+D) for kindly pointing out the second flaw.

Updates in the Extended Version. We have addressed the first flaw by correcting the proof of Theorem 5 and modifying the statement of the theorem accordingly, it lead us to revisit the difference between leakages from the gates and leakages from the wires, which results now in a discussion in subsection 4.1. Finally, we have removed the flawed section as it did not look fixable. This section was independent, so it did not affect the remaining of the paper.

2 Preliminaries

In this paper, we denote sets by calligraphic letters, *e.g.*, $\mathcal{X}$. In particular, the letter $\mathcal{Y}$ denotes a finite field $(\mathcal{Y}, \oplus, \times)$ of characteristic two. Upper-case letters are used to denote random variables, while lower-case letters denote observations of random variables. In this paper, we adopt the following convention: A, B stand for independent random variables uniformly distributed over $\mathcal{Y}$, while G, H denote random variables that are not necessarily uniform over $\mathcal{Y}$, nor assumed to be independent. The letter L will be used to denote a randomized function $\mathcal{Y} \to \mathcal{L}$, were the set $\mathcal{L}$ is assumed without loss of generality to be discrete. When the context does not carry any ambiguity, we will often denote the random variable $\mathrm{L}(\mathrm{Y})$ by omitting the reference to Y. Finally, bold letters denote vectors of random variables.

Mutual Information. Let $\mathrm{Y} \in \mathcal{Y}$ be a discrete random variable. The *entropy* of Y, denoted by $\mathsf{H}(\mathrm{Y})$, defined by: $\mathsf{H}(\mathrm{Y}) = -\sum_{s \in \mathcal{Y}} \Pr(\mathrm{Y} = s) \log_2 \Pr(\mathrm{Y} = s)$. Moreover, we define MI between two discrete random variables Y and L as:

$$\mathsf{MI}(\mathrm{Y}; \mathrm{L}) = \mathsf{H}(\mathrm{Y}) - \mathop{\mathbb{E}}_{l} \left[\mathsf{H}(\mathrm{Y} \mid \mathrm{L} = l) \right] .$$

2.1 Model of Noisy Leaking Computation

We describe hereafter the frame in which Prouff and Rivain's result is established, that is mostly adapted from their seminal work [46].[2]

Block Cipher. A block cipher over a finite field $\mathcal{Y}$ is defined by a pair of inputs $\mathbf{K}, \mathbf{P}$ seen as vectors of $\mathcal{Y}$, and by a sequence of T *elementary calculations* $(\mathsf{C}_i)_{1 \leq i \leq T}$ defined either over $\mathcal{Y}$ or $\mathcal{Y} \times \mathcal{Y}$. More precisely, since $\mathcal{Y}$ is assumed to be a finite field, we consider the elementary calculations to be either an addition $\oplus$ or a field multiplication $\times$, whether the operands are constant or random variables.[3]

Leakage and SCA Adversary. When processed on some input Y (resp. a pair of inputs A, B), an elementary calculation C_i reveals $\mathrm{L}_i(\mathrm{Y})$ (resp. $\mathrm{L}(\mathrm{A}, \mathrm{B})$) to the adversary, for some *noisy leakage* function L_i, that depends both on Y (resp. A, B), and on some internal randomness assumed to be drawn independently each time L_i leaks. Whenever the context does not carry any ambiguity, we may simply denote the leakage $\mathrm{L}_i(\mathrm{Y})$ by L_i. In this paper, we consider an adversary having access to the full leakage induced by each elementary calculation and trying to recover a chunk of secret key.

Definition 1 (SCA key recovery adversary). *An* SCA adversary *for a block cipher defined over $\mathcal{Y}$ is an algorithm that, upon a sequence of N_a plaintexts $\mathbf{P} = (P_1, \ldots, P_{N_a})$, takes as an input a sequence $\{(L_i)_{1 \leq i \leq T}\}_{1 \leq j \leq N_a}$ of leakages induced by each elementary calculation of a block cipher, and that returns a guess $\widehat{K}$ of one chunk $K \in \mathcal{Y}$ of the secret key $\mathbf{K}$. We say that the adversary is* random-plaintext *if $\mathbf{P}$ is chosen randomly and uniformly over $\mathcal{Y}^{N_a}$, whereas we say that the adversary is* chosen-plaintext *if the adversary can arbitrarily choose the sequence $\mathbf{P}$ — possibly adaptively.*

Notice that $\widehat{\mathrm{K}}$ depends on the plaintexts used by the adversary (and on the internal randomness of the leakage functions). Accordingly, the accuracy of the key guessing is expected to increase with the number N_a of queries. We formalize this in the definition hereafter.

Definition 2 (Success Rate). *The* success rate *of an SCA key recovery adversary is the quantity*

$$\mathsf{SR}(N_a) = \Pr\left(\widehat{K} = K\right) \quad . \tag{1}$$

Similarly, for any probability threshold $\frac{1}{|\mathcal{Y}|} \leq \beta \leq 1$, we define the efficiency *$N_a^\star(\beta)$ of an SCA key recovery adversary as the minimal amount of queries necessary to get a success rate higher than β.*

[2] The interested reader may also refer to Rivain's habilitation thesis for a thorough discussion about the leakage model [47].

[3] As argued by Prouff & Rivain, any mapping over a finite field can be decomposed as a sequence of additions and multiplications, using Lagrange interpolation.

MI-Noisy Leakage. The success of an SCA key recovery adversary depends on how informative the leakage is about the underlying secret data processed. To measure this, we assume that the evaluator may determine how *noisy* any leakage function is. To this end, we formally define hereafter the concept of MI-noisy leakage.

Definition 3 (Noisy leakage of one input). *Let $L : \mathcal{Y} \to \mathcal{L}$ be a leakage function. L is said to be δ-*MI-noisy*, for some $\delta \geq 0$, if for a uniformly distributed random variable $A \in \mathcal{Y}$,*

$$\mathsf{MI}(A; L(A)) \leq \delta \ .$$

Definition 3 allows for example to measure the amount of informative leakage that occurs when processing an intermediate computation $\mathrm{A} \to \mathsf{C}(\mathrm{A})$. Likewise, we may define hereafter δ-noisy leakages with respect to two random variables, *e.g.*, when processing a binary gate $\mathrm{A}, \mathrm{B} \to \mathsf{C}(\mathrm{A}, \mathrm{B})$.

Definition 4 (Noisy leakage of two inputs). *Let $L : \mathcal{Y}^2 \to \mathcal{L}$ be a leakage function with two inputs. L is said to be δ-*MI-noisy*, for some $\delta \geq 0$, if for any input random variables A, B of C, independent and uniformly distributed over $\mathcal{Y}$,*

$$\mathsf{MI}(A, B; L(A, B)) \leq \delta \ .$$

Leakage from the Gates vs. Leakages from the Wires. Definition 4 is the generic way to define noisy leakage occurring *from the gates*, as formalized by Prouff and Rivain. In a sense, it consists in viewing a binary gate with inputs over $\mathcal{Y}$ as a unary gate with inputs over $\mathcal{Y}^2$. However, Duc *et al.* have introduced in their groundbreaking work another paradigm, namely that the leakage is assume to occur *from the wires* [25].

Definition 5 (Noisy leakage from the wires). *Let $\mathsf{C} : \mathcal{Y}^2 \to \mathcal{Y}$ be an elementary calculation associated with the leakage function L. $\mathbf{L}$ is said to be a leakage from the wires of $\mathsf{C}(A, B)$ if there exist three functions L', L'' and L''', such that*

$$\mathbf{L}(A, B) = \big(L'(A), L''(B), L'''(\mathsf{C}(A, B))\big) \ ,$$

and such that conditionally to A and B, L', L'' and L''' are independent. Moreover, $\mathbf{L}$ is said to be $(\delta', \delta'', \delta''')$-noisy if L', L'' and L''' are respectively δ', δ'' and δ'''-noisy.

It is straightforward to see from Definition 5 that the "leaky wires" approach is encompassed as a particular case into the "leaky gates" point of view, while the inverse is not true. We will see later in this paper that those points of view lead to different security bounds.

On the Choice of the Metric. We chose the MI as a metric of reference in our proof, because it is at the core of Prouff & Rivain's security bound that we revisit in this paper, and also because we can therefore rely on the recent improvement of Ito *et al.* [35], Masure *et al.* [42] and Béguinot *et al.* [15]. Moreover, the MI

is known to be tightly linked to the complexity of Differential Power Analysis (DPA) attacks [39,40,41,24,19], and "generally carries more intuition (see, *e.g.*, [5] in the context of linear cryptanalysis)" [26]. We discuss this choice of metric in section 4.

2.2 Rivain-Prouff's Masking Scheme

We recall hereafter the definition of masking, mostly taken from Prouff and Rivain's paper [46, Def. 2].

Definition 6. *Let d be a positive integer. The* d-encoding of $\mathrm{Y} \in \mathcal{Y}$ *is a* $(d+1)$-*tuple* $(Y_i)_{0 \leq i \leq d}$ *satisfying* $\bigoplus_{i=0}^{d} Y_i = Y$ *and such that for any strict subset* $\mathcal{I}$ *of* $[\![0, d]\!]$, $(Y_i)_{\mathcal{I}}$ *is uniformly distributed over* $\mathcal{Y}^{|\mathcal{I}|}$.

The parameter d in Definition 6 refers here to the security parameter of the counter-measure. In their paper, Prouff and Rivain explain how to turn any block cipher into a d-order secure implementation — *i.e.* such that any intermediate computation depending on a secret has a $(d+1)$-encoding [46]. First, the plaintext and the secret key are split into $d+1$ shares. Then, each elementary calculation of the block cipher is transformed as follows. If the elementary calculation is linear with respect to its inputs, then it is replaced by the sequence of elementary calculations listed in Algorithm 1. If the elementary calculation is

Algorithm 1 Linear gadget in Prouff & Rivain's proof.

Require: $\mathbf{A}$: $(d+1)$-sharing of A, C: elementary calculation linear with its input.
Ensure: $\mathbf{B}$: $(d+1)$-sharing of C(A).
1: **for** $i = 0, \ldots, d$ **do**
2: $\quad \mathrm{B}_i \leftarrow \mathsf{C}(\mathrm{A}_i)$ ▷ Type 1 or 2
3: **end for**
4: $\mathbf{B} \leftarrow \mathsf{Refresh}(\mathbf{B})$ ▷ Assumed to be leak-free
5: $\mathbf{A} \leftarrow \mathsf{Refresh}(\mathbf{A})$ ▷ Only if A used subsequently.

an Sbox, then it can first be decomposed as a sequence of linear calculations and field multiplications. Then the linear calculations can be processed as in Algorithm 1, and the field multiplications can be replaced by the procedure listed in Algorithm 2. It is a variant of the actual I.S.W. scheme revisited by Rivain and Prouff at CHES 2010, up to a permutation between independent operations, so it does not change the amount of informative leakage. Overall, Rivain-Prouff's masked implementation can be decomposed as subsequences of any of the following types:

1. $(z_i \leftarrow g(x_i))_{0 \leq i \leq d}$, with g being a linear function (of the block-cipher);
2. $(z_i \leftarrow g(x_i))_{0 \leq i \leq d}$, with g being an affine function (within an Sbox evaluation);
3. $(v_{i,j} \leftarrow a_i \times b_j)_{0 \leq i,j \leq d}$ (cross-products computation step in multiplication);

Algorithm 2 Multiplication gadget in Prouff & Rivain's proof.

Require: $\mathbf{A}, \mathbf{B}$: $(d+1)$-sharing of A, B.
Ensure: C : $(d+1)$-sharing of A $\times$ B.
1: **for** $i = 0, \ldots, d$ **do**
2: **for** $j = 0, \ldots, d$ **do**
3: $\mathrm{V}_{i,j} \leftarrow \mathrm{A}_i \times \mathrm{B}_j$ ▷ Cross products (type 3)
4: **end for**
5: **end for**
6: $\mathbf{V} \leftarrow \mathsf{Refresh}(\mathbf{V})$ ▷ Assumed to be leak-free
7: **for** $i = 0, \ldots, d$ **do**
8: $\mathrm{C}_i = 0$
9: **for** $j = 0, \ldots, d$ **do**
10: $\mathrm{C}_i \leftarrow \mathrm{C}_i \oplus \mathrm{V}_{i,j}$ ▷ Compression (type 4)
11: **end for**
12: **end for**
13: $\mathbf{C} \leftarrow \mathsf{Refresh}(\mathbf{C})$ ▷ Assumed to be leak-free
14: $\mathbf{A}, \mathbf{B} \leftarrow \mathsf{Refresh}(\mathbf{A}), \mathsf{Refresh}(\mathbf{B})$ ▷ Only if A, B used subsequently.

4. $(t_{i,j} \leftarrow t_{i,j-1} \oplus v_{i,j})_{0 \leq i,j \leq d}$ (compression step multiplication).

For concreteness, we list two examples of schemes of the AES Sbox (at least its non-linear part) with this method in Algorithms 3 and 4. Algorithm 3 is the one initially proposed by Rivain and Prouff at CHES 2010. Recently, Cardoso *et al.* proposed at CARDIS 2022 an alternative exponentiation scheme depicted in Algorithm 4 which, combined with other implementation tricks, improved upon Rivain-Prouff's exponentiation [49]. Both exponentiations contain the same number of I.S.W. multiplications.[4]

Algorithm 3 R&P's $\mathsf{Exp254}$ [48].

Require: $\mathbf{X}$: $(d+1)$-sharing of X
Ensure: $\mathbf{C}$: $(d+1)$-sharing of X^{254}
1: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{X})$ ▷ $\mathrm{Z} = \mathrm{X}^2$
2: $\mathbf{X} \leftarrow \mathsf{Refresh}(\mathbf{X})$
3: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^3$
4: $\mathbf{V} \leftarrow \mathsf{SecLin}(s \mapsto s^4, \mathbf{Y})$ ▷ $\mathrm{V} = \mathrm{X}^{12}$
5: $\mathbf{V} \leftarrow \mathsf{Refresh}(\mathbf{V})$
6: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{V})$ ▷ $\mathrm{Y} = \mathrm{X}^{15}$
7: $\mathbf{Y} \leftarrow \mathsf{SecLin}(s \mapsto s^{16}, \mathbf{Y})$ ▷ $\mathrm{Y} = \mathrm{X}^{240}$
8: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{W})$ ▷ $\mathrm{Y} = \mathrm{X}^{252}$
9: $\mathbf{C} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{Z})$ ▷ $\mathrm{C} = \mathrm{X}^{254}$

Algorithm 4 Cardoso's $\mathsf{Exp254}$ [49].

Require: $\mathbf{X}$: $(d+1)$-sharing of X
Ensure: $\mathbf{C}$: $(d+1)$-sharing of X^{254}
1: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{X})$ ▷ $\mathrm{Z} = \mathrm{X}^2$
2: $\mathbf{Z} \leftarrow \mathsf{Refresh}(\mathbf{Z})$
3: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^3$
4: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{Y})$ ▷ $\mathrm{Y} = \mathrm{X}^6$
5: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^7$
6: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{Y})$ ▷ $\mathrm{Z} = \mathrm{X}^{14}$
7: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^{15}$
8: $\mathbf{Y} \leftarrow \mathsf{SecLin}(s \mapsto s^{16}, \mathbf{Y})$ ▷ $\mathrm{Y} = \mathrm{X}^{240}$
9: $\mathbf{C} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{Z})$ ▷ $\mathrm{C} = \mathrm{X}^{254}$

3 Revisiting Prouff and Rivain's Bound

We are now ready to revisit Prouff and Rivain's formal security proof in this section. To this end we briefly recall the outline of their proof — that we follow

[4] There are other generic methods to securely compute an Sbox with masking [32], which are out of the scope of this paper.

as well — based on three steps. First, they leverage the assumption that refresh gadgets are leak-free in order to reduce the MI of a sequence of elementary computations to the sum of the MIs between the secret and each subsequence of leakage. Second, some of these elementary computations — *e.g.*, the non-linear operations of the Sbox — may process non-uniform secrets. That is why the authors make an intermediate reduction to the case where every elementary computation processes uniform secrets — and mutually independent as well, in the case of binary gates. Finally, the authors apply some noise amplification lemma from the literature. Our revisited proof applies the same outline. We now dig into the details of these steps.

3.1 Step 1: Decomposition into Subsequences

We first recall that the MI of a sequence of mutually independent leakages can be bounded by the sum of MIs of each leakage.

Theorem 1 (Subsequence decomposition [46]). *Let Y be a random variable over a finite set $\mathcal{Y}$, not necessarily uniform. Let $\mathbf{L} = (L_1, \ldots, L_t)$ be t random variables such that the random variables $(L_i \mid Y = y)_i$ are mutually independent for every $y \in \mathcal{Y}$. Then, we have*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \sum_{i=1}^{t} \mathsf{MI}(Y; L_i) \quad . \tag{2}$$

Although we do not claim any improvement in this first step, we reproduce the proof in section B for completeness.

3.2 Step 2(a): Reduction to Uniform Secrets for Unary Gates

We now revisit the second step of Prouff and Rivain's work, namely the reduction from non-uniform secrets to uniform secrets. To this end, we will split our results into two cases. The first case processed in this subsection deals with non-uniform inputs of unary calculations, such as Line 4 in Algorithm 3. The second case deals with non-uniform and non-independent inputs of binary calculations, such as Line 6 in Algorithm 3, and will be deferred in subsection 3.3.

The results presented in this section aim at bounding the MI between $\mathsf{C}(Y)$, where $\mathsf{C} : \mathcal{Y} \to \mathcal{Y}$ and its corresponding leakage. We first state the following theorem that relies on a technical lemma from Shulman and Feder [50].

Theorem 2 (Generic Bound for Non-Uniform Secrets [50, p. 1360]). *Let $L : \mathcal{Y} \to \mathcal{L}$ be a random function denoting a leakage, and let Y be uniformly distributed over $\mathcal{Y}$. Then, there exists a constant α such that for all random variables G arbitrarily distributed over $\mathcal{Y}$, the following inequality holds true:*

$$\mathsf{MI}(G; L(G)) \leq \alpha \cdot |\mathcal{Y}| \cdot \mathsf{MI}(Y; \mathbf{L}(Y)) \quad . \tag{3}$$

Moreover, the smallest value α such that Equation 3 holds true belongs to the interval $\alpha \in \left[\frac{\log_2(e)}{e}, 1 - e^{-1}\right] \approx [0.53, 0.63]$.

Theorem 2 introduces an overhead scaling with $|\mathcal{Y}|$, which could decrease the final security level by one or several orders of magnitude (*e.g.*, for the AES, $|\mathcal{Y}| = 2^8$). Note that Equation 3 is nearly tight in the general case, in the sense that the range of α is narrow. Shulman and Feder exhibit an example of worst case leakage function, such that Equation 3 becomes an equality, for $\alpha \approx 0.53$ [50].

The Power Map Trick. However, such worst-case C functions are not likely to be used in cryptographic primitives, since, *e.g.*, the input and output of Sbox are expected to be uniformly distributed, for cryptographic reasons. That is why we refine hereafter the generic statement of Theorem 2, and we present some examples where this refinement could remove the dependency on the field size. To this end, we revisit Theorem 2 by relying on an intermediate result of Shulman and Feder's proof.

Lemma 1 ([50, Lemma 6]). *Given a leakage function L and two random variables Y, Y' distributed (non-necessarily uniformly) over the finite set $\mathcal{Y}$, and such that the support of $\Pr(Y')$ contains the support of $\Pr(Y)$. Then, the following inequality holds:*

$$\frac{\mathsf{MI}(Y; L(Y))}{\mathsf{MI}(Y'; L(Y'))} \geq \min_{y \in \mathcal{Y}} \frac{\Pr(Y = y)}{\Pr(Y' = y)} \ .$$

As a result, we straightforwardly get the following corollary.

Corollary 1. *In the same setting as in Lemma 1, if now the support of $\Pr(Y)$ contains the support of $\Pr(Y')$, the following inequality holds true:*

$$\frac{\mathsf{MI}(Y'; L(Y'))}{\mathsf{MI}(Y; L(Y))} \leq \max_{y \in \mathcal{Y}} \frac{\Pr(Y' = y)}{\Pr(Y = y)} \ . \tag{4}$$

Proof. Straightforward, using Lemma 1 and the identity $\max_{x \in \mathcal{X}} x = \frac{1}{\min_{x \in \mathcal{X}} \frac{1}{x}}$, for some finite ordered set $\mathcal{X}$. □

We will leverage Corollary 1 in the case where the Sbox is a monomial, *i.e.* is of the shape $y \mapsto y^k$. Admittedly, this makes our proof slightly more specific than Prouff and Rivain's one, as the latter one can handle any Sbox expressed as a polynomial. Nevertheless, this assumption remains mild, as it covers many Sboxes used in practical ciphers, including the AES, and will allow us to remove a constant factor equal to the field size.

We have seen in Algorithms 3 and 4 that the monomial $y \mapsto y^k$ can be computed in the Rivain-Prouff masking scheme by computing intermediate power maps $y \mapsto y^{k'}$ for some $k' \leq k$, through some square-and-multiply schemes [48]. The bound on the leakage induced by such an intermediate computation is handled by the following corollary.

Corollary 2. *Let Y be a uniform random variable over a finite field $\mathcal{Y}$ of size $M \geq 2$. For any $k \in [\![1, M-1]\!]$, define the function $\mathsf{C} : y \in \mathcal{Y} \mapsto y^k$. Let $L : \mathcal{Y} \to \mathcal{L}$ be a δ-MI-noisy leakage. Then:*

$$\mathsf{MI}\Big(Y; L(Y^k)\Big) \leq \gcd\{k, M-1\} \cdot \delta \ . \tag{5}$$

Proof. Using the Data Processing Inequality (DPI) (stated in Lemma 2 in Appendix A), we are reduced to upper bound $\mathsf{MI}\Big(\mathrm{Y}^k; \mathrm{L}(\mathrm{Y}^k)\Big)$. To this end, we shall compute the p.m.f. of Y^k. The result will then follow from Lemma 1 and Lemma 2. First, notice that by definition of a field, $y^k = 0$ if and only if (i.f.f.) $y = 0$, so $\Pr\Big(\mathrm{Y}^k = 0\Big) = \frac{1}{M}$. Second, notice that since $(\mathcal{Y}, \oplus, \times)$ is a finite field, the group $(\mathcal{Y}^*, \times)$ is cyclic, hence isomorphic with $\mathbb{Z}_{M-1}$. As a result, for any $s \neq 0$ for which there exists $y \in \mathcal{Y}$ verifying $y^k = s$, we have

$$\Pr\Big(\mathrm{Y}^k = s\Big) = \frac{\gcd(k, M-1)}{M} \ ,$$

and $\Pr\Big(\mathrm{Y}^k = s\Big) = 0$ otherwise. To summarize, for all $s \in \mathcal{Y}$, we have

$$\frac{\Pr\Big(\mathrm{Y}^k = s\Big)}{\Pr(\mathrm{Y} = s)} \leq \frac{M}{M} \cdot \gcd(k, M-1) \ . \tag{6}$$

□

Comparing the universal bound of Equation 3 to the specific bound in Equation 5, we can see that we replaced a factor $0.63 \cdot M$ by a factor $\gcd(k, M-1)$. As an example Table 2 reports the different constant factors induced by Equation 5 for the exponentiation scheme of Algorithms 3 and 4, and how they compare to the generic bound of Equation 3. Our power-map-specific bound is between one and two orders of magnitude lower than the generic bound in Equation 3.

Table 2: Factor overheads from Equation 5, and ratio between the generic bound of Equation 3 and the refined bound of Equation 5.

Scheme	k	$\gcd(k, 255)$	$\frac{(1-e^{-1})\cdot 256}{\gcd(k,255)}$
Rivain-Prouff [48]	$2, 3, 12,$ $15, 240, 252$	$1, 3, 3,$ $15, 15, 3$	$\mathbf{161.3}, 53.8, 53.8,$ $\mathbf{10.8}, 10.8, 53.8$
Cardoso *et al.* [49]	$2, 3, 6, 7,$ $14, 15, 240, 252$	$1, 3, 3, 1,$ $1, 15, 15, 3$	$\mathbf{161.3}, 53.8, 53.8, 161.3,$ $161.3, \mathbf{10.8}, 10.8, 53.8$

Admittedly, the numbers reported in Table 2 depend on the exponentiation scheme, and thereby depend on the underlying power-map we aim at computing

— which may differ for other block ciphers with power-map-based Sbox beyond the AES. We may therefore wonder how $\gcd(k, M-1)$ generally scales when M grows. It is not hard to find some integer k such that $\gcd(k, M-1)$ scales linearly with M,[5] so our improved bound could marginally improve the one from Equation 3 in some worst-case exponentiation schemes. Still, the following theorem suggests that this is not likely to happen.

Theorem 3 ([13, Thm. 3.2]). *Let $M > 2$ be an integer. Then, for all $\epsilon > 0$, we have $\mathop{\mathbb{E}}_{k}[\gcd(k, M)] = \mathcal{O}(M^{\epsilon})$, where the expectation is taken with respect to k uniformly distributed in $[\![1, M]\!]$.*

The practical interpretation of Theorem 3 is that if a given exponentiation scheme gives high constant factors, then it should not be hard to modify it, in order to make the constant factor in the right hand-side of Equation 5 arbitrarily low. As a consequence, we may treat the right hand-side of Equation 5 as asymptotically independent of M with high probability. That is why in the remaining of this paper, we will abuse notation by denoting any gcd factor as scaling as $\mathcal{O}(M^{\epsilon})$ — which is confirmed on our implementations of interest by Table 2.

3.3 Step 2(b): Reduction to Uniform Secrets for Binary Gates

We have shown in subsection 3.2 how to significantly decrease the loss in the reduction from non-uniform secrets to uniform secrets for leakage coming from unary gates dealing with power maps. In order to have a complete toolbox for reductions to uniform secrets, we also need to deal with leakages coming from gadgets with two input operands, *e.g.*, I.S.W. multiplications. Hereupon, Theorem 2 straightforwardly applies, although spanning a loss of $0.63\,|\mathcal{Y}|^2$ in the reduction.

That is why we may naturally think of extending the power map trick introduced before. But contrary to Theorem 2, Corollary 2 does not extend as straightforwardly for binary gates. Indeed, calculations with more than one operand add another difficulty: not only the operands may not be uniformly distributed, but they might also be non-independent. This results in the following corollary.

Corollary 3. *Let Y be a random variable uniformly distributed over the finite field $\mathcal{Y}$. For $p, q \in [\![2, M-2]\!]$, let $\mathbf{Z} = (Y^p, Y^q)$. Let $L : \mathcal{Y}^2 \to \mathcal{L}$ be a δ-MI-noisy leakage. Then,*

$$\mathsf{MI}(Y; L(\mathbf{Z})) \leq \min\{\gcd(p, M-1), \gcd(q, M-1)\} \cdot M \cdot \delta \ . \tag{7}$$

Proof. We apply Lemma 1 for the random vector $\mathbf{Z}' = (\mathrm{Y}, \mathrm{Y}')$, where Y' is an independent copy of Y. For any $x, y \in \mathcal{Y}$, the total probability formula implies that

$$\frac{\Pr(\mathrm{Y}^p = x, \mathrm{Y}^q = y)}{\Pr\big(\mathrm{Y} = x, \mathrm{Y}' = y\big)} \leq \frac{\sum_{y'} \Pr(\mathrm{Y}^p = x, \mathrm{Y}^q = y')}{\Pr\big(\mathrm{Y} = x, \mathrm{Y}' = y\big)} = \frac{\Pr(\mathrm{Y}^p = x)}{\Pr(\mathrm{Y} = x)\Pr\big(\mathrm{Y}' = y\big)} \ .$$

[5] As an example, for the AES field $M - 1 = 255$, which is divided by 3 so there exists some k, *e.g.*, $k = 85$, such that $\gcd(k, M-1) = \frac{M-1}{3}$.

Using Equation 6, we get that

$$\frac{\Pr(\mathrm{Y}^p = x, \mathrm{Y}^q = y)}{\Pr(\mathrm{Y} = x)\Pr\big(\mathrm{Y}' = y\big)} \leq \gcd(p, M-1) \cdot M \ . \tag{8}$$

By symmetry, we can obtain the same bound by permuting the roles of p and q, which gives Equation 7. □

Remark 1. Note that the inequality in Equation 8 is tight, *e.g.*, if p divides q, or inversely. Likewise, we argued that Equation 3 is generally tight — unless considering further assumptions on the prior distribution. Nevertheless, both facts do not necessarily imply that Equation 7 is tight. Whether the latter inequality could be refined for binary gates with non-independent operands remains an open-question that we will briefly discuss in subsection 3.4.

3.4 Step 3: The Amplification Theorems

We now revisit the third step of Prouff & Rivain's approach. To this end, like in subsection 3.2 and subsection 3.3, we make a discrepancy between the unary gates and the binary gates.

For Unary Gates. The following amplification theorem is at the core of our direct proof in the noisy leakage model, and holds the name of Mrs. Gerber's Lemma (MGL). It has initially been stated by Wyner and Zyv [54] for binary random variables, and has been recently extended by Jog and Anantharam to random variables in Abelian groups whose size is a power of two [36]. This result has recently been pointed out to the SCA community by Béguinot *et al.* at Cosade 2023 [15].

Theorem 4 (Mrs. Gerber's Lemma (MGL) [15, Cor. 1]). *Let $|\mathcal{Y}| = 2^n$ for some bit-size n and d be a positive integer. Let $Y_0, \ldots, Y_d$ be a $(d+1)$-encoding of the uniform random variable Y over $\mathcal{Y}$, and $\mathbf{L} = (L_0, \ldots, L_d)$ be such that, conditionally to Y_i, the variable L_i is independent of the others. Assume that for all $i \in [\![0, d]\!]$, $\mathsf{MI}(Y_i; L_i) \leq \delta_i$ for some parameter $0 \leq \delta_i \leq 1$. Then*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \mathrm{f}_{\mathsf{MI}}(\delta_0, \ldots, \delta_d) \underset{\delta \to 0}{=} \eta \prod_{i=0}^{d} \frac{\delta_i}{\eta} + \mathcal{O}\left(\prod_{i=0}^{d} \delta_i^2\right) \ , \tag{9}$$

where $\mathrm{f}_{\mathsf{MI}}(\cdot)$ is Mrs. Gerber's function, and $\eta = (2 \log 2)^{-1} \approx 0.72$.

We refer to the works of Béguinot *et al.* for more details about Mrs. Gerber's function [15]. In our context, we only need the properties summarized hereafter, and proven in section A.

Proposition 1 (Properties of the MGL function). *The Mrs. Gerber's Lemma (MGL) function $\mathrm{f}_{\mathsf{MI}}(\cdot)$ is non-decreasing with respect to each of its variables. Furthermore, for all $\delta_0, \ldots, \delta_d \in [0, 1]$, we have*

$$\mathrm{f}_{\mathsf{MI}}(\delta_0, \ldots, \delta_d) \leq \frac{1}{\log(2)} \prod_{i=0}^{d} 2\delta_i \ . \tag{10}$$

The upper bound of Equation 10 is looser than the approximation of Equation 9, yet it is non-asymptotic.

For Binary Gates. We now extend Béguinot *et al.*'s Theorem 4 to the case of binary gates, as stated hereafter by the following theorem that we prove in Appendix B.1, following a similar outline as Prest *et al.* [45, Thm. 6].

Theorem 5 (Binary Gates, Leaky Gates). *Let A, B be two independent and uniform random variables, over a finite field $\mathcal{Y}$. Let $(A_i)_{0\leq i\leq d}, (B_j)_{0\leq j\leq d}$ be d-encodings of A and B respectively. Let $L_{i,j} : A_i, B_j \mapsto L_{i,j}(A_i, B_j)$ be a family of randomized and mutually independent leakage functions such that for every i, j, $L_{i,j}$ is $\delta_{i,j}$-noisy (according to Definition 4). Denote the concatenation of the leakages $\{L_{i,j}\}_{0\leq i,j\leq d}$ by $\mathbf{L}$. Then, $\mathsf{MI}(A, B; \mathbf{L})$ is upper bounded by*

$$\mathrm{f}_{\mathsf{MI}}\left(|\mathcal{Y}| \cdot \sum_{j=0}^{d} \delta_{0,j}, \ldots, |\mathcal{Y}| \cdot \sum_{j=0}^{d} \delta_{d,j}\right) + \mathrm{f}_{\mathsf{MI}}\left(|\mathcal{Y}| \cdot \sum_{i=0}^{d} \delta_{i,0}, \ldots, |\mathcal{Y}| \cdot \sum_{i=0}^{d} \delta_{i,d}\right) \quad . \tag{11}$$

Theorem 6 (Binary Product, Leaky Wires). *Let A, B be two independent and uniform random variables, over a finite field $\mathcal{Y}$. Let $(A_i)_{0\leq i\leq d}, (B_j)_{0\leq j\leq d}$ be d-encodings of A and B respectively. Let $L_{i,j}$ be a family of randomized and mutually independent leakage functions verifying Definition 5, i.e., such that*

$$L_{i,j}(A_i, B_j) = \left(L'_{i,j}(A_i), L''_{i,j}(B_j), L'''_{i,j}(A_i \cdot B_j)\right) \quad ,$$

where $L'_{i,j}, L''_{i,j}, L'''_{i,j}$ are respectively $\delta'_{i,j}$-noisy, $\delta''_{i,j}$-noisy, and $\delta'''_{i,j}$-noisy. Denote the concatenation of the leakages $\{L_{i,j}\}_{0\leq i,j\leq d}$ by $\mathbf{L}$. Then, $\mathsf{MI}(A, B; \mathbf{L})$ is upper bounded by

$$\mathrm{f}_{\mathsf{MI}}\left(2 \cdot \sum_{j=0}^{d} \delta_{0,j}, \ldots, 2 \cdot \sum_{j=0}^{d} \delta_{d,j}\right) + \mathrm{f}_{\mathsf{MI}}\left(2 \cdot \sum_{i=0}^{d} \delta_{i,0}, \ldots, 2 \cdot \sum_{i=0}^{d} \delta_{i,d}\right) \quad . \tag{12}$$

3.5 Security Bound for each Type of Subsequence

In this section, we leverage the noise amplification result to bound the amount of leakage in each subsequence.

Type 1 subsequences occur for linear elementary calculations over uniform secrets, and are already covered by Theorem 4, which is a straightforward application of the MGL.

Corollary 4 (Type 1 subsequences). *Let Y be a uniform random variable over a finite field $\mathcal{Y}$ and $(Y_i)_{0\leq i\leq d}$ be a d-encoding of Y. Let $\delta \geq 0$ and $L_0, \ldots, L_d$ be δ-MI-noisy leakage functions over $\mathcal{Y}$. Denote $(L_0(Y_0), \ldots, L_d(Y_d))$ by $\mathbf{L}$. Then we have:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \frac{1}{\log(2)} \cdot (2 \cdot \delta)^{d+1} \quad . \tag{13}$$

Likewise, type 2 subsequences cover linear elementary calculations over non-uniform secrets, *e.g.*, occurring inside Sboxes. Such subsequences are covered by the following corollary.

Corollary 5 (Type 2 subsequences). *Let Y be a uniform random variable over a finite field $\mathcal{Y}$. Let k, d be positive integers and $(G_i)_{0\leq i\leq d}$ be a $(d+1)$-sharing of Y^k. Let $0 \leq \delta \leq 1$ and let $L_0(G_0), \ldots, L_d(G_d)$ be δ-MI-noisy leakages. Denote the concatenation of the leakages $\{L_i\}_{0\leq i\leq d}$ by $\mathbf{L}$. Then, we have:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \gcd(k, |\mathcal{Y}| - 1) \cdot \frac{1}{\log(2)} \cdot (2 \cdot \delta)^{d+1} \quad . \tag{14}$$

Proof. Straightforward, by combining Theorem 4 with Corollary 2. □

We now focus on the more involved type of subsequences, namely type 3, which is a binary gate. It occurs in the cross-products of the I.S.W. multiplication.

Corollary 6 (Type 3 subsequences). *Let Y be a uniform random variable over a finite field $\mathcal{Y}$, let d, p, q be positive integers. Let $(G_i)_i$, $(H_j)_j$ be $d+1$-additive sharings of Y^p, Y^q respectively. Let $0 \leq \delta$, and $\{G_i, H_j \mapsto L_{i,j}(G_i, H_j)\}_{i,j}$ be δ-MI-noisy leakage functions. Let us denote the concatenation of the leakages $\{L_{i,j}\}_{0\leq i,j\leq d}$ by $\mathbf{L}$, and denote $\varphi(p, q, M) = \min(\gcd(p, M-1), \gcd(q, M-1))$. Then we have:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq 2 \cdot |\mathcal{Y}| \cdot \varphi(p, q, |\mathcal{Y}|) \cdot \frac{1}{\log(2)} \cdot (2 \cdot |\mathcal{Y}| \cdot (d+1) \cdot \delta)^{d+1} \quad . \tag{15}$$

Moreover, if the $L_{i,j}$ are leakage on the wires, then

$$\mathsf{MI}(Y; \mathbf{L}) \leq 2 \cdot |\mathcal{Y}| \cdot \varphi(p, q, |\mathcal{Y}|) \cdot \frac{1}{\log(2)} \cdot (4 \cdot (d+1) \cdot \delta)^{d+1} \quad . \tag{16}$$

Proof. Using Corollary 3,

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}(\mathrm{Y}^p, \mathrm{Y}^q)) \leq |\mathcal{Y}| \cdot \varphi(p, q, |\mathcal{Y}|) \cdot \mathsf{MI}(\mathrm{A}, \mathrm{B}; \mathbf{L}(\mathrm{A}, \mathrm{B})) \quad ,$$

where A, B are uniform and independent random variables over $\mathcal{Y}$. Now invoking Theorem 5 to upper bound $\mathsf{MI}(\mathrm{A}, \mathrm{B}; \mathbf{L}(\mathrm{A}, \mathrm{B}))$ gives

$$\mathsf{MI}(\mathrm{A}, \mathrm{B}; \mathbf{L}(\mathrm{A}, \mathrm{B})) \leq 2\, \mathrm{f}_{\mathsf{MI}}(|\mathcal{Y}| \cdot (d+1) \cdot \delta, \ldots, |\mathcal{Y}| \cdot (d+1) \cdot \delta) \quad ,$$

where $\mathrm{f}(\cdot)$ denotes the MGL function. We conclude the proof by using Proposition 1, which gives Equation 15. For Equation 16, the same proof applies by using Theorem 6 instead of Theorem 5. □

It now remains to give some upper bounds for type 4 subsequences. These subsequences can be observed in the compression phase of I.S.W. multiplications (after cross-products and refreshings). This is the aim of the following result that we prove in Appendix B.1.

Theorem 7. *Let $Y_0, \ldots, Y_d$ be $d+1$ independent uniformly random variables over a finite set $\mathcal{Y}$. Let $L_1, \ldots, L_d$ be a family of δ_i-MI leakage functions, defined over $\mathcal{Y} \times \mathcal{Y}$, for some $0 \leq \delta_i \leq 1$. We have:*

$$\mathsf{MI}(Y_d; L_1(Y_0, Y_1), \ldots, L_d(Y_{d-1}, Y_d)) \leq \delta_d \ . \tag{17}$$

Corollary 7 (Type 4 subsequences). *Let Y be a secret, such that for $p, q \in \mathbb{N}$ the product of the multiplication $Y^p \times Y^q$ is processed by an I.S.W. gadget. For $0 \leq i, j \leq d$ and for $T_{i,j}, V_{i,j} \in \mathcal{Y}$, let $\mathbf{L} = \{L_{i,j}(T_{i,j-1}, V_{i,j})\}_{0 \leq i,j \leq d}$ denote the corresponding type 4 leakages such that for all i, j, the leakage $L_{i,j}(T_{i,j-1}, V_{i,j})$ is $\delta_{i,j}$-MI-noisy, for $\delta_{i,j} \leq \delta \leq 1$. Then the following inequality holds true:*

$$\mathsf{MI}(Y; \mathbf{L}_{i,j}(T_{i,j-1}, V_{i,j})_{0 \leq i,j \leq d}) \leq \gcd(p+q, M-1) \cdot \frac{1}{\log(2)} \cdot (2 \cdot \delta)^{d+1} \ . \tag{18}$$

Proof. Using Corollary 2, we reduce to the case where $\mathrm{Y}^p \times \mathrm{Y}^q$ is uniformly distributed over $\mathcal{Y}$, inducing a $\gcd(p+q, M-1)$ factor overhead. Then, by gathering the leakages $\mathbf{L}_{i,j}$ sharing the same index i by batches, we may notice that each batch of index only depends on one share of Y. We may therefore invoke Theorem 4 as follows:

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \leq \mathrm{f}_{\mathsf{MI}}(\delta'_0, \ldots, \delta'_d) \ , \tag{19}$$

where $\delta'_i = \mathsf{MI}\Big(\mathrm{Y}_i; \{\mathrm{L}_{i,j}(\mathrm{T}_{i,j-1}, \mathrm{V}_{i,j})\}_{0 \leq j \leq d}\Big)$. Finally, we can upper bound each δ'_i by $\delta_{i,d}$ using Theorem 7. □

3.6 From Subsequences to a Complete Computation.

We can now combine the three previous steps to state the main result, in a similar way as Prouff and Rivain [46, Thm. 4] and as Prest *et al.* [45, Sec. 6.3].

Theorem 8. *Consider a $\mathcal{Y}$-block cipher with monomial Sboxes, where a sequence of elementary calculations depends on a random variable Y uniformly distributed. Assume that these elementary calculations are protected by a d-encoding masking scheme as described in subsection 2.2, resulting in T elementary calculations giving access to the leakage $\mathbf{L} = (L_i)_{1 \leq i \leq T}$, where each leakage function L_i is assumed to be δ-MI-noisy. Then, the following inequality is verified:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq t_3 \cdot \frac{1}{\log(2)} \cdot (2 \cdot |\mathcal{Y}| \cdot (d+1) \cdot \delta)^{d+1} + t_{1,2,4} \cdot \frac{1}{\log(2)} \cdot (2 \cdot |\mathcal{Y}| \cdot \delta)^{d+1} \ ,$$

such that

$$t_3 = \sum_{(p,q) \in \mathcal{M}} \varphi(p, q, |\mathcal{Y}|) \,, \ \ t_{1,2,4} = \sum_{(p,q) \in \mathcal{M}} \phi(p, q, |\mathcal{Y}|) + \sum_{k \in \mathcal{S}} \psi(k, |\mathcal{Y}|) \ , \tag{20}$$

where $\mathcal{M}$ is the sequence of pairs (p, q) of exponents in the operands of the I.S.W. multiplication gadgets, $\mathcal{S}$ is the sequence of exponents (k) of operands over which a linear transformation is applied, and

– $\varphi(p,q,M) = 2 \cdot M \cdot \frac{M}{M-1} \cdot \min(\gcd(p, M-1), \gcd(q, M-1))$,
– $\phi(p,q,M) = \frac{M}{M-1} \cdot \gcd(p+q, M-1)$,
– $\psi(k,M) = \gcd(k, M-1)$.

Theorem 9. *In the same conditions as Theorem 8, assuming in addition that the type-3 subsequence occur leakages from their wires, then*

$$\mathsf{MI}(Y;\mathbf{L}) \leq t_3 \cdot \frac{1}{\log(2)} \cdot (4 \cdot (d+1) \cdot \delta)^{d+1} + t_{1,2,4} \cdot \frac{1}{\log(2)} \cdot (2 \cdot \delta)^{d+1} \quad ,$$

Proofs of Theorems 8 and 9. We apply Theorem 1 to decompose the MI into a sum of MIs for each subsequence. Since by assumption Y is uniformly distributed over $\mathcal{Y}$, Corollaries 4, 5, 6, 7 directly apply to bound each term in the sum. □

Note that in (20), $t_3 = \mathcal{O}\Big(|\mathcal{Y}|^{1+\epsilon} \cdot |\mathcal{M}|\Big)$, and $t_{1,2,4} = \mathcal{O}(|\mathcal{Y}|^{\epsilon} \cdot (|\mathcal{M}| + |\mathcal{S}|))$.

Corollary 8. *For any random-plaintext SCA key recovery adversary targeting a $\mathcal{Y}$-block cipher protected by the masking scheme described in subsection 2.2, the efficiency verifies the following bound:*

$$N_a^{\star}(\mathsf{SR}) \geq \frac{f(\mathsf{SR}, |\mathcal{Y}|)}{t_3 + t_{1,2,4}} \cdot \log(2) \cdot (2 \cdot |\mathcal{Y}| \cdot (d+1)\delta)^{-(d+1)} \quad ,$$

where $f(\mathsf{SR}, M) = \log_2(M) - (1-\mathsf{SR})\log_2(M-1) - \mathsf{H}_2(\mathsf{SR})$, where H_2 is the binary entropy function, and where the constants t_3 and $t_{1,2,4}$ are the ones defined in Theorem 8. Moreover, if the leakage is assumed to come from the wires, then

$$N_a^{\star}(\mathsf{SR}) \geq \frac{f(\mathsf{SR}, |\mathcal{Y}|)}{t_3 + t_{1,2,4}} \cdot \log(2) \cdot (4 \cdot (d+1)\delta)^{-(d+1)} \quad ,$$

Proof. Chérisey *et al.*'s security bound allows to link the SCA key recovery efficiency to the MI between $\mathrm{Y} = \mathrm{K} \oplus \mathrm{P}$ and the corresponding leakage:

$$N_a^{\star}(\beta) \geq \frac{f(\mathsf{SR}, |\mathcal{Y}|)}{\mathsf{MI}(\mathrm{Y};\mathbf{L})} \quad .$$

Plugging Theorem 8 or Theorem 9 — depending on the assumptions — into the latter inequality gives the result. □

4 Discussion

We have established our main results in section 3. We propose hereafter to discuss some features of our results, and to compare them to previous works. To this aim, we first argue in subsection 4.1 why the assumption "leakage from the wires vs. leakage from the gates" is very sensitive with respect to the resulting security bounds. Then, we compare in subsection 4.2 our bounds to previous works. Finally, we discuss in subsection 4.4 how we can extend our results to security bounds in terms of chosen plaintext attacks.

4.1 On the Tightness of the Proof

We have seen throughout section 3 that assuming the leakage to come from the wires leads to tighter security bounds. Hereupon, the large gap between the two settings is intriguing. Indeed, the polynomial field size factor from Equation 11 does not seem to come from any physical reason. Unfortunately, we argue hereafter that this gap is unavoidable. To this end, let us consider the leakage function $\mathrm{L} : \mathbb{F}^2 \to \mathbb{F} \cup \{\perp\}$ defined by:

$$\mathrm{L}(\mathrm{A}, \mathrm{B}) = \begin{cases} \mathrm{A}, & \text{if } \mathrm{B} = 0 \\ \perp, & \text{otherwise} \end{cases} . \tag{21}$$

Equation 21 is an example of leakage function that cannot be encompassed into the wire-leakage paradigm. One can verify — as we show in Appendix C — that this leakage model is δ-noisy for $\delta = \mathcal{O}\Big(\frac{\log(|\mathcal{Y}|)}{|\mathcal{Y}|}\Big)$. This means that, should an upper bound of the shape of Equation 12 hold for any type of leakage from the gates, the type 3 subsequences leaking according to Equation 21 are expected to remain leakage-resilient, provided that the field size if big enough, *i.e.*, $|\mathcal{Y}| \gg d$. However, there exists an attack with complexity and success rate polynomial in d and $|\mathcal{Y}|$ whenever the leakage function defined by Equation 21 is applied to a type-3 sequence. More precisely, let us assume an adversary is given the $\{\mathrm{L}(\mathrm{A}_i, \mathrm{B}_j)\}_{i,j}$, for $i, j \in [\![0, d]\!]$, and where L is the leakage function defined by Equation 21. Then it suffices that *at least* one share B_j is not equal to zero to recover *all* the shares of A, which occurs with probability $1 - \Big(1 - \frac{1}{|\mathcal{Y}|}\Big)^d \approx \frac{d}{|\mathcal{Y}|}$. The following proposition, proven in Appendix C .

Proposition 2. *Let $A = \sum_{i=1}^{d} A_i$ and $B = \sum_{i=1}^{d} B_i$, where the random variables A_i and B_j are uniformly distributed over a finite field $\mathbb{F}$ of size M. Suppose that the adversary observes $\mathbf{L} = \{L(A_i, B_j)\}_{i,j}$, for $i, j \in [\![1, d]\!]$, and where L is defined by Equation 21. Then,*

$$\mathsf{MI}(A, B; \mathbf{L}) = \frac{d \cdot \log_2(M)}{M} + \mathcal{O}\left(\frac{\log_2(M)}{M^2}\right) . \tag{22}$$

Hence, Theorem 5 is tight with respect to the field size, and tighter upper bounds require further assumptions — such as leakage from the wires.

4.2 Comparison with Related Works

We compare in this section our security bounds with related works. To this end, we first discuss the noise requirements in the different security bounds in the literature. We synthesize in Table 3 the different noise requirements of masking security bounds.

[6] As pointed out by Béguinot *et al.*, some results of [46,45] are flawed. Some patches are proposed in [16]. Moreover, as explained by Prest *et al.* [45, Appendix E], the Average Relative Error (ARE) is not formally defined for unbounded leakage models like Gaussian noise, unless requiring to a tail-cut argument that adds another constant factor hidden in the $\Omega(\cdot)$ notation.

Table 3: Noise requirements, and illustration on a case study on a Hamming weight leakage model with additive Gaussian noise.

Work (year)	Leak. Model	Leak-free refresh	Requirement	Equiv. Gaussian noise
[46] (2013)[6]	Gates	Yes	$\mathsf{EN} \leq \mathcal{O}\left(\frac{1}{d \cdot M^3}\right)$	
[25] (2014)	Gates	No	$\mathsf{SD} \leq \mathcal{O}\left(\frac{1}{d \cdot M^2}\right)$	
[45] (2019)	Gates	Yes	$\mathsf{RE} \leq \mathcal{O}\left(\frac{1}{d}\right)$	
This work	Gates	Yes	$\mathsf{MI} \leq \mathcal{O}\left(\frac{1}{d \cdot M}\right)$	
[25] (2014)	Wires	No	$\mathsf{SD} \leq \mathcal{O}\left(\frac{1}{d \cdot M}\right)$	$\sigma \geq \Omega\left(dM\sqrt{\log(M)}\right)$
[27] (2015)	Wires	Yes	$\mathsf{SD} \leq \mathcal{O}\left(\frac{1}{d}\right)$	$\sigma \geq \Omega\left(d\sqrt{\log(M)}\right)$
[45] (2019)	Wires	No	$\mathsf{ARE} \leq \mathcal{O}\left(\frac{1}{d}\right)$	$\sigma \geq \Omega(d\log(M))$
This work	Wires	Yes	$\mathsf{MI} \leq \mathcal{O}\left(\frac{1}{d}\right)$	$\sigma \geq \Omega\left(\sqrt{d\log(M)}\right)$

Leakage from the Wires. At first glance, when considering a leakage from the wires, our security bound gets a similar noise requirement as the proofs of Dziembowski *et al.* [27] and Prest *et al.* [45], although stated in different metrics. To clarify this comparison, we extend Prest *et al.*'s case study on the exemplary leakage distribution in which each intermediate calculation is assumed to leak its Hamming weight with an additive Gaussian noise of standard deviation σ [45, Table 1]. We complete Table 3 with our new result, by using the fact that for such a leakage model, $\mathsf{MI} = \Theta\left(\frac{\log(M)}{\sigma^2}\right)$ [7]. It can be noticed that on this particular leakage distribution, our requirement on the minimal noise level is now the weakest of all security proofs based on the I.S.W. masking scheme, with the only drawback of requiring leak-free refreshings.

Notice that the dependency of the noise requirement in d is tight, since it depicts the potential ability of an adversary to increase its success of recovering each share through *horizontal attacks*, as argued by Battistello *et al.* [7] and Grosso and Standaert [33]. Nevertheless, it is still possible to relax this dependency by using other multiplication gadgets [1,3,2,8,30,31].

Leakage from the Gates. When considering the more general paradigm of leakage from the gates, Table 3 suggests that Prest *et al.*'s RE-based security bound remains the best one on the noise requirements. However, we emphasize that the RE is a *worst-case* metric, whereas all the other metrics in Table 3 are *averaged* metrics. Estimating worst-case metrics may not always be efficiently tractable by practitioners, especially for high-dimensional leakage. In addition, worst-case metrics are by definition more conservative than averaged metrics. This is usual in theoretical cryptography, but it contrasts with the concrete SCA security metrics like the GE or the SR [51] that are also averaged metrics.

To illustrate this, let us consider another example of leakage distribution, namely the now famous *random probing* model considered by Duc *et al.* in their groundbreaking work [25]. In this leakage model, the adversary can recover the

exact value of any intermediate calculation, each with probability $0 \leq \kappa \leq 1$, where the parameter κ denotes the baseline noise level here. It can be verified that the MI of this leakage model is $\log |\mathcal{Y}| \cdot \kappa$, whereas its RE is always fixed to $|\mathcal{Y}|-1$ regardless of the value of κ, so that the MI can be set arbitrarily close to zero — by setting κ accordingly — while the RE remains constant. In other words, the random probing model can *never* be proven secure with masking by using a security bound involving the RE, whereas our masking security bound remains meaningful for the random probing leakage model, as long as $\kappa \leq \mathcal{O}\left(\frac{1}{\log|\mathcal{Y}| \cdot d}\right)$.[7]

As a result, the only security bound comparable with ours in terms of noise requirements remains Dziembowski *et al.*'s bound [27]. Their bound is obtained using Azuma's concentration inequalities [12]. Although a similar approach using Chernoff's concentration inequality has been fruitful in Duc *et al.*'s elegant reduction to the probing model [25] due to its genericity, it has a major drawback since the convergence rate of the security bound no longer depends on the actual baseline noise level δ. This is highlighted in their final bound [27, Eq. (42)]: it can be verified that the bound is even always increasing for values of d between 0 and 8, and becomes non-trivial — *i.e.*, lower than one — only for $d \geq 142$ if $|\mathcal{Y}| = 256$. On the opposite, our security bounds do not suffer from this caveat, since they depend on the actual baseline noise level δ, which makes our bounds non-trivial for arbitrarily small value of d, provided that δ is small enough.

4.3 Beyond Monomial Sboxes

So far, our proof has focused on the particular case of monomial Sboxes, covering the AES Sbox. We may therefore wonder to which extent the latter assumption is sensitive to derive our proof. Thankfully, relaxing the monomial assumption is still possible, at the cost of an additional constant factor scaling with the field size, by simply using the generic reduction to non-uniform secrets stated by Theorem 2 instead of refining its refined variant.

Corollary 9. *Let Y be a random variable* arbitrarily *distributed over $\mathcal{Y}$, and protected by a masking scheme with $d+1$ shares as described in subsection 2.2, resulting in T elementary calculations. Assume that the scheme protects $|\mathcal{S}|$ linear operations, and $|\mathcal{M}|$ I.S.W. multiplications. Let $\mathbf{L} = (L_i)_{1 \leq i \leq T}$ be the random vector denoting the leakage of the full masking scheme, and let $\delta \geq 0$ be such that every L_i is δ-MI-noisy. Then, the inequalities of Theorem 8 and Theorem 9 are verified for:*

$$t_3 = 2 \cdot \left(1 - e^{-1}\right) \cdot |\mathcal{Y}|^2 \cdot |\mathcal{M}|\,, \;\; t_{1,2,4} = \quad \left(1 - e^{-1}\right) \cdot |\mathcal{Y}| \cdot (|\mathcal{S}| + |\mathcal{M}|)\,.$$

Proof. We apply Theorem 1, then we group the type 1, 2, and 4 subsequences together and we apply the reduction to uniform secrets using Theorem 2. Likewise, we apply Theorem 2 for type 3 subsequences over the domain $\mathcal{Y} \times \mathcal{Y}$. We can then directly apply Theorem 4 and Theorem 5 respectively. □

[7] This condition could even be relaxed to $\kappa \leq \mathcal{O}\left(\frac{1}{d}\right)$ in the particular case of leakage in the random probing model, if one would directly state a security bound for this leakage model, *e.g.*, by extending Eq. (9) of Duc *et al.* [26].

4.4 Beyond Random Plaintext Attacks

Likewise, one may argue that the latter comparison with the works of Dziembowski *et al.* is not completely fair, since their bound is stated for SCA adversary with chosen plaintext. Hereupon, the authors stated later at Tcc 2016 that by leveraging a reduction from non-uniform secrets to uniform secrets [28, Lemma 2],

> *"The cryptographic interpretation of [reductions from non-uniform to uniform secrets] is that it suffices to consider only random-plaintext attacks, instead of chosen-plaintext attacks"* [28, p. 297].

We notice that our Theorem 2 actually represents such a reduction. Accordingly, Theorems 8 and 9 can be extended to cover adversaries with chosen plaintexts, by multiplying the constant factors by $(1 - e^{-1}) \cdot |\mathcal{Y}|$, instead of using the power map trick. In other words, Corollary 9 also holds for chosen-plaintexts attacks.

Table 4 synthesizes the different constant factors t_3, whether the SCA adversary is assumed to operate with random or chosen plaintexts. We may notice

Table 4: Constant factor overhead, depending on the attack scenario, and on the multiplication gadget used.

Sbox \ Plaintext	Random	Chosen
Any Sbox	$\mathcal{O}(\lvert\mathcal{Y}\rvert^2)$	$\mathcal{O}(\lvert\mathcal{Y}\rvert^2)$
Monomial Sbox	$\mathcal{O}(\lvert\mathcal{Y}\rvert^{1+\epsilon})$	$\mathcal{O}(\lvert\mathcal{Y}\rvert^2)$

that the constant factor of Corollary 9 scaling quadratically with the field size seems at first glance worse than the one of Dziembowski *et al.* [27, Thm. 1], whereas their security bound only scales linearly with the field size $|\mathcal{Y}|$. Nevertheless, their bound is stated in terms of SD, whereas ours is stated in terms of MI, which does not behave the same as recalled in Table 3.

4.5 Perspectives

The main limitation of our work remains the leak-free assumption for the mask refreshings, like in the previous works [46,27,45]. It remains an open problem whether this assumption could be relaxed. Likewise, our masking security proof only covers the I.S.W. masking scheme, as in the previous works, whereas the generic approach through the probing model can cover any type of masking scheme. Nevertheless, we do not see any prior reason why our security proof could not be used to extend over different masking gadgets, beyond the I.S.W. multiplication gadget, and in particular for table-based masking schemes [20,22], that are known to be efficiently secure in the probing model, but much less in the noisy leakage [52,14]. Overall, this leaves the door open for good opportunities of improvement in the next few years.

Acknowledgments. François-Xavier Standaert is a Senior Associate Researcher of the Belgian Fund for Scientific Research (FNRS-F.R.S.). This work has been funded in part by the ERC project number 724725 (acronym SWORD).

A Utilitary Lemma

Proof of Proposition 1. By definition of the MGL function, we have

$$\mathrm{f}_{\mathsf{MI}}(\delta_0, \ldots, \delta_d) = 1 - \mathsf{H}\left(\frac{1}{2} - 2^d \cdot \prod_{i=0}^{d} \left(\frac{1}{2} - \mathsf{H}^{-1}(1-\delta_i)\right)\right) \ ,$$

where H stands here for the binary entropy function. Therefore, to upper bound the MGL function, we need to lower bound the binary entropy function, and to lower bound its inverse. The former one it straightforward to derive. Using the inequality $\log(1+x) \leq x$, it comes from the definition of the binary entropy that $\mathsf{H}\big(\frac{1}{2} - \epsilon\big) \geq 1 - \frac{4\epsilon^2}{\log(2)}$ for any $0 \leq \epsilon \leq \frac{1}{2}$.

We now focus on the upper bound. The binary entropy function is bijective from $[0, 1/2]$ to $[0, 1]$, and upper bounded by $2\sqrt{p(1-p)}$ — which is itself bijective from $[0, 1/2]$ to $[0, 1]$. Since it is also non-decreasing over $[0, 1/2]$, so is its inverse over $[0, 1]$, which means that

$$\mathsf{H}(p) \leq 2\sqrt{p(1-p)} \implies p \leq \mathsf{H}^{-1}\Big(2\sqrt{p(1-p)}\Big) \ .$$

Let us make the following change of variable $p = \frac{1}{2} - \epsilon$ and let $1 - \delta = 2\sqrt{p(1-p)}$. This implies that $\epsilon = \sqrt{\frac{1-(1-x)^2}{4}}$. In other words,

$$\mathsf{H}^{-1}(1-x) \geq \frac{1}{2} - \sqrt{\frac{1-(1-x)^2}{4}} \ .$$

Thus,

$$\begin{aligned}
&\frac{1}{2} - \mathsf{H}^{-1}(1-x) \leq \sqrt{\frac{1-(1-x)^2}{4}} \\
\implies &\prod_{i=0}^{d} \left(\frac{1}{2} - \mathsf{H}^{-1}(1-\delta_i)\right) \leq \sqrt{\prod_{i=0}^{d} \left(\frac{1-(1-\delta_i)^2}{4}\right)} \\
\implies &\frac{1}{2} - 2^d \prod_{i=0}^{d} \left(\frac{1}{2} - \mathsf{H}^{-1}(1-\delta_i)\right) \geq \frac{1}{2} - 2^d \sqrt{\prod_{i=0}^{d} \left(\frac{1-(1-\delta_i)^2}{4}\right)} \\
\implies &\mathsf{H}\left(\frac{1}{2} - 2^d \prod_{i=0}^{d} \left(\frac{1}{2} - \mathsf{H}^{-1}(1-\delta_i)\right)\right) \geq 1 - \frac{4}{\log(2)} \cdot \left(2^d \sqrt{\prod_{i=0}^{d} \left(\frac{1-(1-\delta_i)^2}{4}\right)}\right)^2 \\
&= 1 - \frac{1}{\log(2)} \prod_{i=0}^{d} (2\delta_i - \delta_i^2) \ .
\end{aligned}$$

Hence,

$$\mathrm{f}_{\mathsf{MI}}(\delta_0, \ldots, \delta_d) \leq \frac{1}{\log(2)} \prod_{i=0}^{d} (2\delta_i - \delta_i^2) \leq \frac{1}{\log(2)} \prod_{i=0}^{d} 2\delta_i$$

□

Lemma 2. *Let $Y \in \mathcal{Y}$ be a discrete random variable, and let $g : Y \mapsto g(Y)$ be a mapping $\mathcal{Y} \to \mathcal{Y}$. Let $\mathbf{L} : \mathcal{Y} \to \mathcal{L}$ be a noisy leakage function. Then:*

$$\mathsf{MI}(Y; \mathbf{L}(g(Y))) = \mathsf{MI}(g(Y); \mathbf{L}(g(Y))) \ .$$

Proof of Lemma 2. First, notice that we have the two following Markov chains:

$$\mathrm{Y} \to g(\mathrm{Y}) \to \mathbf{L}(g(\mathrm{Y})) \ ,$$
$$g(\mathrm{Y}) \leftarrow \mathrm{Y} \to \mathbf{L}(g(\mathrm{Y})) \ .$$

By the DPI [23, Sec. 2.8] on the first two chains, we have $\mathsf{MI}(\mathrm{Y}; \mathbf{L}(g(\mathrm{Y}))) \lesseqgtr \mathsf{MI}(g(\mathrm{Y}); \mathbf{L}(g(\mathrm{Y})))$, hence:

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}(g(\mathrm{Y}))) = \mathsf{MI}(g(\mathrm{Y}); \mathbf{L}(g(\mathrm{Y}))) \ .$$

□

Lemma 3 (Leakage from binary gates). *Let A, B be two independent uniform random variables over a finite set of size M, and let $L : \mathbb{F}^2 \to \mathcal{L}$ be a δ-noisy function, i.e. $\mathsf{MI}(A, B; L(A, B)) \leq \delta$. Then, for any realization b of B, we have*

$$\mathsf{MI}(A; L(A, b)) \leq M \cdot \delta \ . \tag{23}$$

Proof. Using the chain rule of MI [23, Thm. 2.5.2] the other way around, we get the following inequality:

$$\mathop{\mathbb{E}}_{b} [\mathsf{MI}(\mathrm{A}; \mathrm{L}(\mathrm{A}, b))] = \mathsf{MI}(\mathrm{A}; \mathrm{L}(\mathrm{A}, \mathrm{B}) \mid \mathrm{B}) \leq \mathsf{MI}(\mathrm{A}, \mathrm{B}; \mathrm{L}(\mathrm{A}, \mathrm{B})) = \delta \ .$$

Since the expectation is taken over a finite set of size M, we have that

$$\max_{b} \mathsf{MI}(\mathrm{A}; \mathrm{L}(\mathrm{A}, b)) \leq M \cdot \mathop{\mathbb{E}}_{b} [\mathsf{MI}(\mathrm{A}; \mathrm{L}(\mathrm{A}, b))] \leq M \cdot \delta \ ,$$

hence the result. □

Lemma 4 (Leakage from wires). *Let A be a uniform random variable over a finite set of size M, and let $L : \mathbb{F} \to \mathcal{L}$ be a δ-noisy function, i.e. $\mathsf{MI}(A; L(A)) \leq \delta$. Then, for any $b \in \mathbb{F}$, we have*

$$\mathsf{MI}(A; L(A \cdot b)) \leq \delta \ . \tag{24}$$

Proof. The inequality is trivial for $b = 0$, so we assume hereafter that $b \neq 0$. Observe first that for any random variable A (non-necessarily uniform), since $b \neq 0$, the PMF $\Pr(\mathrm{A} \cdot b)$ is equal to $\Pr(\mathrm{A})$ up to a fixed permutation characterized by b. This means in particular that if A is uniformly distributed, $\mathrm{A}' = \mathrm{A} \cdot b$ and A are equally distributed as a uniform random variable over $\mathbb{F}$. Likewise, $\mathrm{A}' \overset{d}{=} \mathrm{A}$ implies that $\mathrm{L}(\mathrm{A}') \overset{d}{=} \mathrm{L}(\mathrm{A})$, as for any $l \in \mathcal{L}$,

$$\begin{aligned}
\Pr(\mathrm{L}(\mathrm{A}) = l) &= \sum_a \Pr(\mathrm{L}(\mathrm{A}) = l \ \mid \ \mathrm{A} = a) \cdot \Pr(\mathrm{A} = a) \\
&= \frac{1}{M} \sum_a \Pr(\mathrm{L}(a) = l) \\
&= \frac{1}{M} \sum_{a' = a \cdot b} \Pr(\mathrm{L}(a') = l) \\
&= \sum_{a'} \Pr\big(\mathrm{L}(\mathrm{A}') = l \ \mid \ \mathrm{A}' = a'\big) \cdot \Pr\big(\mathrm{A}' = a'\big) \\
&= \Pr\big(\mathrm{L}(\mathrm{A}') = l\big) \ .
\end{aligned}$$

Furthermore, for every $a \in \mathbb{F}, l \in \mathcal{L}$,

$$\begin{aligned}
\Pr(\mathrm{A} = a, \mathrm{L}(\mathrm{A} \cdot b) = l) &= \Pr(\mathrm{A} = a) \cdot \Pr(\mathrm{L}(\mathrm{A} \cdot b) = l \ \mid \ \mathrm{A} = a) \\
&= \Pr(\mathrm{A} \cdot b = a \cdot b) \cdot \Pr(\mathrm{L}(\mathrm{A} \cdot b) = l \ \mid \ \mathrm{A} \cdot b = a \cdot b) \\
&= \Pr\big(\mathrm{A}' = a \cdot b\big) \cdot \Pr\big(\mathrm{L}(\mathrm{A}') = l \ \mid \ \mathrm{A}' = a \cdot b\big) \\
&= \Pr\big(\mathrm{A}' = a \cdot b, \mathrm{L}(\mathrm{A}') = l\big) \ ,
\end{aligned}$$

where the first and the last equalities come from the definition of conditional probabilities, the second equality comes from the fact that $\mathrm{A} = a$ if and only if $\mathrm{A} \cdot b = a \cdot b$ since $b \in \mathbb{F}^\star$, the third equality comes from the definition of A'. As a result of those facts, denoting $\mathsf{MI}(\mathrm{A}; \mathrm{L}(\mathrm{A} \cdot b))$ by δ_b, we have

$$\begin{aligned}
\delta_b &= \mathsf{D}_{\mathsf{KL}}(\Pr(\mathrm{A}, \mathrm{L}(\mathrm{A} \cdot b)) \ \| \ \Pr(\mathrm{A}) \otimes \Pr(\mathrm{L}(\mathrm{A} \cdot b))) \\
&= \sum_{\mathrm{a} \in \mathbb{F}} \sum_{l \in \mathcal{L}} \Pr(\mathrm{A} = a, \mathrm{L}(\mathrm{A} \cdot b) = l) \cdot \log\left(\frac{\Pr(\mathrm{A} = a, \mathrm{L}(\mathrm{A} \cdot b) = l)}{\Pr(\mathrm{A} = a) \cdot \Pr(\mathrm{L}(\mathrm{A} \cdot b) = l)}\right) \\
&= \sum_{\mathrm{a} \in \mathbb{F}} \sum_{l \in \mathcal{L}} \Pr\big(\mathrm{A}' = a \cdot b, \mathrm{L}(\mathrm{A}') = l\big) \cdot \log\left(\frac{\Pr\big(\mathrm{A}' = a \cdot b, \mathrm{L}(\mathrm{A}') = l\big)}{\Pr(\mathrm{A} = a) \cdot \Pr\big(\mathrm{L}(\mathrm{A}') = l\big)}\right) \\
&= \sum_{a' = a \cdot b \in \mathbb{F}} \sum_{l \in \mathcal{L}} \Pr\big(\mathrm{A}' = a', \mathrm{L}(\mathrm{A}') = l\big) \cdot \log\left(\frac{\Pr\big(\mathrm{A}' = a', \mathrm{L}(\mathrm{A}') = a'\big)}{\Pr\big(\mathrm{A}' = a'\big) \cdot \Pr\big(\mathrm{L}(\mathrm{A}') = l\big)}\right) \\
&= \mathsf{MI}\big(\mathrm{A}'; \mathrm{L}(\mathrm{A}')\big) = \delta \ ,
\end{aligned}$$

where the first and second equalities are by definition of the MI and the Kullback - Leibler (KL) divergence. The third equality is obtained by replacing the probabilities thanks to the previous computations, and the fourth equality is obtained

by the change of variable $a' = a \cdot b$. Finally, the last equality comes from the fact that $\mathrm{A}' \overset{d}{=} \mathrm{A}$. □

Corollary 10. *Assume* $\mathbf{L}$ *is a leakage from wires on a binary gate, i.e., there exist three* δ*-noisy functions* L', L'' *and* L''' *such that*

$$\mathbf{L}(A, B) = \big(L(A), L''(B), L'''(A \cdot B)\big) \ ,$$

and such that conditionally to A *and* B, L_{i_1}, L_{i_2} *and* L_o *are independent. Then, for any* $b \in \mathbb{F}^\star$*, we have:*

$$\mathsf{MI}(A; L(A, b)) \leq 2 \cdot \delta \ . \tag{25}$$

Moreover, for $b = 0$*, we have* $\mathsf{MI}(A; L(A, b)) \leq \delta$.

Proof. According to the chaining rule for MIs [23, Thm. 2.5.2], we have

$$\begin{aligned}\mathsf{MI}(\mathrm{A}; \mathbf{L}(\mathrm{A}, b)) = \mathsf{MI}\big(\mathrm{A}; \mathrm{L}'(\mathrm{A})\big) &+ \mathsf{MI}\big(\mathrm{A}; \mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{L}'(\mathrm{A})\big) \\ &+ \mathsf{MI}\big(\mathrm{A}; \mathrm{L}''(b) \ \big| \ \mathrm{L}'(\mathrm{A}), \mathrm{L}'''(\mathrm{A} \cdot b)\big)\end{aligned}$$

By assumption, the first term of the right-hand side above can be upper bounded by δ. Moreover, since $\mathrm{L}''(b)$ is independent of A, the last term of the right-hand side above is equal to zero. Finally observe for the second term that

$$\begin{aligned}\mathsf{MI}\big(\mathrm{A}; \mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{L}'(\mathrm{A})\big) &= \mathsf{H}\big(\mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{L}'(\mathrm{A})\big) - \mathsf{H}\big(\mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{A}, \mathrm{L}'(\mathrm{A})\big) \\ &= \mathsf{H}\big(\mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{L}'(\mathrm{A})\big) - \mathsf{H}\big(\mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{A}\big) \\ &\leq \mathsf{H}\big(\mathrm{L}'''(\mathrm{A} \cdot b)\big) - \mathsf{H}\big(\mathrm{L}'''(\mathrm{A} \cdot b) \ \big| \ \mathrm{A}\big) \\ &= \mathsf{MI}\big(\mathrm{A}; \mathrm{L}'''(\mathrm{A} \cdot b)\big) \\ &\leq \delta \ ,\end{aligned}$$

where the first equality holds by definition of the conditional mutual information, the second equality holds as conditionally to A, $\mathrm{L}'(\mathrm{A})$ is independent of $\mathrm{L}'''(\mathrm{A} \cdot b)$, the first inequality holds as conditioning reduces the entropy, the last equality holds by definition of the mutual information, and the last inequality holds by virtue of Lemma 4. □

B Proofs of Main Results

Proof of Theorem 1. By definition, we have

$$\mathsf{H}(\mathbf{L} \ \big| \ \mathrm{Y}) = \mathop{\mathbb{E}}_{y} \left[\mathsf{H}(\mathrm{L}_1, \ldots, \mathrm{L}_t \ \big| \ \mathrm{Y} = y)\right] \ . \tag{26}$$

By assumption, all the leakages, conditioned to $\mathrm{Y} = y$ are mutually independent so

$$\mathsf{H}(\mathbf{L} \ \big| \ \mathrm{Y} = y) = \sum_{i=1}^{t} \mathsf{H}(\mathrm{L}_i \ \big| \ \mathrm{Y} = y) \ .$$

Hence, combining with Equation 26, $\mathsf{H}(\mathbf{L} \mid \mathrm{Y}) = \sum_{i=1}^{t} \mathsf{H}(\mathrm{L}_i \mid \mathrm{Y})$. Thereby,

$$\mathsf{MI}(\mathbf{L}; \mathrm{Y}) \leq \sum_{i=1}^{t} \mathsf{MI}(\mathrm{L}_i; \mathrm{Y})$$

□

Proof of Theorem 2. Now, we can see $\mathbf{L}$ as an — undesired — communication channel. By definition of the capacity C of the channel $\mathbf{L}$, and using Lemma 2, we get that

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}(g(\mathrm{Y}))) = \mathsf{MI}(g(\mathrm{Y}); \mathbf{L}(g(\mathrm{Y}))) \leq \max_{\Pr(\mathrm{Z})} \mathsf{MI}(\mathrm{Z}; \mathbf{L}(\mathrm{Z})) = C \ .$$

Using [50, Thm. 1, Eq. (17)], we get that

$$\frac{C}{\mathsf{MI}(\mathrm{Y}; \mathbf{L}(\mathrm{Y}))} \leq |\mathcal{Y}| \cdot \min\left\{2^{-C}, 1 - e^{-1}\right\} \leq |\mathcal{Y}| \cdot \left(1 - e^{-1}\right)$$

□

B.1 Proofs of Theorem 5 and Theorem 6

Proof. Using the chain rule of MI [23, Thm. 2.5.2], we have:

$$\mathsf{MI}((\mathrm{A}, \mathrm{B}); \mathbf{L}) = \mathsf{MI}(\mathrm{A}; \mathbf{L}) + \mathsf{MI}(\mathrm{B}; \mathbf{L} \mid \mathrm{A}) \ . \tag{27}$$

Let us bound the first term of Equation 27. The bound on the second term will straightforwardly follow.

Bounding $\mathsf{MI}(A; \mathbf{L})$. Observe that since A and B are independent, it follows that A and $\mathbf{B}$ are also independent. As a result, $\mathsf{H}(\mathrm{A} \mid \mathbf{B}) = \mathsf{H}(\mathrm{A})$. Furthermore, since conditioning decreases the entropy, we have $\mathsf{H}(\mathrm{A} \mid \mathbf{L}, \mathbf{B}) \leq \mathsf{H}(\mathrm{A} \mid \mathbf{L})$. The two latter facts imply that

$$\mathsf{MI}(\mathrm{A}; \mathbf{L}) \leq \mathsf{MI}(\mathrm{A}; \mathbf{L} \mid \mathbf{B}) = \mathop{\mathbb{E}}_{\mathbf{b}} [\mathsf{MI}(\mathrm{A}; \mathbf{L} \mid \mathbf{B} = \mathbf{b})] \ . \tag{28}$$

Let $\mathbf{b} = (b_0, \ldots, b_d)$ be fixed for now, and let us bound $\mathsf{MI}(\mathrm{A}; \mathbf{L} \mid \mathbf{B} = \mathbf{b})$. To this end, notice that we may now gather the leakages $\mathbf{L}_{i,j}$ by batches sharing the same index i as follows:

$$\mathsf{MI}(\mathrm{A}; \mathbf{L} \mid \mathbf{B} = \mathbf{b}) = \mathsf{MI}\Big(\mathrm{A}; \{\mathbf{L}_{0,j}(\mathrm{A}_0, b_j)\}_{0 \leq j \leq d}, \ldots, \{\mathbf{L}_{d,j}(\mathrm{A}_d, b_j)\}_{0 \leq j \leq d}\Big) \ . \tag{29}$$

By assumption, each batch of leakages $\{\mathbf{L}_{i,j}(\mathrm{A}_i, b_j)\}_{0 \leq j \leq d}$ only depends on the share A_i. Hence, we may use Theorem 4 to bound the right hand-side of Equation 29 as follows. Let us define $\mathsf{MI}\Big(\mathrm{A}_i; \{\mathbf{L}_{i,j}(\mathrm{A}_i, b_j)\}_{0 \leq j \leq d}\Big) = \delta'_i$ — notice that δ'_i depends on all the b_j. Then we have

$$\mathsf{MI}\Big(\mathrm{A}; \{\mathbf{L}_{0,j}\}_{0 \leq j \leq d}, \ldots, \{\mathbf{L}_{d,j}\}_{0 \leq j \leq d}\Big) \underset{(9)}{\leq} \mathrm{f}_{2^n}(\delta'_0, \ldots, \delta'_d) \ . \tag{30}$$

Substituting Equation 30 in Equation 29, and then plugging into Equation 28 gives

$$\mathsf{MI}(\mathrm{A};\mathbf{L}) \leq \mathop{\mathbb{E}}_{\mathbf{b}}\left[\mathrm{f}_{2^n}(\delta'_0,\ldots,\delta'_d)\right] \ . \tag{31}$$

We are then reduced to upper bound Equation 31. To this end, notice that for i fixed, the batch of leakages $\{\mathbf{L}_{i,j}(\mathrm{A}_i,b_j) \mid \mathrm{A}_i\}_{0\leq j\leq d}$ are mutually independent. Hence, we can now leverage Theorem 1 to upper bound δ'_i, as follows:

$$\delta'_i = \mathsf{MI}\Big(\mathrm{A}_i;\{\mathbf{L}_{i,j}(\mathrm{A}_i,b_j)\}_{0\leq j\leq d}\Big) \underset{(2)}{\leq} \sum_{j=0}^{d}\mathsf{MI}(\mathrm{A}_i;\mathbf{L}_{i,j}(\mathrm{A}_i,b_j)) \ . \tag{32}$$

Depending on the assumptions on the leakage, *i.e.*, whether the leakage is assumed to come from the gates or from the wires, we get the following upper bounds:

$$\mathsf{MI}(\mathrm{A}_i;\mathbf{L}_{i,j}(\mathrm{A}_i,b_j)) \leq \begin{cases} M\cdot\delta_{i,j} & , \text{ leakage from the gates (Lemma 3);} \\ 2\cdot\delta_{i,j} & , \text{ leakage from the wires (Corollary 10) .} \end{cases} \tag{33}$$

Hence, combining Equation 32 with Equation 33, we get that

$$\delta'_i \leq \begin{cases} M\cdot\sum_{j=0}^{d}\delta_{i,j} & , \text{ leakage from the gates ;} \\ 2\cdot\sum_{j=0}^{d}\delta_{i,j} & , \text{ leakage from the wires .} \end{cases} \tag{34}$$

Finally, plugging Equation 34 into Equation 31 gives the first term in the right hand-side of Equation 11.

Bounding $\mathsf{MI}(B;\mathbf{L} \mid A)$. Using the chain rule of the MI again, we may bound $\mathsf{MI}(\mathrm{B};\mathbf{L} \mid \mathrm{A})$ by conditioning on the d last shares of $\mathbf{A}$ (except the share of index 0):

$$\mathsf{MI}(\mathrm{B};\mathbf{L} \mid \mathrm{A}) \leq \mathsf{MI}\Big(\mathrm{B};\mathbf{L} \mid \mathrm{A},\{\mathrm{A}_i\}_{1\leq i\leq d}\Big)$$

Using the same argument as Dziembowski *et al.* [28, Lemma 3], we may notice that since A is assumed to be uniform:

$$\Big(\mathrm{A},\{\mathrm{A}_i\}_{1\leq i\leq d}\Big) \overset{d}{=} \left(\mathrm{A}\oplus\left(\bigoplus_{i=1}^{d}\mathrm{A}_i\right),\{\mathrm{A}_i\}_{1\leq i\leq d}\right) \overset{d}{=} \{\mathrm{A}_i\}_{0\leq i\leq d} \ ,$$

it implies that $\mathsf{MI}\Big(\mathrm{B};\mathbf{L} \mid \mathrm{A},\{\mathrm{A}_i\}_{1\leq i\leq d}\Big) = \mathsf{MI}(\mathrm{B};\mathbf{L} \mid \mathbf{A})$. By symmetry of the roles, the latter term can be bound in the same way as the right hand-side of Equation 28, by permuting the roles of the indices i and j. □

B.2 Proof of Theorem 7

Proof of Theorem 7. Let $\mathbf{L} = (\mathrm{L}_1(\mathrm{Y}_0,\mathrm{Y}_1),\ldots,\mathrm{L}_d(\mathrm{Y}_{d-1},\mathrm{Y}_d))$ for short. Expanding $\mathsf{MI}(\mathrm{Y}_d;\mathbf{L})$, we have

$$\begin{aligned}\mathsf{MI}(\mathrm{Y}_d;\mathbf{L}) = \ & \mathsf{MI}(\mathrm{Y}_d;\mathrm{L}_d(\mathrm{Y}_{d-1},\mathrm{Y}_d) \mid \mathrm{L}_{d-1}(\mathrm{Y}_{d-2},\mathrm{Y}_{d-1}),\ldots,\mathrm{L}_1(\mathrm{Y}_0,\mathrm{Y}_1)) \\ & + \mathsf{MI}(\mathrm{Y}_d;\mathrm{L}_{d-1}(\mathrm{Y}_{d-2},\mathrm{Y}_{d-1}),\ldots,\mathrm{L}_1(\mathrm{Y}_0,\mathrm{Y}_1))\end{aligned}$$

Notice first that the second term in the right hand-side equals 0, since by assumption Y_d is independent of the $\{\mathrm{Y}_i\}_{0\leq i\leq d-1}$. Likewise, the first term of the right hand-side can be upper bounded by $\overline{\mathsf{MI}}(\mathrm{Y}_d; \mathrm{L}_d(\mathrm{Y}_{d-1}, \mathrm{Y}_d) \mid \mathrm{Y}_{d-1})$, which can in turn be upper bounded by δ_d. □

C Proofs and Technical Lemma from the Example

Proposition 3. *Let L be the leakage function defined by Equation 21. Then*

$$\frac{\max_{b\in\mathbb{F}} \mathsf{MI}(A; L(A,b))}{\mathsf{MI}(A,B; L(A,B))} = \Omega(M) \quad . \tag{35}$$

Proof of Proposition 3. The leakage function is deterministic, so the MI coincides with the entropy. On the one hand, we have

$$\mathsf{MI}(\mathrm{A}; \mathrm{L}(\mathrm{A}, b)) = \begin{cases} \log_2(M), & \text{if } b = 0 \\ 0, & \text{otherwise} \end{cases} \tag{36}$$

On the other hand, let us compute $\mathsf{MI}(\mathrm{A}, \mathrm{B}; \mathrm{L}(\mathrm{A}, \mathrm{B}))$. The leakage function takes $M+1$ values, namely $\perp$ and all the values $a \in \mathbb{F}$. More precisely,

$$\begin{aligned} \Pr(\mathrm{L} = a) &= \Pr(\mathrm{A} = a, \mathrm{B} = 0) = \frac{1}{M^2} \quad , \\ \Pr(\mathrm{L} = \perp) &= \Pr(\mathrm{B} \neq 0) = \frac{M-1}{M} \quad . \end{aligned}$$

As a result,

$$\begin{aligned} \mathsf{H}(\mathrm{L}(\mathrm{A}, \mathrm{B})) &= M \cdot \frac{1}{M^2} \cdot 2\log_2(M) + \frac{M-1}{M} \cdot \log_2\left(\frac{M}{M-1}\right) \\ &= 2\frac{\log_2(M)}{M} - \left(1 - \frac{1}{M}\right) \cdot \log_2\left(1 - \frac{1}{M}\right) \\ &= \mathcal{O}\left(\frac{\log_2(M)}{M}\right) \end{aligned}$$

□

Lemma 5. *Let $B_1, \ldots, B_d$ be d random variables, each being uniformly distributed over $\mathbb{F}^\star$, and let $B = \sum_{i=1}^{d} B_i$. Then,*

$$\mathsf{H}(B) = \log_2(M) + \mathcal{O}\left(\frac{1}{M^d}\right) \tag{37}$$

Proof. Let p be the p.m.f. of one non-zero share, *i.e.*, $\mathsf{p} = (0, \frac{1}{M-1}, \ldots, \frac{1}{M-1})$. Denoting the uniform p.m.f. by u and $\mathsf{m} = (1, 0, \ldots, 0)$, observe that $\mathsf{u} = \frac{1}{M} \cdot \mathsf{m} + \frac{M-1}{M} \cdot \mathsf{p}$, *i.e.*, $\mathsf{p} = \frac{1}{M-1} \cdot (M \cdot \mathsf{u} - \mathsf{m})$. Let $\mathsf{p}^{*\,d}$ be the p.m.f. p convoluted d

times with itself. Using the fact that for any p.m.f. p', we have that $\mathsf{p}' * \mathsf{u} = \mathsf{u}$, and that $\mathsf{m} * \mathsf{m} = \mathsf{m}$, we get that:

$$\begin{aligned}
\mathsf{p}^{*\,d} &= \frac{1}{(M-1)^d} \cdot \sum_{k=0}^{d} \binom{d}{k} \cdot M^k \cdot \overset{*\,k}{\mathsf{u}} * \mathsf{m}^{*\,d-k} \cdot (-1)^{d-k} \\
&= \frac{1}{(M-1)^d} \cdot \left((-1)^d \cdot \mathsf{m} + (-1)^d \cdot \sum_{k=1}^{d} \binom{d}{k} \cdot (-M)^k \cdot \mathsf{u} \right) \\
&= \frac{1}{(M-1)^d} \cdot \left((-1)^d \cdot \mathsf{m} + \left[(M-1)^d - (-1)^d \right] \cdot \mathsf{u} \right) \\
&= \mathsf{u} \cdot \left(1 - \frac{(-1)^d}{(M-1)^d} \right) + \mathsf{m} \cdot \frac{(-1)^d}{(M-1)^d} \quad .
\end{aligned}$$

In other words, $\Pr(\mathrm{B}=0) = \frac{1}{M}\left(1 + \frac{(-1)^d}{(M-1)^{d-1}}\right)$, and for any $b \neq 0$, we have $\Pr(\mathrm{B}=b) = \frac{1}{M}\left(1 - \frac{(-1)^d}{(M-1)^d}\right)$. Plugging this into Lemma 6 gives the desired result. □

Lemma 6. *Let* $\mathsf{p} = \mathsf{u} + \frac{1}{M} \cdot (\epsilon, \epsilon', \ldots, \epsilon')$*, such that* $\epsilon' = \mathcal{O}\left(\frac{\epsilon}{M}\right)$*. Then,* $\mathsf{H}(\mathsf{p}) = \log_2(M) + \mathcal{O}\left(\frac{\epsilon}{M}\right)$*.*

Proof. Let $p = \frac{1}{M}(1+\epsilon)$. Observe first that

$$\begin{aligned}
-p\log(p) &= p \cdot \log\left(\frac{M}{1+\epsilon}\right) \\
&= p \cdot \log(M) - p \cdot \left(\epsilon + \mathcal{O}\left(\epsilon^2\right)\right) \\
&= p \cdot \log(M) - \left(\frac{1}{M} + \frac{\epsilon}{M}\right) \cdot \left(\epsilon + \mathcal{O}\left(\epsilon^2\right)\right) \\
&= p \cdot \log(M) - \frac{\epsilon}{M} + \mathcal{O}\left(\frac{\epsilon^2}{M}\right) \quad .
\end{aligned}$$

Summing the latter expansion for every probability of p, we get that

$$\begin{aligned}
\mathsf{H}(\mathsf{p}) = -\sum_{i=0}^{M-1} \mathsf{p}_i \log(\mathsf{p}_i) &= \log(M) \cdot \sum_{i=0}^{M-1} \mathsf{p}_i - \sum_{i=0}^{M-1} \left(\frac{\epsilon_i}{M} + \mathcal{O}\left(\frac{\epsilon_i^2}{M}\right) \right) \\
&= \log(M) + \mathcal{O}\left(\frac{\epsilon}{M}\right) + \sum_{i=1}^{M-1} \mathcal{O}\left(\frac{\epsilon'}{M}\right) \\
&= \log(M) + \mathcal{O}\left(\frac{\epsilon}{M}\right) \quad .
\end{aligned}$$

□

Proof of Proposition 2. The adversary may observe $d+1$ types of leakage, namely when $0 \leq k \leq d$ columns of the cross-product matrix are revealed.

– When $k = d$ columns are revealed, we know all the shares of A and since all the columns are revealed, this also means that all the shares of B are equal to zero. Hence, the conditional entropy is equal to 0.
– When $0 < k < d$ columns are revealed, A is fully revealed to the adversary, along with the k shares of B that are equal to zero. In other words, B is the sum of $d - k$ shares uniformly distributed over $\mathbb{F}^\star$. Hence,the entropy is equal to $\log_2(M) + \mathcal{O}\big(\frac{1}{M^{d-k}}\big)$, according to Lemma 5.
– When $k = 0$ columns are revealed, A remains fully uniform to the adversary, and B is the sum of d shares uniformly distributed over $\mathbb{F}^\star$. Hence, the entropy is equal to $2\log_2(M) + \mathcal{O}\big(\frac{1}{M^d}\big)$, according to Lemma 5.

Furthermore, the probability that k columns are revealed follows a binomial distribution, *i.e.*, $\Pr(\text{“}k \text{ columns revealed”}) = \binom{d}{k}\frac{1}{M^k}\left(1 - \frac{1}{M}\right)^{d-k}$. Hence,

$$\begin{aligned}\mathsf{H}(\mathrm{A},\mathrm{B} \mid \mathbf{L}) = \left(1 - \frac{1}{M}\right)^d \cdot \left(2\log_2(M) + \mathcal{O}\left(\frac{1}{M^d}\right)\right) \\ + \sum_{k=1}^{d-1}\left(\log_2(M) + \mathcal{O}\left(\frac{1}{M^{d-k}}\right)\right) \cdot \binom{d}{k} \cdot \frac{1}{M^k}\left(1 - \frac{1}{M}\right)^{d-k} \quad (38)\end{aligned}$$

That is,

$$\begin{aligned}\mathsf{H}(\mathrm{A},\mathrm{B} \mid \mathbf{L}) &= \log_2(M) \cdot \left(2\left(1 - \frac{1}{M}\right)^d + \sum_{k=1}^{d-1}\binom{d}{k}\frac{1}{M^k}\left(1 - \frac{1}{M}\right)^{d-k}\right) + \mathcal{O}\left(\left(\frac{2}{M}\right)^d\right) \\ &= \log_2(M) \cdot \left(2\left(1 - \frac{1}{M}\right)^d + 1 - \frac{1}{M^d} - \left(1 - \frac{1}{M}\right)^d\right) + \mathcal{O}\left(\left(\frac{2}{M}\right)^d\right) \\ &= 2\log_2(M) - \frac{d \cdot \log_2(M)}{M} + \mathcal{O}\left(\frac{\log_2(M)}{M^2}\right)\end{aligned}$$

This implies that $\mathsf{MI}(\mathrm{A},\mathrm{B};\mathbf{L}) = \frac{d\cdot\log_2(M)}{M} + \mathcal{O}\left(\frac{\log_2(M)}{M^2}\right)$.

□

References

1. Ajtai, M.: Secure computation with information leaking to an adversary. In: Fortnow, L., Vadhan, S.P. (eds.) 43rd Annual ACM Symposium on Theory of Computing. pp. 715–724. ACM Press, San Jose, CA, USA (Jun 6–8, 2011). https://doi.org/10.1145/1993636.1993731 21
2. Ananth, P., Ishai, Y., Sahai, A.: Private circuits: A modular approach. In: Shacham, H., Boldyreva, A. (eds.) Advances in Cryptology – CRYPTO 2018, Part III. Lecture Notes in Computer Science, vol. 10993, pp. 427–455. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 19–23, 2018). https://doi.org/10.1007/978-3-319-96878-0_15 21
3. Andrychowicz, M., Dziembowski, S., Faust, S.: Circuit compilers with $O(1/\log(n))$ leakage rate. In: Fischlin, M., Coron, J.S. (eds.) Advances in Cryptology – EUROCRYPT 2016, Part II. Lecture Notes in Computer Science, vol. 9666, pp. 586–615. Springer, Heidelberg, Germany, Vienna, Austria (May 8–12, 2016). https://doi.org/10.1007/978-3-662-49896-5_21 21
4. Azouaoui, M., Bellizia, D., Buhan, I., Debande, N., Duval, S., Giraud, C., Jaulmes, E., Koeune, F., Oswald, E., Standaert, F.X., Whitnall, C.: A systematic appraisal of side channel evaluation strategies. Cryptology ePrint Archive, Report 2020/1347 (2020), https://eprint.iacr.org/2020/1347 5
5. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In: Lee, P.J. (ed.) Advances in Cryptology – ASIACRYPT 2004. Lecture Notes in Computer Science, vol. 3329, pp. 432–450. Springer, Heidelberg, Germany, Jeju Island, Korea (Dec 5–9, 2004). https://doi.org/10.1007/978-3-540-30539-2_31 9
6. Barthe, G., Belaïd, S., Dupressoir, F., Fouque, P.A., Grégoire, B., Strub, P.Y., Zucchini, R.: Strong non-interference and type-directed higher-order masking. In: Weippl, E.R., Katzenbeisser, S., Kruegel, C., Myers, A.C., Halevi, S. (eds.) ACM CCS 2016: 23rd Conference on Computer and Communications Security. pp. 116–129. ACM Press, Vienna, Austria (Oct 24–28, 2016). https://doi.org/10.1145/2976749.2978427 3
7. Battistello, A., Coron, J.S., Prouff, E., Zeitoun, R.: Horizontal side-channel attacks and countermeasures on the ISW masking scheme. In: Gierlichs, B., Poschmann, A.Y. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2016. Lecture Notes in Computer Science, vol. 9813, pp. 23–39. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–19, 2016). https://doi.org/10.1007/978-3-662-53140-2_2 21
8. Belaïd, S., Coron, J.S., Prouff, E., Rivain, M., Taleb, A.R.: Random probing security: Verification, composition, expansion and new constructions. In: Micciancio, D., Ristenpart, T. (eds.) Advances in Cryptology – CRYPTO 2020, Part I. Lecture Notes in Computer Science, vol. 12170, pp. 339–368. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–21, 2020). https://doi.org/10.1007/978-3-030-56784-2_12 3, 21
9. Belaïd, S., Rivain, M., Taleb, A.R.: On the power of expansion: More efficient constructions in the random probing model. In: Canteaut, A., Standaert, F.X. (eds.) Advances in Cryptology – EUROCRYPT 2021, Part II. Lecture Notes in Computer Science, vol. 12697, pp. 313–343. Springer, Heidelberg, Germany, Zagreb, Croatia (Oct 17–21, 2021). https://doi.org/10.1007/978-3-030-77886-6_11 3
10. Belaïd, S., Rivain, M., Taleb, A.R., Vergnaud, D.: Dynamic random probing expansion with quasi linear asymptotic complexity. In: Tibouchi, M., Wang, H. (eds.)

Advances in Cryptology – ASIACRYPT 2021, Part II. Lecture Notes in Computer Science, vol. 13091, pp. 157–188. Springer, Heidelberg, Germany, Singapore (Dec 6–10, 2021). https://doi.org/10.1007/978-3-030-92075-3_6 3
11. Belaïd, S., Cassiers, G., Mutschler, C., Rivain, M., Roche, T., Standaert, F.X., Taleb, A.R.: A methodology to achieve provable side-channel security in real-world implementations. Cryptology ePrint Archive, Paper 2023/1198 (2023), https://eprint.iacr.org/2023/1198, https://eprint.iacr.org/2023/1198 4
12. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities: A nonasymptotic theory of independence. Oxford university press (2013) 22
13. Broughan, K.A.: The gcd-sum function. J. Integer Seq. **4**(2.2) (2001) 14
14. Bruneau, N., Guilley, S., Najm, Z., Teglia, Y.: Multivariate high-order attacks of shuffled tables recomputation. Journal of Cryptology **31**(2), 351–393 (Apr 2018). https://doi.org/10.1007/s00145-017-9259-7 23
15. Béguinot, J., Cheng, W., Guilley, S., Liu, Y., Masure, L., Rioul, O., Standaert, F.X.: Removing the field size loss from duc et al.'s conjectured bound for masked encodings. Cryptology ePrint Archive, Paper 2022/1738 (2022), https://eprint.iacr.org/2022/1738, https://eprint.iacr.org/2022/1738 5, 8, 15
16. Béguinot, J., Cheng, W., Guilley, S., Rioul, O.: Formal security proofs via doeblin coefficients: Optimal side-channel factorization from noisy leakage to random probing. Cryptology ePrint Archive, Paper 2024/199 (2024), https://eprint.iacr.org/2024/199, https://eprint.iacr.org/2024/199 6, 20
17. Cassiers, G., Faust, S., Orlt, M., Standaert, F.X.: Towards tight random probing security. In: Malkin, T., Peikert, C. (eds.) Advances in Cryptology – CRYPTO 2021, Part III. Lecture Notes in Computer Science, vol. 12827, pp. 185–214. Springer, Heidelberg, Germany, Virtual Event (Aug 16–20, 2021). https://doi.org/10.1007/978-3-030-84252-9_7 3
18. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards sound approaches to counteract power-analysis attacks. In: Wiener [53], pp. 398–412. https://doi.org/10.1007/3-540-48405-1_26 2
19. Cheng, W., Liu, Y., Guilley, S., Rioul, O.: Attacking masked cryptographic implementations: Information-theoretic bounds. In: IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022. pp. 654–659. IEEE (2022). https://doi.org/10.1109/ISIT50566.2022.9834556, https://doi.org/10.1109/ISIT50566.2022.9834556 9
20. Coron, J.S.: Higher order masking of look-up tables. In: Nguyen and Oswald [44], pp. 441–458. https://doi.org/10.1007/978-3-642-55220-5_25 23
21. Coron, J.S., Prouff, E., Rivain, M., Roche, T.: Higher-order side channel security and mask refreshing. In: Moriai [43], pp. 410–424. https://doi.org/10.1007/978-3-662-43933-3_21 3
22. Coron, J.S., Rondepierre, F., Zeitoun, R.: High order masking of look-up tables with common shares. IACR Transactions on Cryptographic Hardware and Embedded Systems **2018**(1), 40–72 (2018). https://doi.org/10.13154/tches.v2018.i1.40-72, https://tches.iacr.org/index.php/TCHES/article/view/832 23
23. Cover, T.M., Thomas, J.A.: Elements of information theory (2. ed.). Wiley (2006) 25, 27, 28
24. de Chérisey, E., Guilley, S., Rioul, O., Piantanida, P.: Best information is most successful. IACR Transactions on Cryptographic Hardware and Embedded Systems **2019**(2), 49–79 (2019). https://doi.org/10.13154/tches.v2019.i2.49-79, https://tches.iacr.org/index.php/TCHES/article/view/7385 9

25. Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: From probing attacks to noisy leakage. In: Nguyen and Oswald [44], pp. 423–440. https://doi.org/10.1007/978-3-642-55220-5_24 3, 4, 8, 21, 22
26. Duc, A., Faust, S., Standaert, F.X.: Making masking security proofs concrete - or how to evaluate the security of any leaking device. In: Oswald, E., Fischlin, M. (eds.) Advances in Cryptology – EUROCRYPT 2015, Part I. Lecture Notes in Computer Science, vol. 9056, pp. 401–429. Springer, Heidelberg, Germany, Sofia, Bulgaria (Apr 26–30, 2015). https://doi.org/10.1007/978-3-662-46800-5_16 4, 5, 9, 22
27. Dziembowski, S., Faust, S., Skorski, M.: Noisy leakage revisited. In: Oswald, E., Fischlin, M. (eds.) Advances in Cryptology – EUROCRYPT 2015, Part II. Lecture Notes in Computer Science, vol. 9057, pp. 159–188. Springer, Heidelberg, Germany, Sofia, Bulgaria (Apr 26–30, 2015). https://doi.org/10.1007/978-3-662-46803-6_6 3, 4, 21, 22, 23
28. Dziembowski, S., Faust, S., Skórski, M.: Optimal amplification of noisy leakages. In: Kushilevitz, E., Malkin, T. (eds.) TCC 2016-A: 13th Theory of Cryptography Conference, Part II. Lecture Notes in Computer Science, vol. 9563, pp. 291–318. Springer, Heidelberg, Germany, Tel Aviv, Israel (Jan 10–13, 2016). https://doi.org/10.1007/978-3-662-49099-0_11 5, 23, 29
29. Goubin, L., Patarin, J.: DES and differential power analysis (the "duplication" method). In: Koç, Çetin Kaya., Paar, C. (eds.) Cryptographic Hardware and Embedded Systems – CHES'99. Lecture Notes in Computer Science, vol. 1717, pp. 158–172. Springer, Heidelberg, Germany, Worcester, Massachusetts, USA (Aug 12–13, 1999). https://doi.org/10.1007/3-540-48059-5_15 2
30. Goudarzi, D., Joux, A., Rivain, M.: How to securely compute with noisy leakage in quasilinear complexity. In: Peyrin, T., Galbraith, S. (eds.) Advances in Cryptology – ASIACRYPT 2018, Part II. Lecture Notes in Computer Science, vol. 11273, pp. 547–574. Springer, Heidelberg, Germany, Brisbane, Queensland, Australia (Dec 2–6, 2018). https://doi.org/10.1007/978-3-030-03329-3_19 21
31. Goudarzi, D., Prest, T., Rivain, M., Vergnaud, D.: Probing security through input-output separation and revisited quasilinear masking. IACR Transactions on Cryptographic Hardware and Embedded Systems **2021**(3), 599–640 (2021). https://doi.org/10.46586/tches.v2021.i3.599-640, https://tches.iacr.org/index.php/TCHES/article/view/8987 21
32. Goudarzi, D., Rivain, M.: How fast can higher-order masking be in software? In: Coron, J.S., Nielsen, J.B. (eds.) Advances in Cryptology – EUROCRYPT 2017, Part I. Lecture Notes in Computer Science, vol. 10210, pp. 567–597. Springer, Heidelberg, Germany, Paris, France (Apr 30 – May 4, 2017). https://doi.org/10.1007/978-3-319-56620-7_20 10
33. Grosso, V., Standaert, F.X.: Masking proofs are tight and how to exploit it in security evaluations. In: Nielsen, J.B., Rijmen, V. (eds.) Advances in Cryptology – EUROCRYPT 2018, Part II. Lecture Notes in Computer Science, vol. 10821, pp. 385–412. Springer, Heidelberg, Germany, Tel Aviv, Israel (Apr 29 – May 3, 2018). https://doi.org/10.1007/978-3-319-78375-8_13 5, 21
34. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: Securing hardware against probing attacks. In: Boneh, D. (ed.) Advances in Cryptology – CRYPTO 2003. Lecture Notes in Computer Science, vol. 2729, pp. 463–481. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–21, 2003). https://doi.org/10.1007/978-3-540-45146-4_27 2, 3

35. Ito, A., Ueno, R., Homma, N.: On the success rate of side-channel attacks on masked implementations: Information-theoretical bounds and their practical usage. In: Yin, H., Stavrou, A., Cremers, C., Shi, E. (eds.) ACM CCS 2022: 29th Conference on Computer and Communications Security. pp. 1521–1535. ACM Press, Los Angeles, CA, USA (Nov 7–11, 2022). https://doi.org/10.1145/3548606.3560579 5, 8
36. Jog, V.S., Anantharam, V.: The entropy power inequality and mrs. gerber's lemma for groups of order 2^n. IEEE Trans. Inf. Theory **60**(7), 3773–3786 (2014). https://doi.org/10.1109/TIT.2014.2317692, https://doi.org/10.1109/TIT.2014.2317692 15
37. Kocher, P.C.: Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In: Koblitz, N. (ed.) Advances in Cryptology – CRYPTO'96. Lecture Notes in Computer Science, vol. 1109, pp. 104–113. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 18–22, 1996). https://doi.org/10.1007/3-540-68697-5_9 1
38. Kocher, P.C., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener [53], pp. 388–397. https://doi.org/10.1007/3-540-48405-1_25 1
39. Mangard, S.: Hardware countermeasures against DPA – A statistical analysis of their effectiveness. In: Okamoto, T. (ed.) Topics in Cryptology – CT-RSA 2004. Lecture Notes in Computer Science, vol. 2964, pp. 222–235. Springer, Heidelberg, Germany, San Francisco, CA, USA (Feb 23–27, 2004). https://doi.org/10.1007/978-3-540-24660-2_18 9
40. Mangard, S., Oswald, E., Popp, T.: Power analysis attacks - revealing the secrets of smart cards. Springer (2007) 9
41. Mangard, S., Oswald, E., Standaert, F.: One for all - all for one: unifying standard differential power analysis attacks. IET Inf. Secur. **5**(2), 100–110 (2011). https://doi.org/10.1049/iet-ifs.2010.0096, https://doi.org/10.1049/iet-ifs.2010.0096 9
42. Masure, L., Rioul, O., Standaert, F.X.: A nearly tight proof of duc et al.'s conjectured security bound for masked implementations. Cryptology ePrint Archive, Paper 2022/600 (2022), https://eprint.iacr.org/2022/600, https://eprint.iacr.org/2022/600 5, 8
43. Moriai, S. (ed.): Fast Software Encryption – FSE 2013, Lecture Notes in Computer Science, vol. 8424. Springer, Heidelberg, Germany, Singapore (Mar 11–13, 2014) 34, 37
44. Nguyen, P.Q., Oswald, E. (eds.): Advances in Cryptology – EUROCRYPT 2014, Lecture Notes in Computer Science, vol. 8441. Springer, Heidelberg, Germany, Copenhagen, Denmark (May 11–15, 2014) 34, 35
45. Prest, T., Goudarzi, D., Martinelli, A., Passelègue, A.: Unifying leakage models on a Rényi day. In: Boldyreva, A., Micciancio, D. (eds.) Advances in Cryptology – CRYPTO 2019, Part I. Lecture Notes in Computer Science, vol. 11692, pp. 683–712. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 18–22, 2019). https://doi.org/10.1007/978-3-030-26948-7_24 3, 4, 16, 18, 20, 21, 23
46. Prouff, E., Rivain, M.: Masking against side-channel attacks: A formal security proof. In: Johansson, T., Nguyen, P.Q. (eds.) Advances in Cryptology – EUROCRYPT 2013. Lecture Notes in Computer Science, vol. 7881, pp. 142–159. Springer, Heidelberg, Germany, Athens, Greece (May 26–30, 2013). https://doi.org/10.1007/978-3-642-38348-9_9 2, 3, 4, 5, 7, 9, 11, 18, 20, 21, 23
47. Rivain, M.: On the provable security of cryptographic implementations : Habilitation thesis. Personal website (2022), https://www.matthieurivain.com/hdr.html, https://www.matthieurivain.com/hdr.html 4, 7

48. Rivain, M., Prouff, E.: Provably secure higher-order masking of AES. In: Mangard, S., Standaert, F.X. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2010. Lecture Notes in Computer Science, vol. 6225, pp. 413–427. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–20, 2010). https://doi.org/10.1007/978-3-642-15031-9_28 3, 10, 12, 13
49. Cardoso dos Santos, L., Gérard, F., Großschädl, J., Spignoli, L.: Rivain-Prouff on steroids: Faster and stronger masking of the AES. In: Buhan, I., Schneider, T. (eds.) Smart Card Research and Advanced Applications. pp. 123–145. Springer International Publishing, Cham (2023) 10, 13
50. Shulman, N., Feder, M.: The uniform distribution as a universal prior. IEEE Trans. Inf. Theory **50**(6), 1356–1362 (2004). https://doi.org/10.1109/TIT.2004.828152, https://doi.org/10.1109/TIT.2004.828152 11, 12, 28
51. Standaert, F.X., Malkin, T., Yung, M.: A unified framework for the analysis of side-channel key recovery attacks. In: Joux, A. (ed.) Advances in Cryptology – EUROCRYPT 2009. Lecture Notes in Computer Science, vol. 5479, pp. 443–461. Springer, Heidelberg, Germany, Cologne, Germany (Apr 26–30, 2009). https://doi.org/10.1007/978-3-642-01001-9_26 3, 21
52. Tunstall, M., Whitnall, C., Oswald, E.: Masking tables - an underestimated security risk. In: Moriai [43], pp. 425–444. https://doi.org/10.1007/978-3-662-43933-3_22 23
53. Wiener, M.J. (ed.): Advances in Cryptology – CRYPTO'99, Lecture Notes in Computer Science, vol. 1666. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 15–19, 1999) 34, 36
54. Wyner, A.D., Ziv, J.: A theorem on the entropy of certain binary sequences and applications-i. IEEE Trans. Inf. Theory **19**(6), 769–772 (1973). https://doi.org/10.1109/TIT.1973.1055107, https://doi.org/10.1109/TIT.1973.1055107 15